.

Lecture Notes for

Math 341: Introduction to Analysis

Michael Frazier
Department of Mathematics
University of Tennessee

Contents

Chapter 1

Analysis: Introduction and Motivation

This course is sometimes titled "Advanced Calculus" and described as the theory behind calculus: limits, convergence, continuity, differentiation and integration, presented logically with proofs, all based on a minimal set of assumptions about the real numbers. This description is accurate but misses the main point. If analysis were to end with justifying calculus, we would not bother to ask every math major to take this course. Instead, we would accept the word of the earlier scholars that calculus is valid, just as we don't ask every physics major to personally carry out every experiment that supports the atomic theory of matter. The real reason math majors take this course is that the ideas of advanced calculus are needed in order to go further in most fields of mathematics - for example in differential equations, applied and computational mathematics, and mathematical physics. In this course we study ideas like convergence in the context of points (numbers) on the real line. This allows students to get an understanding of these concepts in a familiar setting, where their intuition is usually valid. Later in mathematics one needs to study convergence in more complicated, infinite dimensional spaces, such as the space $\mathcal{C}([0,1])$ consisting of all continuous functions on the interval $[0,1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$, where $\mathbb{R}$ denotes the set of real numbers. By learning to reason rigorously first in relatively simple situations, students develop the skills to reason accurately in more complex settings which may be unintuitive or even counter-intuitive.

Let's start with an example which shows why the notions of limits and convergence are important. For this and the following 2 examples, it is not expected that you will understand them completely at this time. If you do understand them fully, then you probably don't need to take this course. The examples are designed to give you a glimpse of how some of the topics that we will consider later are used in applications and more advanced mathematics.

Example 1.0.1 Calculating square roots

Suppose that you were working at the beginning of the computational age, and your assignment was to find an algorithm for computers or calculators to efficiently calculate square roots. You assume the calculator can already do addition, subtraction, multiplication, and division, and now you want to find square roots. Some of you may have learned an algorithm, somewhat like long division, but more complicated, to directly compute square roots by hand. That algorithm is very slow and difficult to automate. Another procedure is successive approximation: make an educated guess, and repeatedly increase or decrease the guess if squaring it gives too small or large a result, respectively. That method is also slow and hard to program. Here is an alternative. Given a postive number C, our goal is to compute $\sqrt{C}$. (Remark: in math, $\sqrt{C}$ always denotes the *positive* number whose square is C; the negative number with square C is denoted $-\sqrt{C}$).

Step 1: Make a reasonable approximation to $\sqrt{C}$ by guessing. Call that number x_1. Part of the definition of "reasonable" here is that x_1 should be positive.

Step 2: Let $x_2 = \dfrac{1}{2}\left(x_1 + \dfrac{C}{x_1}\right)$.

Step 3: Let $x_3 = \frac{1}{2}\left(x_2 + \frac{C}{x_2}\right)$.

Now continue in this way: for each $n \in \mathbb{N}$ (here $\mathbb{N}$ is the set of natural numbers $\{1, 2, 3, \ldots\}$; i.e., the strictly positive integers), given the previous value x_n, let

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{C}{x_n}\right). \tag{1.1}$$

In this way we obtain a sequence $(x_1, x_2, x_3, \ldots) = (x_n)_{n=1}^{\infty}$. A sequence obtained by determining each value from the previous one is said to be defined *recursively*. Recursive sequences are common in computer science. We remark that the formula (1.1) does not come out of nowhere; it is what one gets in this example from Newton's method for approximating roots of functions.

It is worth trying this process with $C = 2$, for example. Let't take $x_1 = 1$, for simplicity. You will get $x_2 = 1.5, x_3 = 1.4166666$ and $x_4 = 1.4142156$. We see that pretty quickly, the values of x_n start to approach $1.4142135\ldots$, which you may know is approximately $\sqrt{2}$. This process raises several questions. First, how do we know x_n is really approaching ("converging") to $\sqrt{2}$? Second, will it work for any number $C > 0$? You don't want to program an algorithm into your computer that sometimes gives the wrong answer. We can't obtain a guarantee from the fact that the formula comes from Newton's method, because there are examples where Newton's method does not converge.

To get some insight, let's assume for the moment that the sequence x_n converges to some number x (the definition of "converges," which is surprisingly difficult to make precise, will come later in the course, but let's work with the idea intuitively right now). Then x_n approaches x as $n \to \infty$, which means also that x_{n+1} approaches x as $n \to \infty$. Let's take the limit as $n \to \infty$ on both sides of the equation

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{C}{x_n}\right).$$

Assuming the basic properties of limits that we will prove later, we get

$$x = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} \frac{1}{2}\left(x_n + \frac{C}{x_n}\right) = \frac{1}{2}\left(\lim_{n\to\infty} x_n + \lim_{n\to\infty} \frac{C}{x_n}\right) = \frac{1}{2}\left(x + \frac{C}{x}\right).$$

Solving this equation for x, we get $2x = x + \frac{C}{x}$, or $x = \frac{C}{x}$, or $x^2 = C$, so (since x is postive, since every x_n is positive) we get $x = \sqrt{C}$. So, if we know that x_n converges, then we know the limit is $\sqrt{C}$. Then by taking n large enough, we can approximate $\sqrt{C}$ as accurately as needed.

Do we know that x_n always converges, for any C and no matter what positive number x_1 we choose? It may seem reasonable to just assume this always works, but there are recursively defined sequences which do not converge. For a simple example, define $x_1 = 2$ and define x_n for all $n \in \mathbb{N}$ recursively by the formula $x_{n+1} = 6 - x_n$. Then we get

$x_2 = 6 - 2 = 4,$
$x_3 = 6 - 4 = 2,$
$x_4 = 6 - 2 = 4,$
$x_5 = 6 - 4 = 2,$

and so on. That is, for n even, we get $x_n = 4$, and for n odd, we get $x_n = 2$. In other words, the sequence is $(2, 4, 2, 4, 2, 4, \ldots)$, which obviously does not converge. So there is no general guarantee that recursively defined sequences converge.

Nevertheless, it may be that the sequence defined by (1.1) converges for any choice $x_1 > 0$. In fact, with some work we can show that the sequence x_n has two key properties, as follows.

(a): Except possibly for x_1, we have $x_n \geq \sqrt{C}$. To see this, we claim that for any $x > 0$, we have $\frac{1}{2}\left(x + \frac{C}{x}\right) \geq \sqrt{C}$, which proves that $x_2 \geq \sqrt{C}, x_3 \geq \sqrt{C}$, etc., since each is of the form $\frac{1}{2}\left(x + \frac{C}{x}\right)$ for some $x > 0$. To see the claim, start from the fact that $(x - \sqrt{C})^2 \geq 0$. Expanding, we get $x^2 - 2x\sqrt{C} + C \geq 0$, hence $x^2 + C \geq 2x\sqrt{C}$. Dividing by $2x$ gives $\frac{1}{2}\left(x + \frac{C}{x}\right) \geq \sqrt{C}$.

(b) For $n \geq 2$, we have $x_{n+1} \leq x_n$. To see this fact, by (a) we know that $x_n \geq \sqrt{C}$. Therefore $x_n^2 \geq C$, so $x_n \geq \frac{C}{x_n}$ (here we are using the fact that $x_n > 0$). Then $2x_n \geq x_n + \frac{C}{x_n}$, so $x_n \geq \frac{1}{2}\left(x_n + \frac{C}{x_n}\right) = x_{n+1}$.

Except possibly for x_1, whose value doesn't affect whether the sequence x_n converges as $n \to \infty$, we have by (a) that the sequence x_n is bounded below, and by (b) the sequence is decreasing (in the sense of being non-increasing). An important theorem that we will prove later states that a decreasing, bounded below sequence of real numbers (or an increasing, bounded above sequence) must converge to some real number. This may seem obvious: if you are heading toward a brick wall, either you stop before the brick wall, or at the brick wall, but you certainly stop. However, we will see that this fact depends on the deep property that the real numbers are "complete," which means intuitively that there are no gaps or missing numbers on the real line. In any case, once we have this theorem, we will know that the algorithm above for computing square roots always works. For a practical algorithm, one would also need estimates on how rapidly the algorithm converges, which requires further analysis.

To understand the algorithm better, let's define the function

$$f(x) = \frac{1}{2}\left(x + \frac{C}{x}\right),$$

for $x > 0$. Then our sequence above is defined by $x_{n+1} = f(x_n)$. Taking the limit as $n \to \infty$ on both sides of the equation $x_{n+1} = f(x_n)$ (using the continuity of f for $x > 0$), gives $x = f(x)$, where x is the limit of x_n. So the desired value x is a "fixed point" of f, which means a value x such that

$$f(x) = x. \tag{1.2}$$

Note that if we start with $x = \sqrt{C}$ we get $f(x) = f(\sqrt{C}) = \frac{1}{2}\left(\sqrt{C} + \frac{C}{\sqrt{C}}\right) = \sqrt{C} = x$, so the number we are trying to calculate is in fact a fixed point of the function f.

Example 1.0.2 Existence of Solutions of Ordinary Differential Equations

Now let's consider a more complicated problem that turns out to have the same underlying structure as in Example 1.0.1.

Suppose we want to find a function y, which is a function of a real number x (which we denote by saying $y = y(x)$) satisfying

$$y' = xy^2 + y + x, \text{ and } y(0) = 2 \tag{1.3}$$

for x in an interval containing $x = 0$. (From Math 231 we know that we can't always have solutions which are defined for all x; for example, the function $y = \frac{1}{1-x}$ satisfies $y' = y^2$ and $y(0) = 1$, but this function is undefined at $x = 1$. So the most we can hope for in general is a local solution.) The problem cannot be solved just by integration because the unknown function y occurs on both sides of equation (1.3).

The more general problem is of the form

$$y' = f(x, y), \text{ and } y(x_0) = y_0, \tag{1.4}$$

where the function f and the numbers x_0 and y_0 are given. Except in simple cases, we can't write down a formula for a solution y. However it is still of interest to know whether there is a solution. Why? If we know there is a solution, but we can't find an explicit formula for it, we may still be able to approximate the solution numerically to good accuracy. But if there is no solution, then there is no point trying to find one by computational or any other methods. It turns out that there is a theorem (called the fundamental existence and uniqueness theorem for ordinary differential equations) that states that under pretty general conditions on f, which cover most naturally occuring examples, there is a unique solution to (1.4). We won't try to detail the precise statement of that theorem now, but we will sketch the idea behind the proof.

Let's replace the variable x by t in (1.4): $y'(t) = f(t, y(t))$, where we have made explicit the dependence of y on t on the right side of the equation. Then let's integrate from the base point x_0 to a general x near x_0:

$$\int_{x_0}^{x} y'(t)\, dt = \int_{x_0}^{x} f(t, y(t))\, dt.$$

From calculus (specifically, the fundamental theorem of calculus, whose proof is one of the main objectives of this course), we know that

$$\int_{x_0}^{x} y'(t)\,dt = y(x) - y(x_0).$$

So if y satisfies (1.4), we have $y(x_0) = y_0$, so $y(x) - y_0 = \int_{x_0}^{x} f(t, y(t))\,dt$, or

$$y(x) = y_0 + \int_{x_0}^{x} f(t, y(t))\,dt. \tag{1.5}$$

Conversely, if y satisfies (1.5) then $y(x_0) = y_0$ (since $\int_{x_0}^{x_0}(\cdots)\,dt = 0$ always), and taking the derivative on both sides of (1.5) and using the other part of the fundamental theorem of calculus, we get that y satisfies (1.4). Thus the problems (1.4) and (1.5) are equivalent in the sense that if y is a solution of either one of them, then y is a solution of the other. It may seem that reformulating the problem doesn't help, because we don't know how to explicitly solve (1.5) either, since again y appears on both sides of the equation. However, the second formulation has some advantages, as follows.

For any reasonable function $g = g(x)$, let's define another function $T(g) = T(g)(x)$ by the formula

$$T(g)(x) = y_0 + \int_{x_0}^{x} f(t, g(t))\,dt. \tag{1.6}$$

Looking at (1.5), then, we are looking for a function y such that

$$y = T(y). \tag{1.7}$$

That is, we are looking for function y which is a fixed point of T, which seems at first to be just like (1.2). However, there are important differences. In (1.2) we have a function f that takes as its input a number x, and returns as output the number $f(x)$. In (1.6), T takes as input a function g and returns as output a function $T(g)$. We call something like T that takes functions to functions an *operator*. But given the similarity of (1.7) and (1.2), we can try to find a solution to (1.7) the way we found a solution to (1.2), namely by iteration. That is, start with some function y_1, maybe a constant function for simplicity. Then let

$$y_2(x) = T(y_1)(x) = y_0 + \int_{x_0}^{x} f(t, y_1(t))\,dt.$$

Then continue iteratively:

$$y_3(x) = T(y_2)(x) = y_0 + \int_{x_0}^{x} f(t, y_2(t))\,dt,$$

etc., letting

$$y_{n+1}(x) = T(y_n)(x) = y_0 + \int_{x_0}^{x} f(t, y_n(t))\,dt,$$

for each $n = 2, 3, \ldots$. If it happens that y_n converges to some function y as $n \to \infty$, and if it also happens that $T(y_n)$ converges to $T(y)$, then we can let $n \to \infty$ in the equation $y_{n+1} = T(y_n)$ to get $y = T(y)$. Then y is a fixed point of T, and we get a solution y to (1.5), hence to (1.4).

But now we have to see if, in fact, y_n converges as $n \to \infty$. But each y_n is a function, so we first have to decide what it means for a sequence $(y_n)_{n=1}^{\infty}$ of functions to converge to a function y. It turns out that there are many possible ways to define such convergence, including pointwise convergence ($\lim_{n\to\infty} y_n(x) = y(x)$ for each x), uniform convergence ($\max_x |y_n(x) - y(x)| \to 0$ as $n \to \infty$), convergence in mean ($\int |y_n(x) - y(x)|\,dx \to 0$ as $n \to \infty$), and many others. The usual idea is to define a "distance" $d(f, g)$ between two functions f and g and say that y_n converges to y if $d(y_n, y) \to 0$ as $n \to \infty$. What notion of distance is right for our problem of solving ordinary differential equations? What class of functions should we work with; e.g., for what g is $T(g)$ defined? These are the kinds of questions which we build the background to answer in this course. It turns out that the right notion is uniform convergence, and we should work with the space of continuous functions on a interval around the point x_0, which, as alluded to earlier, is an infinite dimensional space. So ultimately it becomes necessary to understand convergence of elements in an infinite

dimensional space. But, in this class, we will stick with finite dimensional spaces, in fact the 1 dimensional space $\mathbb{R}$, which makes matters quite a bit easier.

We remark that although the fundamental existence and uniqueness theorem for ordinary differential equations is more than 100 years old, the technique of solving a differential equation by finding a fixed point of an associated operator is still being used to this day in research related to nonlinear partial differential equations.

By the way, the method for solving differential equations iteratively as described above not only provides a way to prove the existence of a solution, it provides a algorithm for computing the solution. That method is called *Picard iteration*. Moreover the same method applies to systems of ordinary differential equations. Going even further, first order partial differential equations (meaning equations relating a function of several variables and its first order partial derivatives) can be reduced to systems of ordinary differential equations, so the theory even shows us how to handle first order partial differential equations. However, the partial differential equations of most importance in basic mathematical physics are usually second order equations (i.e., equations involving at most second order partial derivatives), because of Newton's law $F = ma$ and the fact thats the acceleration a is the second derivative of position. In the next example we consider a second order partial differential equation of great importance.

Example 1.0.3 The one-dimensional heat equation and Fourier series

Diffusion is one of the most common phenomena in nature. A chemical dumped into a body of water will diffuse throughout the water. Medicines (or poisons) will diffuse in our bloodstream and body. Pollutants will diffuse in the atmosphere. Heat will diffuse in any object. Although these phenomena involve different materials, they have a common underlying nature, which is epressed mathematically in that the equations describing them have the same form. Thus once the mathematics of one of these phenomena is understood, so are the others. This is the fundamental efficiency of mathematics: mathematics discovers and exhibits the deep underlying unity of apparently diverse phenomena. To be specific, let's consider the diffusion of heat.

We would like to understand heat flow in a 3 dimensional region, but it is usually best in mathematics to start with a simpler case and work our way up to the most general. So we start with 1 dimensional heat conduction. We can visualize this problem by considering a thin rod or wire. Suppose the wire is wrapped with insulation, so that no heat escapes laterally from the wire. We start with some initial distribution of heat in the wire, and we want to see how that distribution evolves over time. There are different assumptions that can be made at the two ends of the wire. For simplicity, we consider "Dirichlet" boundary conditions, namely that the two ends are kept at the constant temperature of 0 for all time (for example, the insulated wire may be in room whose ambient temperature is 0). To make all of this mathematical, let $u = u(x,t)$ be the temperature in the wire at position x and time $t > 0$. Whatever length the wire is, we can rescale it to any number we want; it turns out to be most convenient to assume the length of the wire is π (this point is not obvious right now). So we let the position x vary from 0 to π, and we let time t satisfy $t \geq 0$. Some considerations from physics suggest that u should satisfy the equation $u_t = c^2 u_{xx}$, where u_t is the partial derivative of u with respect to t (i.e., $u_t = \frac{\partial u}{\partial t}$), u_{xx} is the second partial derivative of u with respect to x, and c is a postive constant relating to the material of the wire (c is the "thermal conductivity" of the wire). Our Dirichlet boundary conditions can be formulated as saying that $u(0,t)$ and $u(\pi,t)$ are both 0 for all time $t \geq 0$. We have to assume that we know the initial temperature, i.e., that $u(x,0)$ is some given function $f(x)$. So we can write our problem mathematically as trying to find a function $u(x,t)$ satisfying

$$\begin{cases} u_t = c^2 u_{xx} & \text{for } 0 \leq x \leq \pi, t \geq 0, \\ u(0,t) = u(\pi,t) = 0 & \text{for } t \geq 0, \\ u(x,0) = f(x) & \text{for } 0 \leq x \leq \pi. \end{cases} \tag{1.8}$$

It seems reasonable physically that a solution u should exist and it should be unique, but we would like to prove that this is the true; otherwise something is wrong with our model. The problem (1.8) is called the heat equation with Dirichlet boundary conditions for the interval $[0,\pi]$.

This problem was of great interest to mathematicians and physicists around 1800. The French Academy of Sciences offered a prize for the best paper on this topic. In 1807, the former civil servant Jean-Baptiste

Joseph Fourier (not to be confused with the utopian socialist Charles Fourier) submitted his paper on the question to the Academy. To very quickly summarize Fourier's ideas, he noticed that for each positive integer n, the function

$$u_n(x,t) = e^{-c^2n^2t}\sin(nx)$$

satisfies the first two conditions of (1.8). This fact is easy to check:

$$(u_n)_t = -c^2n^2e^{-c^2n^2t}\sin(nx),$$

$$(u_n)_x = ne^{-c^2n^2t}\cos(nx),$$

and hence

$$(u_n)_{xx} = -n^2e^{-c^2n^2t}\sin(nx),$$

so we can see that $c^2(u_n)_{xx} = (u_n)_t$, and also $u_n(x,0) = \sin(0) = 0$ and $u_n(x,\pi) = e^{-c^2n^2\pi}\sin(n\pi) = 0$ (this last point is why we let the length of the wire be π; if it is length L we need to replace $\sin(nx)$ by $\sin(n\pi x/L)$ and $e^{-c^2n^2t}$ by $e^{-c^2n^2\pi^2t/L^2}$, which is awkward). One way to find the form of u_n is to guess a solution of the form $F(t)G(x)$, called a "separated" solution because it's form is a function of t multiplied by a function of x. Substituting this form into the equation and using the boundary conditions leads to a pair of ordinary differential equations which can be solved to get the u_n's. This method, called "separation of variables," can be applied to many other equations, including the wave equation and Schrödinger's equation.

Fourier then suggested that any solution of (1.8) could be written as a superposition of these u_n. Since there are infinitely many u_n, this means that the solution u would have the form of an infinite series:

$$u(x,t) = \sum_{n=1}^{\infty} c_n u_n(x,t) = \sum_{n=1}^{\infty} c_n e^{-c^2n^2t}\sin(nx), \tag{1.9}$$

for some coefficients (numbers) c_n, $n = 1, 2, 3, \ldots$. We hope that the first and second conditions are still satisfied, by linearity, for such a sum (this would be true if the sum were finite, but it is not so clear for an infinite sum - that is one of the points that would have to be investigated). But the main problem is to satisfy the last condtion $u(x,0) = f(x)$ in (1.8). However, we are allowed to choose any coefficients c_n to try to attain $u(x,0) = f(x)$. Fourier set $t = 0$ in (1.9) to obtain

$$u(x,0) = \sum_{n=1}^{\infty} c_n u_n(x,0) = \sum_{n=1}^{\infty} c_n \sin(nx). \tag{1.10}$$

At this point Fourier made a radical statement: he asserted, without proof, that "any" function f can be written in the form (1.10) for some choice of c_n, $n \in \mathbb{N}$. He then derived an explicit formula for c_n in terms of f (in fact, $c_n = \frac{2}{\pi}\int_0^\pi f(x)\sin(nx)\,dx$), and claimed that he therefore had a solution to the original problem (1.8). The series of the form (1.10) with Fourier's choice of coefficients is now called the "Fourier series" of f.

The French Academy of Sciences, which included the famous mathematicians Lagrange and Laplace, did not know what to make of Fourier's assertions, especially the one that any function f can be written in the form (1.10). After a long delay, they eventually decided Fourier might be right, and he received the prize for his paper. But the question of validating or invalidating Fourier's assertion remained. The mathematical techniques to analyze this question simply did not exist in Fourier's time. Gradually, over the next 150 years, the tools were developed (including developing an entire new, more general theory of integration, called Lebesgue integration, around 1900) that allowed mathematicians to obtain a pretty good understanding of the question. It turns out to be quite subtle: the answer to the question "Does the Fourier series of f converge and equal f" depends on what class of functions f you are considering and exactly what you mean by "converge" - there are many possible senses of convergence for a sequence of functions. The theory of Fourier series developed over the previous 150 years was exposited in a large, classic volume called "Trigonmetric Series" by Antoni Zygmund, published in 1959. The biggest remaining open question about 1 dimensional Fourier series was settled in a notoriously difficult 1966 paper by the great mathematician Lennart Carleson.

Meeting Fourier's challenge of determining whether a function agrees with its Fourier series required mathematicians to develop precise definitions of many concepts central to analysis, such as limits, convergence, continuity, and the distance between functions in various senses. In fact, the modern definition of a function evolved from the questions raised by Fourier. You are not expected to understand all of this now. Our point is just that it turns out that to resolve questions relating to natural phenomena like diffusion, it is necessary to develop the basic tools of mathematical analysis. In this course we will introduce the fundamental concepts of analysis in the one-dimensional setting where we can best comprehend them at first. Once these ideas have been suffiently well understood in that context, they can be extended to deal with more advanced questions raised by deeper issues in physics and mathematics.

Chapter 2

Logic and Proofs

The fundamental structure of mathematics is to start with a set of assumptions (called "axioms") and proceed via logically valid steps to reach conclusions which we then know to be true (assuming the axioms are true). A logically valid argument establishing a conclusion is called a "proof." Proofs are the heart of pure mathematics. In this course we assume you have some familiarity with proofs, so here we just give a summary of the basic points about proofs that we will be using regularly.

A statement, usually denoted by lower case letters like p or q, is something that is either true or false. Sentences that are subjective ("life is sweet"), vague ("I am the greatest"), or ambiguous ("you can't put too much water into a nuclear reactor"), are not considered statements for our purposes. Suppose that x is a real number. An example of a statement p is that $x > 2$, which we write as:

$$p : x > 2. \tag{2.1}$$

If we know x, we can determine whether p is true or false. Another statment is

$$q : x^2 > 1. \tag{2.2}$$

A.) New Statements: negation, implication, converse, "and," "or," equivalences

Given a statement or statements, we can form additional statements in various ways. We can then determine whether each of these new statements is true or false if we know whether each original statement is true or false. This way of thinking can be formalized using Boolean algebra, or "truth tables," which, as we will see soon, provide a rigorous way of calculating with logical statements. In this course we will only use truth tables for a couple of key foundational principles (namely proof by contradiction and by contrapositive), instead dealing with logical statements more intuitively.

One statement that we can form from a given statement p is the "negation" of p, written $\sim p$ and often called "not p." The statement $\sim p$ is the statement that p is false. Formally, if p is true then $\sim p$ is false, and it p is false, then $\sim p$ is true. This definition is illustrated by the following truth table:

p	$\sim p$
T	F
F	T

The entries to the left of the double vertical bars are the given variables and their possible truth values. In this simple case, there is just one given variable, namely p. To the right of the double vertical bars are the conclusions we can draw about the truth values of the statements at the top of each column. Each row of the truth table represents one possible set of truth values. For this example, the first row states that if p is true, then $\sim p$ is false. The second line states that if p is false, then $\sim p$ is true.

Given two statements p and q, we can form the compound statement $p \Longrightarrow q$, which means that p implies q; that is, if p is true then q is true. Such a statement is called an implication. For the examples of p and q in (2.1) and (2.2), the statement $p \Longrightarrow q$ is true: if $x > 2$ then we know that $x^2 > 1$ (we will see how to

prove this fact based on axioms for the real numbers later, but for the current discussion we assume such basic facts).

The most important point of logic to keep in mind is that in order to prove $p \Longrightarrow q$, we assume only the truth of p and then make logical inferences that lead us to the truth of q. One of the most common logical mistakes is to start with the conclusion q, reason with it for a while, reach a result we know is true, and conclude that q must have been true because it led us to a valid conclusion. Such thinking is "inductive" reasoning, which is crucial in science. If a physical model leads to valid predictions, we have more faith in the model. However, making one valid prediction does not guarantee that a model is absolutely true; it just gives supporting evidence which suggests that the model might be true. In science you cannot be absolutely sure of anything. Mathematics is fundamentally different from science because it is based on deductive reasoning, in which the validity of the conclusion is certain if the assumptions are true and the argument is correct. A trivial example of the logical fallacy of inductive reasoning in mathematics is the following fallacious "proof" that $2 = 3$: Start with $2 = 3$. Then multiply both sides of the equation $2 = 3$ by -1 to get $-2 = -3$. Then add this equation to the equation $2 = 3$ to get the conclusion $0 = 0$. Since $0 = 0$ is certainly true, we conclude (inductively) that $2 = 3$. Obviously this argument is incorrect. Inductive reasoning is not conclusive in mathematics. In math, the validity of the assumptions guarantees the validity of the conclusion, but the validity of the conclusion does not guarantee the validity of the assumptions.

If p is true and q is true, the statement $p \Longrightarrow q$ is true, because the truth of p gives us the truth of q. Also, if p is true and q is false, we see that $p \Longrightarrow q$ is false; we have the truth of p but we don't get the truth of q, so $p \Longrightarrow q$ is not correct. It gets a bit more tricky to analyze the truth of the statement $p \Longrightarrow q$ in the case where p is false. If p is false, then $p \Longrightarrow q$ seems to be meaningless, because it only gives a conclusion if p is true. So it is tempting to not assign $p \Longrightarrow q$ any truth value, that is, to not say whether $p \Longrightarrow q$ is true or false. But, if we do that, then $p \Longrightarrow q$ is no longer a statement, since it is not true or false. Instead, the convention in mathematics (which turns out to be the most useful) is to make the convention that if p is false, then $p \Longrightarrow q$ is true whether q is true or false. This convention makes $p \Longrightarrow q$ a statement because it is either true or false in all possible cases for p and q. We say $p \Longrightarrow q$ is "vacuously true" in the case that p is false; the statement $p \Longrightarrow q$ is technically true but not meaningful. These conclusions are represented in the following truth table.

p	q	$p \Longrightarrow q$
T	T	T
T	F	F
F	T	T
F	F	T

The interpretation of this truth table is as before: each line tells what truth value is assigned to the quantities on the right of the double vertical lines, assuming the values on the left. For example, the second line tells us that if p is true and q is false, then the implication $p \Longrightarrow q$ is false.

The "converse" of the statement $p \Longrightarrow q$ is the statement $q \Longrightarrow p$. The converse is an independent statement. Knowing that $p \Longrightarrow q$ is true (or knowing it is false) does not in general tell us anything about whether $q \Longrightarrow p$ is true or false. In the specific example above, the statement $q \Longrightarrow p$ states that if $x^2 > 1$, then $x > 2$, which is false: for example, if $x = 1.5$ then $x^2 = 2.25 > 1$ but it is not true that x is greater than 2. Notice that in this example, the statement $q \Longrightarrow p$ can be stated using the quantifier "for all" as: for all real numbers x satisfying $x^2 > 1$, we have (or "it follows that") $x > 2$. To show that a "for all" statement about x is true, we would have to verify it for every possible choice of x, but to show that a "for all" statement is false, we just had to exhibit one value of x (a "counterexample," in this case $x = 1.5$) which satisfies the assumption q but not the conclusion p of the implication $q \Longrightarrow p$. In other words, the negation of a "for all" statement is a "there exists" statement. In this case, the negation of the statement that "for all x satifying $x^2 > 1$ we have $x > 2$" is the statement "there exists x satisfying $x^2 > 1$ but x is not greater than 2."

Perhaps the reason for our convention regarding vacuously true statements can be understood by considering "for all" statements. If a "for all" statement is false, there must be an example that violates the statement. For a real number x, consider the statement $x^2 < 0 \Longrightarrow x = 6$. Admittedly, this statement is bizarre, because there are no real numbers x satisfying $x^2 < 0$ (we will prove this fact later), and the conclusion that $x = 6$ seems ridiculous because $x = 6$ does not satisfy $x^2 < 0$. But, if the statement

$x^2 < 0 \implies x = 6$ is false, then there would have to exist a counterexample, namely some real number x satisfying the assumption $x^2 < 0$ but not the conclusion $x = 6$. But such a counterexample is impossible, since there is no real number x satisfying $x^2 < 0$ to start with. So we say the statement $x^2 < 0 \implies x = 6$ is true, albeit vacuously true. Another true statement would be: $x^2 < 0 \implies x = 6$ and $x = 7$, which is perhaps even more ridiculous.

Given two statements p and q, we can form the statement "p and q," written symbolically as $p \wedge q$, which means that both p and q are true. The statement $p \wedge q$ is only true if both p and q are true; $p \wedge q$ is false in any of the cases: (i) p is true and q is false, (ii) p is false and q is true, (iii) p is false and q is false. In other words, the following truth table holds.

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

Sometimes the notation "$\wedge$" is used to denote the minimum of two real numbers: e.g., $3 \wedge 4 = 3, 5 \wedge 4 = 4$, which we can write formally as $a \wedge b = \min(a, b)$. It may seem coincidental that the same notation is used for "and," but that is not a coincidence. If we associate the number 1 with a statement being true, and the number 0 with a statement being false, then the number associated with $p \wedge q$ is the minimum of the numbers associated with p and q separately, since the minimum will be 0 except for the one case where both p and q are true and hence are both associated with the number 1. In this way, the truth value of complicated statements can be determined by a numerical calculation that can be carried out via a computer program. Other statements can be dealt with computationally as well: for example, if the truth value of a statement p is x (where either $x = 1$ or $x = 0$), then the truth value of $\sim p$ is $1 - x$.

Another statement that we can form is "p or q," written $p \vee q$, which means that at least one of the statements p or q is true. The statement $p \vee q$ is true in the cases (i) p is true and q is true, (ii) p is true and q is false, (iii) p is false and q is true. The only case where $p \vee q$ is false is when p is false and q is false. That is, we have the truth table:

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

The symbol "$\vee$" is also used to denote the maximum of two real numbers: $a \vee b = \max(a, b)$. If a true statement is given the value 1 and a false statement is given the value 0, then the truth value of $p \vee q$ is the maximum of the truth values of p and q, since the truth value of $p \vee q$ is 1 unless the truth values of p and q are both 0.

In ordinary conversation, the word "or" can be used in different ways. In the sentence "you or I should take the part of Lincoln in the play," the word "or" is exclusive (i.e., it excludes both events): we can not both play Lincoln. But in the sentence "to ride this ride, you must be over 5 ft. tall or weigh more than 100 lbs," the word "or" is inclusive (i.e., allows both events), since one may well meet both criteria. Mathematics cannot tolerate such ambiguity, so the convention is to always use "or" inclusively: the statement "$p \vee q$" is true in the case where both p and q are true.

If $p \implies q$ is true and the converse $q \implies p$ is also true, we write $p \iff q$ and we say that p and q are logically equivalent (or just "equivalent," as in "p is equivalent to q"), since one is true if and only if the other is true. In other words, $p \iff q$ is equivalent to $(p \implies q) \wedge (q \implies p)$. In the example above given by (2.1) and (2.2), the statement $p \iff q$ is false, because, although $p \implies q$ is true, the converse $q \implies p$ is false. For a real number x, for example, the statements $x > 2$ and $x^3 > 8$ are equivalent. That doesn't mean that either statement $x > 2$ or $x^3 > 8$ is necessarily true; if $x = 1$ neither statement is true. It just means that if one of the statements is true, so is the other. An example of a valid equivalence is $p \iff \sim (\sim p)$.

The statement $p \iff q$ can be read as "p if and only if q." It is tempting but wrong to think that "$p \implies q$" is the same as "p if q." In fact "p if q" is the same as $q \implies p$, because both statements say that

when q holds, so must p. The statement $p \Longrightarrow q$ is the statement "p only if q," because both say that q must be true if p is true.

By the way, the statement $p \Longleftrightarrow \sim (\sim p)$ is always true; it is true if p is true and it is true if p is false. Such a statement is called a "tautology." A tautology gives no information in the sense that a tautology cannot imply that another statement is true or false, unless that statement is another tautology (always true) or is contradictory (always false).

B.) Proof by contradiction

Suppose we are trying to prove $p \Longrightarrow q$ for some statements p and q. A fairly common strategy is to argue by contradiction. This means that in addition to assuming p, we assume $\sim q$, the negation of q. After some logical steps, we reach some contradiction (i.e., we derive a result which we know is false). At this point we can conclude that q must be true, if p is true, because we have ruled out the possibility that $\sim q$ holds since $\sim q$ leads to a false statement. This approach is sometimes called "reducto ad absurdum," or reduction to the absurd. It is often used in arguments, e.g., "if what you say is true, then we would all be dead, but we're not all dead, so you must be wrong."

Most people regard proof by contradiction as reasonable and are willing to accept it in general. The precise formulation of proof by contradiction is the equivalence

$$p \Longrightarrow q \qquad \Longleftrightarrow \qquad \sim (p \wedge \sim q). \tag{2.3}$$

The right side of this equivalence, namely $\sim (p \wedge \sim q)$, says that if you assume p and the negation of q, you reach a contradiction (a statement that is false), which is exactly how proof by contradiction proceeds. To verbally justify the equivalence in (2.3), the right side can be stated as "it is not true that p is true and q is false;" in other words, when p is true, then q must be true, which is the same as $p \Longrightarrow q$. However, if this explanation is not enough for you, it can be put on a firm foundation using truth tables, as follows:

p	q	$p \Longrightarrow q$	$\sim q$	$p \wedge \sim q$	$\sim (p \wedge \sim q)$
T	T	T	F	F	T
T	F	F	T	T	F
F	T	T	F	F	T
F	F	T	T	F	T

We observe that the columns for $p \Longrightarrow q$ and $\sim (p \wedge \sim q)$ coincide. That is, in all cases, the statement $p \Longrightarrow q$ is true (or "holds") if and only if the statement $\sim (p \wedge \sim q)$ holds. Thus $p \Longrightarrow q$ and $\sim (p \wedge \sim q)$ are logically equivalent; if you prove one of them you have the other automatically. The proof by truth tables has the advantage of being a reliable, even automatic, computational process, which can be easily verified (even programmed). However, proof by truth tables has the disadvantage that the intuition behind the statements and their implications tends to get lost in the computations.

Sometimes proof by contradiction is necessary, and sometimes it is simpler or more convenient than directly proving an implication. However, it has the drawback that ultimately the argument leads to a contradiction, meaning that the intermediate steps are not true. One is writing a sequence of wrong statements to see that they are all wrong. It is more satisfying to write true statements. It is often possible to rewrite a proof by contradiction as a direct proof, which is usually more intuitive because the intermediate steps are actually true statements. So although sometimes proof by contradiction is necessary, it tends to be overused by students. We encourage students who are using proof by contradiction and getting confused in the process to try reformulating the proof as a direct proof. Often the reasoning become much more intuitive and transparent in direct form.

C.) Contrapositive

The "contrapositive" of the implication $p \Longrightarrow q$ is the statement $\sim q \Longrightarrow \sim p$. However, unlike the converse, the contrapositive is actually equivalent to the original implication. That is,

$$p \Longrightarrow q \qquad \Longleftrightarrow \qquad \sim q \Longrightarrow \sim p. \tag{2.4}$$

Here's a proof of (2.4) in the style of the proofs we will do in this course:

PROOF. To prove the equivalence, we must prove two implications: $(p \Longrightarrow q) \Longrightarrow (\sim q \Longrightarrow \sim p)$ and $(\sim q \Longrightarrow \sim p) \Longrightarrow (p \Longrightarrow q)$.

(i) $(p \Longrightarrow q) \Longrightarrow (\sim q \Longrightarrow \sim p)$: Assume $p \Longrightarrow q$. To show $\sim q \Longrightarrow \sim p$, we argue by contradiction. That is, we assume $\sim q$ and $\sim (\sim p) = p$. Since p holds, then from $p \Longrightarrow q$, we conclude q. But q contradicts our assumption $\sim q$. This contradiction shows that our assumption p is false, so we have $\sim p$, as required.

(ii) $(\sim q \Longrightarrow \sim p) \Longrightarrow (p \Longrightarrow q)$: Assume $\sim q \Longrightarrow \sim p$. Applying (i), which we have already proved, with p replaced by $\sim q$ and q replaced by $\sim p$, we get that $\sim (\sim p) \Longrightarrow \sim (\sim q)$, which is the same as $p \Longrightarrow q$. ■

One can also use truth tables, as follows:

p	q	$p \Longrightarrow q$	$\sim q$	$\sim p$	$\sim q \Longrightarrow \sim p$
T	T	T	F	F	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

Because the columns under $p \Longrightarrow q$ and $\sim q \Longrightarrow \sim p$ are identical, we conclude that these two statements are equivalent, i.e., (2.4).

D.) Exercise

Let's do a typical sort of homework problem regarding the logical concepts we have introduced.

Example 2.0.1 *Let p, q, and r be statements. Prove that*

$$p \Longrightarrow (q \vee r) \qquad \Longleftrightarrow \qquad (p \wedge \sim q) \Longrightarrow r.$$

PROOF. ($\Longrightarrow$) (Remark: this notation is shorthand for saying: we first prove the forward direction of the equivalence, namely $(p \Longrightarrow (q \vee r)) \Longrightarrow ((p \wedge \sim q) \Longrightarrow r)$.)

Assume $p \Longrightarrow (q \vee r)$. To show $(p \wedge \sim q) \Longrightarrow r$, assume $p \wedge \sim q$. Then p holds (by $p \wedge \sim q$), so from our assumption $p \Longrightarrow (q \vee r)$ we conclude $q \vee r$. But $\sim q$ is true (by $p \wedge \sim q$), so r must hold (from $q \vee r$, since q is false).

($\Longleftarrow$) (Remark: this notation is shorthand for saying that now we prove the converse direction of the equivalence, namely $((p \wedge \sim q) \Longrightarrow r) \Longrightarrow (p \Longrightarrow (q \vee r))$.)

Assume $(p \wedge \sim q) \Longrightarrow r$. To show $p \Longrightarrow (q \vee r)$, assume p. We consider the two possibilities for q:
(i) if q is true, then $q \vee r$ is true, as required;
(ii) if $\sim q$ is true, then since p is true, we have $p \wedge \sim q$. Then from $(p \wedge \sim q) \Longrightarrow r$ we deduce r, so $q \vee r$ is true, as required. ■

Of course there is a truth table proof of this exercise as well, but there are 8 different possibilities for the truth values of p, q, and r, which makes the truth table tedious to work out. The proof above is more intuitively clear.

Chapter 3

Sets

A.) Introduction to sets; subsets

Sets and set notation are used throughout mathematics. It would take us too far into mathematical logic to discuss the axioms of set theory, so we will deal with sets in an intuitive way. That will be sufficient for our purposes. A *set* is a collection of (mathematical) objects, which are called the *elements* or *members* of the set. If x is an element of the set A, we write $x \in A$. We write $\sim (x \in A)$, the negation of the statement that x is an element of A, in the shorter notation $x \notin A$, which is stated as "x is not in A."

Sometimes sets are described by listing the elements, such as

$$D = \{2, 46, -\pi\}.$$

We will use the standard notation $\mathbb{N}$ for the natural numbers:

$$\mathbb{N} = \{1, 2, 3, \ldots\},$$

and we will denote the real numbers by $\mathbb{R}$. Sometimes a set is described by the properties its members have; for example,

$$C = \{x \in \mathbb{R} : x \text{ is rational and } 0 < x < 1\},$$

which is read as the set of real numbers x such that x is a rational number and $0 < x < 1$. It is useful to define a set with no elements, called the *empty set*, which is denoted $\emptyset$.

If every member of a set A is also a member of B, we write $A \subseteq B$ or $A \subset B$ (read "A is contained in B" or "A is a subset of B"). Then $\sim (A \subseteq B)$, the negation of the statement $A \subseteq B$, is written $A \not\subseteq B$. The statement $A \subseteq B$ can be written as a "for all" statement in the form: "for all $x \in A$, we have $x \in B$." Therefore the statement $A \not\subseteq B$ is a "there exists" statement: $A \not\subseteq B$ is the same as "there exists $x \in A$ such that $x \notin B$."

This leads to the question: given a set A, is it true that $\emptyset \subseteq A$? We (mathematicians) hold the statement $\emptyset \subseteq A$ to be true. It may seem meaningless to say that every element of the empty set is an element of A, since there are no elements in the empty set. In fact, $\emptyset \subseteq A$ is a good example of a statement that is vacuously true, as described in Chapter 2. The truth of the statement $\emptyset \subseteq A$ can just be regarded as a mathematical or logical convention. Or, to look at the question another way, if it were not true that $\emptyset \subseteq A$, then by the last paragraph, there would have to be an element of the empty set which is not an element of A, which is impossible because there are no elements of the empty set.

Often a mathematical problem is formulated as: "Prove that $A \subseteq B$," for two given sets A and B. Question: How does one prove $A \subseteq B$? Answer: by taking an arbitrary $x \in A$ and showing that $x \in B$.

Example 3.0.1 *Let $A = \{x \in \mathbb{R} : x > 2\}$. Let $B = \{x \in \mathbb{R} : x^2 > 1\}$. Prove that $A \subseteq B$.*

Proof. Let $x \in A$. Then $x > 2$. Then $x^2 > 4$ (the justification for this step will be given when we study the order properties of $\mathbb{R}$, but let's assume it for now). Since $4 > 1$, we get $x^2 > 1$, so $x \in B$. Hence $A \subseteq B$. ■

The last sentence of the solution incorporates the idea above: we have shown that an arbitrary element of A is an element of B, so we have shown $A \subseteq B$. It is very important to note that we had to show that $x \in B$ for an arbitrary $x \in A$. A common mistake that students make is to pick a particular element of A, say $x = 6$, and note that $x^2 = 36 > 1$, so $x \in B$. This observation does not show that $A \subseteq B$ because there might in principle be other elements of A (besides $x = 6$) which are not in B. To show that $A \subseteq B$ we have to show that *all* elements of A also belong to B. But by taking an arbitrary element x of A (hence using only the property that $x \in A$), and showing $x \in B$, we cover all possible $x \in A$, and hence we show that $A \subseteq B$.

Similarly, how does one show that one set is not a subset of another? For example, for the A and B just considered, how do we show $B \not\subseteq A$? For $B \not\subseteq A$ to be true, there must be an element of B which is not an element of A. The easiest (but not necessarily the only) way to prove that such an element exists is to find such an element and show that it belongs to B but not to A. In particular, $x = 1.5$ satisfies $x \in B$, since $x^2 = 2.25 > 1$, but $x \notin A$, since x is not greater than 2. This example by itself shows that $B \not\subseteq A$. Notice that there are some elements of B, such as $x = 5$, which are elements of A. But, as we noted above, showing that some element of B belongs to A does not imply that $B \subseteq A$, and in this case it is not true that $B \subseteq A$.

If these examples seem familiar, it is because we considered the same mathematical point in the previous section. There we let p be the assertion that $x > 2$, and q the assertion that $x^2 > 1$. We then considered the possible implicatins $p \Longrightarrow q$ (which we saw to be true) and $q \Longrightarrow p$ (which we saw to be false). The statement $p \Longrightarrow q$ is the same as $A \subseteq B$, and $q \Longrightarrow p$ is the same as $B \subseteq A$. From these examples we see that set notation provides an alternate way of formulating mathematical statements. Why is such a reformulation useful? Sometimes set notation is just a way to shorten statements and avoid repetition. If we are going to be writing about real numbers x satisfying $x^2 > 1$ extensively, it is tedious to write "let x be a real number such that $x^2 > 1$" repeatedly. So we define the set B as above and just write: "let $x \in B$." However, convenience is not the only reason to use set notation. Once we establish some general principles about sets, we will be able to apply them without having to repeat the reasoning in each circumstance.

Two sets A and B are equal (written $A = B$), which means that they have exactly the same elements, if and only if every element of A is an element of B and every element of B is an element of A. More succinctly, $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$. This observation is a triviality, but we often use this principle, by breaking the proof of $A = B$ into two parts: $A \subseteq B$ and $B \subseteq A$.

Another triviality is the property that if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$. As obvious as this may be, it is important to have a proof, which goes as follows. To show $A \subseteq C$, let $x \in A$. We must show that $x \in C$. Since $A \subseteq B$ and $x \in A$, it follows that $x \in B$. Then because $B \subseteq C$ and $x \in B$, we have $x \in C$, as required.

B.) Unions, Intersections, Differences, Complements, and Products of Sets

Definition 3.0.2 *Suppose A and B are sets. Then*

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

We call $A \cup B$ the union *of A and B. Also,*

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

We call $A \cap B$ the intersection *of A and B.*

In words, $A \cup B$ consists of all elements that belong to either A or B or both (because "or" is used inclusively, as discussed in Section 2). Also, $A \cap B$ is the set of elements that belong to both A and B. If there are no such elements, then $A \cap B = \emptyset$, in which case we say A and B are *disjoint*. One of the reasons it is useful to define the empty set is so that $A \cap B$ is always defined, for any sets A and B.

We can take the union or intersection of any number of sets, but for that we need some more notation. Suppose that for some set Λ and for each element $\lambda \in \Lambda$, there is a set A_λ. We define the union of all set A_λ as

$$\cup_{\lambda \in \Lambda} A_\lambda = \{x : x \in A_\lambda \text{ for some } \lambda \in \Lambda\} = \{x : \text{ there exists } \lambda \in \Lambda \text{ such that } x \in A_\lambda\}.$$

We also define the intersection of all sets A_λ to be

$$\cap_{\lambda \in \Lambda} A_\lambda = \{x : x \in A_\lambda \text{ for all } \lambda \in \Lambda\}.$$

We call Λ the *index* set; it is just needed to give names to the sets in the union. The union $A \cup B$ of two sets can be written in this form by renaming the sets A_1 and A_2 and letting $\Lambda = \{1, 2\}$; then $A \cup B = \cup_{i \in \{1,2\}} A_i$. When the index set is the natural numbers $\mathbb{N}$, we often write $\cup_{i=1}^{\infty} A_i$ instead of $\cup_{i \in \mathbb{N}} A_i$ and $\cap_{i=1}^{\infty} A_i$ instead of $\cap_{i \in \mathbb{N}} A_i$. Similarly, when we have finitely many sets $A_1, A_2, \ldots, A_n$, we write $\cup_{i=1}^{n} A_i$ instead of $\cup_{i \in \{1,2,\ldots,n\}} A_i$ and $\cap_{i=1}^{n} A_i$ instead of $\cap_{i \in \{1,2,\ldots,n\}} A_i$

We give full details in the solution to the following example not because they are difficult, but because we want to show how to write out, in a logically correct proof, things that may seem obvious.

Example 3.0.3 *For $n \in \mathbb{N}$, let $A_n = [\frac{1}{n}, 1] \equiv \{x \in \mathbb{R} : \frac{1}{n} \leq x \leq 1\}$. Find $\cup_{n=1}^{\infty} A_n$ and $\cap_{n=1}^{\infty} A_n$.*

Solution (i): $\cup_{n=1}^{\infty} A_n = (0, 1] \equiv \{x \in \mathbb{R} : 0 < x \leq 1\}$. Proof: we first show that $\cup_{n=1}^{\infty} A_n \subseteq (0, 1]$. Suppose $x \in \cup_{n=1}^{\infty} A_n$. By definition of union, then, there exists $n \in \mathbb{N}$ such that $x \in A_n = [\frac{1}{n}, 1]$ Hence $0 < \frac{1}{n} \leq x \leq 1$, so $x \in (0, 1]$. (For now we are assuming basic facts about the ordering of the real numbers, but these will be elaborated later). Second we show that $(0, 1] \subseteq \cup_{n=1}^{\infty} A_n$. Suppose $x \in (0, 1]$. Since $x > 0$, there exists $n_0 \in \mathbb{N}$ such that $\frac{1}{n_0} < x$ (this fact follows from the Archimedean property of the real numbers, which we will discuss later.) Since $x \leq 1$, we have $x \in [\frac{1}{n_0}, 1] = A_{n_0}$. Hence $x \in \cup_{n=1}^{\infty} A_n$. Since we have shown that $\cup_{n=1}^{\infty} A_n \subseteq (0, 1]$ and $(0, 1] \subseteq \cup_{n=1}^{\infty} A_n$, we have shown that $\cup_{n=1}^{\infty} A_n = (0, 1]$.

(ii) $\cap_{n=1}^{\infty} A_n = \{1\}$. First, if $x \in \cap_{n=1}^{\infty} A_n$, then $x \in A_1 = [1, 1] = \{1\}$, so $\cap_{n=1}^{\infty} A_n \subseteq \{1\}$. Second, if $x \in \{1\}$, then $x = 1 \in [\frac{1}{n}, 1] = A_n$ for each $n \in \mathbb{N}$, so $x \in \cap_{n=1}^{\infty} A_n$, which shows that $\{1\} \subseteq \cap_{n=1}^{\infty} A_n$. The two inclusions ($\cap_{n=1}^{\infty} A_n \subseteq \{1\}$ and $\{1\} \subseteq \cap_{n=1}^{\infty} A_n$) imply that $\cap_{n=1}^{\infty} A_n = \{1\}$.

Notice that the sets A_n are increasing: $A_n \subseteq A_{n+1}$ for each $n \in \mathbb{N}$. So for part (i) it is tempting to "let $n \to \infty$" and conclude that $\cup_{n=1}^{\infty} A_n =$ "A_∞" where $A_\infty = [\lim_{n \to \infty} \frac{1}{n}, 1] = [0, 1]$. But that reasoning is wrong because the element 0 does not belong to any A_n and hence does not belong to $\cup_{n=1}^{\infty} A_n$, by the definition of union.

Definition 3.0.4 *For sets A and B, we define the* difference *of A and B to be*

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

If all sets that we are considering are assumed to be subsets of some given set X (e.g., if we are considering only sets of real numbers, so that $X = \mathbb{R}$), we often use the notation

$$A^c = X \setminus A.$$

Here "c" stands for "complement" and we say A^c is the "complement" of A (in X, but this part is usually understood and not stated). In this case, we can write

$$A \setminus B = A \cap B^c.$$

Here the "order of operations" in set notation is that complements are taken first, so that, for example $A \cap B^c$ means $A \cap (B^c)$, not $(A \cap B)^c$. Here are some simple facts, which are left to the reader to prove: for any set A, we have $(A^c)^c = A$, $A \cup A^c = X$, and $A \cap A^c = \emptyset$.

One more way to form new sets given sets A and B is to form their *product*, as follows.

Definition 3.0.5 *Given two sets A and B, an ordered pair is an element of the form (a, b), where $a \in A$ and $b \in B$. For $a_1, a_2 \in A$ and $b_1, b_2 \in B$, then $(a_1, b_1) = (a_2, b_2)$ if and only if $a_1 = a_2$ and $b_1 = b_2$. The product of A and B, denoted $A \times B$, is the set of all such ordered pairs; i.e.,*

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

The ordered pair (a, b) is not the same as the set $\{a, b\}$; for example if $A = B$, and $a_1, a_2 \in A$ with $a_1 \neq a_2$, then $\{a_1, a_2\} = \{a_2, a_1\}$ whereas $(a_1, a_2) \neq (a_2, a_1)$. That distinction is the reason for the word "ordered" in "ordered pair."

C.) De Morgan's Laws

De Morgan's laws are two frequently used principles about sets. We assume that all sets considered are subsets of a given set X, and complements are taken with respect to X, in the sense described above (so $A^c = X \setminus A$, etc.). For two sets A, B, De Morgan's laws state:

$$(A \cup B)^c = A^c \cap B^c \tag{3.1}$$

and

$$(A \cap B)^c = A^c \cup B^c. \tag{3.2}$$

There is a more general form of De Morgan's laws, which apply to arbitrary unions and intersections, as follows. Let Λ be an index set, so that for each $\lambda \in \Lambda$, there is a set $A_\lambda \subseteq X$. Then

$$(\cup_{\lambda\in\Lambda} A_\lambda)^c = \cap_{\lambda\in\Lambda} A_\lambda^c \tag{3.3}$$

and

$$(\cap_{\lambda\in\Lambda} A_\lambda)^c = \cup_{\lambda\in\Lambda} A_\lambda^c. \tag{3.4}$$

These relations are often remembered as "the complement of the union is the intersection of the complements" and "the complement of the intersection is the union of the complements." The statements (3.1) and (3.2) follow from (3.3) and (3.4) by using an index set with two elements. We prove (3.1) just to give an understanding of how these proofs work, and then leave (3.3) and (3.4) as exercises.

Proof. To prove (3.1), we first show that $(A\cup B)^c \subseteq A^c \cap B^c$. Let $x \in (A\cup B)^c$. This statement means that $x \notin A\cup B$. By definition of union, it follows that $x \notin A$ and $x \notin B$. Since $x \notin A$, we have $x \in A^c$. Since $x \notin B$, we have $x \in B^c$. Since $x \in A^c$ and $x \in B^c$, we have shown $x \in A^c \cap B^c$.

Now we show that $A^c \cap B^c \subseteq (A\cup B)^c$. Let $x \in A^c \cap B^c$. Then $x \in A^c$ and $x \in B^c$. Since $x \in A^c$, we have $x \notin A$. Since $x \in B^c$, we have $x \notin B$. Since $x \notin A$ and $x \notin B$, we have $x \notin A\cup B$. Hence $x \in (A\cup B)^c$. Thus $A^c \cap B^c \subseteq (A\cup B)^c$.

The two inclusions ($(A\cup B)^c \subseteq A^c \cap B^c$ and $A^c \cap B^c \subseteq (A\cup B)^c$) show that $(A\cup B)^c = A^c \cap B^c$. ■

In the future we will write proofs with less detail, leaving a bit more to reader's understanding. For example, the last sentence of the last proof might be omitted, with the point of proving both inclusions being regarded as obvious. At this stage we are being a bit pedantic just to exhibit all steps of the logic.

Chapter 4

Functions

A.) Definition of a Function

Analysis is largely the study of functions, often functions that have reasonable properties, such as continuity or differentiability. We begin with the general concept of a function. Functions are often described (or defined) as "rules of assignment," as follows. Given sets X (the "domain" of the function) and Y (the "co-domain" of the function), a function $f : X \to Y$ (or just a function f if X and Y are understood) is a rule that assigns, to each $x \in X$, one element, called $f(x)$, of Y.

The exact meaning of the words "rule" and "assigns" may not be entirely clear in that definition, so there is a more rigorous definition. Formally a function is a subset f of $X \times Y$ with two properties: (i) for each $x \in X$, there exists a pair $(x, y) \in f$, and (ii) if $(x, y_1) \in f$ and $(x, y_2) \in f$, then $y_1 = y_2$. Together these two conditions mean that for each $x \in X$, there is exactly one $y \in Y$ such that (x, y) is in the subset of ordered pairs called f, and we then define $y = f(x)$. This formality is just needed to make the definition precise; for all practical purposes the intuition associated with the "rule of assignment" description of a function is perfectly fine.

The two key points about a function is that there is always an element $f(x) \in Y$ for every $x \in X$, and there is only one such element $f(x)$, for each $x \in X$. Intuitively, a function f is just a way of associating a point $f(x)$ in Y to each point x in X.

Here is a simple example.

Example 4.0.1 *Let $X = \{a, b, c, d\}$ and $Y = \{\alpha, \beta, \gamma\}$. Define a function $f : X \to Y$ by defining*

$$f(a) = \alpha, f(b) = \beta, f(c) = \beta, \text{ and } f(d) = \beta.$$

Since each point of X is assigned a single point of Y, this assignment defines a function.

We can picture functions by putting the points of the domain on the left, the points of the co-domain on the right, and drawing an arrow from each point x in the domain to the point $f(x)$ in Y. We can have several arrows ending at the same point (b, c and d in the example above), and we can have points of Y which are not at the end of any arrow (γ in the example above). Two properties are required for a function: (i) we can not have two or more arrows starting from the same point of X, and (ii) we can not have a point of X which does not have any arrow starting from it. We sometimes call a function a "mapping," because we think of the arrows as mapping the points of X to points of Y. This viewpoint can help with visualizing functions.

B.) Images and Inverse Images

Definition 4.0.2 *Suppose $f : X \to Y$ is a function. For any subset A of X, the image (sometimes the "forward image") of A is*

$$f(A) = \{f(a) : a \in A\}.$$

In the example above, $f(\{a, b\}) = \{\alpha, \beta\}$ and also $f(\{a, b, c, d\}) = \{\alpha, \beta\}$. We call $f(X)$, the image of the entire domain, the *range* of f. As in the example above, the range (in this case $\{\alpha, \beta\}$) may not be all of the co-domain (in this case $\{\alpha, \beta, \gamma\}$).

It is worthwhile to see how the images of sets behave under unions and intersections. Suppose X, Y are sets, $f : X \to Y$ is a function, and $A_\lambda \subseteq X$ for each $\lambda \in \Lambda$, where Λ is some index set. Then

$$f(\cup_{\lambda\in\Lambda} A_\lambda) = \cup_{\lambda\in\Lambda} f(A_\lambda) \tag{4.1}$$

and

$$f(\cap_{\lambda\in\Lambda} A_\lambda) \subseteq \cap_{\lambda\in\Lambda} f(A_\lambda). \tag{4.2}$$

We note that there are examples where the sets in (4.2) are not equal; containment is the most that is true in general (constructing such an example will be part of an exercise for the student).

We will prove one direction of (4.1) and leave the rest as exercise. Let us prove that $f(\cup_{\lambda\in\Lambda} A_\lambda) \subseteq \cup_{\lambda\in\Lambda} f(A_\lambda)$.

PROOF. To show containment, we start with a general $y \in f(\cup_{\lambda\in\Lambda} A_\lambda)$ (warning: don't start with $x \in \cup_{\lambda\in\Lambda} A_\lambda$; that is confusing because we want to start with a general element of the left side). Then by definition of the image, there exists $x \in \cup_{\lambda\in\Lambda} A_\lambda$ such that $f(x) = y$. Since $x \in \cup_{\lambda\in\Lambda} A_\lambda$, there exists $\lambda_0 \in \Lambda$ such that $x \in A_{\lambda_0}$, by definition of union (we use the notation λ_0 for the particular $\lambda \in \Lambda$ just to distinguish it from the general λ). Then $y = f(x) \in f(A_{\lambda_0})$. Hence $y \in \cup_{\lambda\in\Lambda} f(A_\lambda)$. Thus we have shown that $f(\cup_{\lambda\in\Lambda} A_\lambda) \subseteq \cup_{\lambda\in\Lambda} f(A_\lambda)$. ■

Definition 4.0.3 *For a function $f : X \to Y$ and a subset B of Y, the "inverse image" of B is*

$$f^{-1}(B) = \{x \in X : f(x) \in B\}.$$

That is,, $f^{-1}(B)$ is the set of all points in X which are mapped into B by f. By definition, $x \in f^{-1}(B)$ if and only if $f(x) \in B$. In the example above, $f^{-1}(\{\beta\}) = \{b, c, d\}$ whereas $f^{-1}(\{\gamma\}) = \emptyset$. Notice that $f^{-1}(B)$ is defined to be a set.

To see how inverse images behave under unions and intersections, suppose $f : X \to Y$ is a function, and $B_\lambda \subseteq X$ for each $\lambda \in \Lambda$, where Λ is some index set. Then

$$f^{-1}(\cup_{\lambda\in\Lambda} B_\lambda) = \cup_{\lambda\in\Lambda} f^{-1}(B_\lambda) \tag{4.3}$$

and

$$f^{-1}(\cap_{\lambda\in\Lambda} B_\lambda) = \cap_{\lambda\in\Lambda} f^{-1}(B_\lambda). \tag{4.4}$$

To give a feeling for how the proofs go, we will prove that $f^{-1}(\cup_{\lambda\in\Lambda} B_\lambda) \subseteq \cup_{\lambda\in\Lambda} f^{-1}(B_\lambda)$, leaving the other direction of (4.3) and all of (4.4) as exercises.

PROOF. To prove $f^{-1}(\cup_{\lambda\in\Lambda} B_\lambda) \subseteq \cup_{\lambda\in\Lambda} f^{-1}(B_\lambda)$, suppose $x \in f^{-1}(\cup_{\lambda\in\Lambda} B_\lambda)$. Then $f(x) \in \cup_{\lambda\in\Lambda} B_\lambda$ (by the definition of inverse images). Then there exists $\lambda_0 \in \Lambda$ such that $f(x) \in B_{\lambda_0}$ (by the definition of unions). Hence $x \in f^{-1}(B_{\lambda_0})$. Therefore $x \in \cup_{\lambda\in\Lambda} f^{-1}(B_\lambda)$. ■

C.) $1-1$ Functions, Onto Functions

We noted that for a function, it is possible that different points of the domain are mapped to the same point of the co-domain. Also, it may be that there is some point of the co-domain that is not the image of any point in the domain. However, for some functions one or the other of these things may not happen. Such functions then have special properties which are worth studying.

Definition 4.0.4 *A function $f : X \to Y$ is $1-1$ (read "one to one") or injective if, for $x_1, x_2 \in X$, we have $f(x_1) \neq f(x_2)$ if $x_1 \neq x_2$.*

In other words, f is $1-1$ if different points of the domain always get mapped to different points of the range. We can write the $1-1$ condition as the implication $x_1 \neq x_2 \Longrightarrow f(x_1) \neq f(x_2)$. In our visualization of a function using as mappings using arrows starting in X and ending in Y, a function is $1-1$ if no two arrows (i.e., arrows starting at different points) end at the same point of Y.

Although this definition is the most intuitive way to think of $1-1$ functions, it is awkward to apply to actually demonstrate that a given function is $1-1$. Such a proof usually goes something like:

"Suppose $x_1 \neq x_2$. We want to show that $f(x_1) \neq f(x_2)$. By way of contradiction, suppose $f(x_1) = f(x_2)$. Then (after some computation, depending on what f is) we see that $x_1 = x_2$. But this contradicts our assumption that $x_1 \neq x_2$. Therefore our assumption $f(x_1) = f(x_2)$ must be wrong, so we have $f(x_1) \neq f(x_2)$, which is what we wanted to show."

Although this reasoning is correct, it is somewhat convoluted. It can also be confusing. We started by assuming $x_1 \neq x_2$, but then we deduced that $x_1 = x_2$. So is $x_1 = x_2$ or not? That is the problem with proofs by contradiction. It is hard to keep track of what is true or false as we go along. This difficulty can be avoided if we reformulate the definition of a $1-1$ function using the contrapositive (recall that the contrapositive of $p \Longrightarrow q$ is $\sim q \Longrightarrow \sim p$, and that a statement and its contrapositive are logically equivalent: see (2.4)). The main point in the definition that f is $1-1$ is that $x_1 \neq x_2 \Longrightarrow f(x_1) \neq f(x_2)$. The contrapositive of $x_1 \neq x_2 \Longrightarrow f(x_1) \neq f(x_2)$ is $f(x_1) = f(x_2) \Longrightarrow x_1 = x_2$. So we can reformulate the definition of f being $1-1$ as:

a function $f : X \to Y$ is $1-1$ if, for $x_1, x_2 \in X$, we have $f(x_1) = f(x_2) \Longrightarrow x_1 = x_2$.

This formulation may seem odd ("why do we call x_1 and x_2 by different names if they are going to turn out to be the same?" - the answer is that we don't know they are the same at the start, so we have to use different names, and our goal is to show that they are the same). However, this formulation is almost always the most convenient in actually proving a given function is $1-1$; we know that we can start by assuming $f(x_1) = f(x_2)$, which gives us something concrete to work with, and, using the properties of f, we must show $x_1 = x_2$.

Example 4.0.5 *Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = 3x + 4$. Prove that f is $1-1$.*

PROOF. Let $x_1, x_2 \in \mathbb{R}$ and suppose $f(x_1) = f(x_2)$; that is, $3x_1 + 4 = 3x_2 + 4$. Subtracting 4 from both sides (we will discuss steps like this later) gives $3x_1 = 3x_2$. Dividing both sides by 3 gives $x_1 = x_2$. Hence f is $1-1$. ■

To show that a function $f : X \to Y$ is not $1-1$, we need to show that there exist x_1, x_2 in X with $x_1 \neq x_2$ but $f(x_1) = f(x_2)$. For example, the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is not $1-1$; to prove f is not $1-1$ it is enough to give an example, such as $x_1 = -2, x_2 = 2$, and note that $f(x_1) = 4 = f(x_2)$.

Now we consider another special property that a function can satisfy.

Definition 4.0.6 *A function $f : X \to Y$ is onto or surjective if, for all $y \in Y$, there exists at least one $x \in X$ such that $f(x) = y$.*

In our arrow visualization of a function, a function is onto if every point in the co-domain is at the end of at least one arrow starting in the domain.

As an example, we can show that the function $f : \mathbb{R} \to \mathbb{R}$ defined above by $f(x) = 3x + 4$ is onto by taking an arbitrary $y \in \mathbb{R}$, and observing that for $x = \frac{1}{3}(y - 4)$, which is always a real number, we have $f(x) = 3x + 4 = 3 \cdot \frac{1}{3}(y - 4) + 4 = y$. To show that a function is not onto, we just have to demonstrate that there is a point y in the co-domain which is not the image of any x under f. For example, $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is not onto because for $y = -1$, there is no real number x such that $x^2 = f(x) = -1$ (as we will see when we discuss the properties of the real numbers).

D.) Composition of Functions, Bijections, and Inverse Functions

Definition 4.0.7 *Given sets X, Y, and Z, and functions $f : X \to Y$ and $g : Y \to Z$, the composition $g \circ f : X \to Z$ is the function defined by $g \circ f(x) = g(f(x))$.*

We have to check that this makes sense: since $x \in X$, we have $f(x) \in Y$ (since $f : X \to Y$), and hence $g(y) \in Z$ (since $g : Y \to Z$). Moreover $g \circ f$ is a function, because, for each $x \in X$, there is one and only

one value $f(x)$ assigned to x by f, and hence there is one and only one value $g(f(x))$ assigned to $f(x)$ by g, with the result that there is one and only one value $g(f(x))$ assigned to x by $g \circ f$. Our arrow visualization of functions is useful here; if you think of the arrow taking x to $f(x)$ as a trip, we then follow with the trip g which takes $f(x)$ to $g(f(x))$. The composition taking x to $g(f(x))$ is just the concatenation of the two trips, thought of as a single trip.

Notice that for $f : X \to Y$ and $g : Y \to Z$, the composition $f \circ g(y) = f(g(y))$ may not be defined, because $g(y)$ is an element of Z, which may not be in the domain X of f. So care has to be used with the notation of composition; just because you write down a composition, it doesn't mean that the composition is actually defined. Also, even when both $f \circ g$ and $g \circ f$ are defined, they are almost never the same. For example, for $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 3x + 4$ and $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^2$, we have $g \circ f(x) = g(f(x)) = g(3x + 4) = (3x + 4)^2$ whereas $f(g(x)) = f(x^2) = 3x^2 + 4$. In particular, $g \circ f(0) = 16$ and $f \circ g(0) = 4$.

It is useful to note the following.

Proposition 4.0.8 *Suppose $f : X \to Y$ and $g : Y \to Z$ are functions.*

(i) If f and g are $1-1$, then $g \circ f$ is $1-1$;
and
(ii) if f and g are onto, then $g \circ f$ is onto.

We prove (i) and leave (ii) as an exercise. The proof of (i) is very quick using the contrapositive formulation of $1-1$ definition as described above, as follows.

PROOF. Suppose $x_1, x_1 \in X$ and $g \circ f(x_1) = g \circ f(x_2)$. That is, $g(f(x_1)) = g(f(x_2))$. Since g is $1-1$, it follows that $f(x_1) = f(x_2)$. Since f is $1-1$, then $x_1 = x_2$, which completes the proof. ■

Definition 4.0.9 *A function $f : X \to Y$ is called a bijection if f is $1-1$ and onto. A bijection is often called a "$1-1$ correspondence."*

As an example, let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 3x + 4$. We have already shown that f is $1-1$ and onto, so f is a bijection.

Note that if $f : X \to Y$ and $g : Y \to Z$ are bijections, then $g \circ f : X \to Z$ is a bijection. This fact follows because we have already observed in Proposition 4.0.8 that $g \circ f$ is both $1-1$ and onto.

Suppose $f : X \to Y$ is a bijection. Then we can define a function, called f^{-1}, from Y to X, as follows: given $y \in Y$, there exists a unique $x \in X$ such that $f(x) = y$ (one such x exists because f is onto, and this x is unique since f is $1-1$). Define $f^{-1}(y) = x$. This defines $f^{-1}(y)$ as an element of X for each $y \in Y$, so $f^{-1} : Y \to X$ is a function. We observe that f^{-1} satisfies the two properties:

$$f^{-1}(f(x)) = x \text{ for all } x \in X, \text{ and } f(f^{-1}(y)) = y \text{ for all } y \in Y. \tag{4.5}$$

To see the first property, suppose $x \in X$. Let $y = f(x)$. Then by definition of the inverse function, $f^{-1}(f(x)) = f^{-1}(y) = x$. For the second property, suppose $y \in Y$. Since f is onto, there exists $x \in X$ such that $f(x) = y$. Then $x = f^{-1}(y)$ by definition of the inverse. Hence $f(f^{-1}(y)) = f(x) = y$.

Moreover, the properties (4.5) characterize f^{-1}, in the following sense.

Proposition 4.0.10 *Suppose $f : X \to Y$ is a function, and there exists a function $g : Y \to X$ such that $g \circ f(x) = x$ for all $x \in X$ and $f \circ g(y) = y$ for all $y \in Y$. Then f is a bijection and $g = f^{-1}$.*

PROOF. To show that f is $1-1$, suppose $f(x_1) = f(x_2)$. Then

$$x_1 = g \circ f(x_1) = g(f(x_1)) = g(f(x_2)) = g \circ f(x_2) = x_2,$$

hence f is $1-1$ (the equality $g(f(x_1)) = g(f(x_2))$ holds just because $f(x_1) = f(x_2)$). To show that f is onto, suppose $y \in Y$. Then $f(g(y)) = f \circ g(y) = y$, so f is onto. Therefore f is a bijection. Therefore an inverse function f^{-1} exists. To show that $g = f^{-1}$, let $y \in Y$. Then

$$f(g(y)) = f \circ g(y) = y = f \circ f^{-1}(y) = f(f^{-1}(y)).$$

Since $f(g(y)) = f(f^{-1}(y))$ and f is $1-1$, we conclude that $g(y) = f^{-1}(y)$. Since $y \in Y$ is arbitrary, we conclude that $g = f^{-1}$. ■

Therefore a function f is a bijection if and only if it has an inverse f^{-1}. For that reason, functions which are bijections are sometimes called *invertible*. We warn the student that this is another example where just because one can write down the notation f^{-1}, it doesn't mean that a function f^{-1} actually exists. Before writing f^{-1}, one must show that f is a bijection, which then guarantees that there is an inverse function f^{-1}.

Note that if $f : X \to Y$ is a bijection with inverse f^{-1}, we can apply the previous proposition with f replaced by f^{-1} and g replaced with f (with X and Y interchanged), because we already know that $f \circ f^{-1}(y) = y$ for all $y \in Y$, and $f^{-1}(f(x)) = x$ for all $x \in X$. We can conclude, then, that f^{-1} is a bijection with inverse f. In particular, $(f^{-1})^{-1} = f$.

We warn the student about a possible confusion relating to the notation f^{-1}, which has been used above in reference to the inverse image of a set, and just recently to the inverse function, when it exists. Some students think that using the notation f^{-1} implies that f is invertible, and try to use properties of inverse functions to answer questions about inverse images of sets. Recall from Definition 4.0.3 that for $f : X \to Y$, the inverse image $f^{-1}(B) = \{x \in X : f(x) \in B\}$ is defined, whether or not f is $1-1$ or onto. The inverse image of a set is always defined, and is defined to be a set, not an element of X. Even if the inverse image of some set B has only one element, call it x_0, then $f^{-1}(B)$ is the set $\{x_0\}$, not the element x_0. It is tempting to identify the set $\{x_0\}$ with the element x_0, but they are not the same; one is a set, the other is an element. (This distinction may seem pedantic, but it is important because sets and elements are different types of mathematical objects and must be distinguished. For example, the set whose only element is the empty set, that is, $\{\emptyset\}$, is not the same as the empty set $\emptyset$, because the set $\{\emptyset\}$ is not empty - it has one element, namely $\emptyset$.) Given an element $y \in Y$, the quantity $f^{-1}(y)$ is only defined if f is invertible, but $f^{-1}(\{y\})$ is defined for any function f. If f is invertible, then $f^{-1}(\{y\}) = \{f^{-1}(y)\}$; that is, the element $f^{-1}(y)$ is the only element in the set $f^{-1}(\{y\})$.

At this point we could go directly into the study of the "cardinality" of sets, in particular to the discussion of countable and uncountable sets, but we prefer to first study the natural numbers $\mathbb{N}$, the integers $\mathbb{Z}$, the rational numbers $\mathbb{Q}$, and the real numbers $\mathbb{R}$. Then we can study cardinality with these examples in mind.

Chapter 5

The Natural Numbers $\mathbb{N}$ and Induction

Our goal in the next three sections is to establish the standard algebraic properties of the real numbers $\mathbb{R}$. For example, a standard property that we learn in elementary school is that $-(-x) = x$, or that $(-a)(-b) = ab$ (if a and b are positive, this principle is often drilled into students as "a negative times a negative is a positive"). Why is that true? At least one scholar gave up on mathematics because no one could explain to him why this principle was true, claiming that math is not as logical as it claims to be. So, how does one prove that $(-a)(-b) = ab$? What does it even mean to prove such a statement? Prove based on what?

To make basic algebra logically sound, we need to establish axioms for $\mathbb{R}$ and then derive all further properties from the axioms. Wwe would like the axioms to be as elementary and non-controversial as possible. However, it turns out that finding axiomatic properties characterizing $\mathbb{R}$ is not so easy, and we need to work up to that. We start with the natural numbers $\mathbb{N}$. The basic principles of $\mathbb{N}$ are that (i) there is a first natural number, which we call 1, (ii) every natural number n has a successor, which we call $n+1$, (iii) different natural numbers have different successors, and (iv) $\mathbb{N}$ is as small as it can be with these properties. To be more precise, we elaborate the axioms for the set N as follows.

Axioms for $\mathbb{N}$:

N1: $1 \in \mathbb{N}$,
N2: Each element $n \in \mathbb{N}$ has a *successor*, which we denote by $n + 1$,
N3: 1 is not the successor of any $n \in \mathbb{N}$,
N4: If $n, m \in \mathbb{N}$ and n and m have the same successor, then $n = m$,
N5: Suppose S is a set such that: (a) $S \subseteq \mathbb{N}$, (b) $1 \in S$, and (c) $n \in S \Longrightarrow n + 1 \in S$. Then $S = \mathbb{N}$.

Certainly our intuition tells us that $\mathbb{N}$ satisfies N1-N5. One might wonder why N5 is needed. N5 guarantees that $\mathbb{N}$ is the minimal set that satisfies N1-N4. For example, if

$$M = \left\{ \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, 3, \dots \right\},$$

then M satisfies N1-N4 but not N5. The assumptions N1-N5 guarantee that $\mathbb{N}$ includes the elements $1, 2, 3, \dots$, and nothing more.

N5 is also used to establish a key method of proof called the *Principle of Induction*, as follows.

Lemma 5.0.1 *(Induction Principle) Suppose that for each $n \in \mathbb{N}$, P_n is a statement (depending on n). Suppose (i) P_1 is true, and (ii) $P_n \Longrightarrow P_{n+1}$. Then P_n is true for all $n \in \mathbb{N}$.*

Proof. Let $S = \{n \in \mathbb{N} : P_n \text{ is true } \}$. Then (a) $S \subseteq \mathbb{N}$, (b) $1 \in S$ by (i), and (c) $n \in S \Longrightarrow n + 1 \in S$ by (ii). Hence by N5, $S = \mathbb{N}$, which means (by definition of S) that P_n is true for all $n \in \mathbb{N}$. ■

Step (i) is sometimes referred to as the *base step* of the induction, whereas (ii) is often called the *inductive step.*

For the following examples of proof by induction, we will assume the basic rules of arithmetic (commutativity and assoiciativity of addition and multiplication, etc.). These properties will be part of the axioms for $\mathbb{R}$ when we reach the point of defining $\mathbb{R}$, and basic algebra priniciples will follow from these axioms.

Example 5.0.2 *Prove that*

$$1+2+3+\cdots+n=\frac{n(n+1)}{2},$$

for each $n\in\mathbb{N}$.

PROOF. Let P_n be the statement that $1+2+3+\cdots+n=\frac{n(n+1)}{2}$. Then:
(i): P_1 is true: P_1 states that $1=\frac{1\cdot 2}{2}$, which is true,
and
(ii) $P_n\Longrightarrow P_{n+1}$: Suppose P_n is true, for some $n\in\mathbb{N}$, which means that $1+2+3+\cdots+n=\frac{n(n+1)}{2}$. We add $n+1$ to both sides of the equation, to obtain

$$1+2+3+\cdots+n+(n+1)=\frac{n(n+1)}{2}+(n+1)=\left(\frac{n}{2}+1\right)(n+1)=\frac{n+2}{2}\cdot(n+1)=\frac{(n+1)(n+2)}{2}.$$

Hence $1+2+3+\cdots+(n+1)=\frac{(n+1)(n+2)}{2}$, which is the statement P_{n+1}.
By (i), (ii), and the induction principle, P_n holds for all $n\in\mathbb{N}$, as required. ■

If one feels that the notation $\cdots$ in the expression $1+2+3+\cdots+n$ is not quite clear or rigorous enough, one can use the more precise notation $\sum_{j=1}^{n} j$ instead. At this stage we are comfortable with the more intuitive notation $1+2+3+\cdots+n$.

Example 5.0.3 *Suppose* $x\in\mathbb{R}$ *and* $x\neq 1$. *Prove that*

$$1+x+x^2+x^3+\cdots+x^n=\frac{1-x^{n+1}}{1-x},$$

for each $n\in\mathbb{N}$.

PROOF. For $x\in\mathbb{R}$ with $x\neq 1$, let P_n be the statement that $1+x+x^2+x^3+\cdots+x^n=\frac{1-x^{n+1}}{1-x}$. Then
(i): P_1 is true: P_1 states that $1+x=\frac{1-x^2}{1-x}$, which is true because $1-x^2=(1+x)(1-x)$,
and
(ii) $P_n\Longrightarrow P_{n+1}$: Suppose P_n is true, for some $n\in\mathbb{N}$, which means that $1+x+x^2+x^3+\cdots+x^n=\frac{1-x^{n+1}}{1-x}$. Then

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

$$=\frac{1-x^{n+1}}{1-x}+x^{n+1}=\frac{1-x^{n+1}+x^{n+1}-x^{n+2}}{1-x}=\frac{1-x^{n+2}}{1-x},$$

where we used P_n to establish the second equality. The conclusion $1+x+x^2+x^3+\cdots+x^{n+1}=\frac{1-x^{n+2}}{1-x}$ is P_{n+1}.
By (i), (ii), and the induction principle, P_n holds for all $n\in\mathbb{N}$, which establishes the result. ■

The following is a slight variant of the induction principle, which is sometimes useful. It allows one to assume $P_1, P_2, \ldots, P_n$ in the induction step, instead of just P_n. This version may seem stronger than the usual induction principle, but it is logically equivalent to it.

Lemma 5.0.4 *(Generalized Induction Principle) Suppose that for each* $n\in\mathbb{N}$, P_n *is a statement (depending on* n*). Suppose (i)* P_1 *is true, and (ii) if* P_k *is true for all* $k=1,2,\ldots,n$, *then* P_{n+1} *is true. Then* P_n *is true for all* $n\in\mathbb{N}$.

PROOF. Suppose (i) and (ii) hold. For $n \in \mathbb{N}$, let Q_n be the statement that P_k is true for all $k \in \{1, 2, \ldots, n\}$. We apply the usual induction principle (that is, Lemma 5.0.1) to Q_n. First, Q_1 is true because Q_1 is the same as P_1, which is true by (i). For the inductive step, suppose Q_n is true. This means that $P_1, P_2, \ldots, P_n$ are all true. By (ii), then, P_{n+1} is true. Since we already know that $P_1, P_2, \ldots, P_n$ are true, we have that P_k is true for all $k \in \mathbb{N}$ satisfying $1 \leq k \leq n+1$. Hence Q_{n+1} is true. Thus by the induction principle, Q_n is true for all $n \in \mathbb{N}$. But P_n follows from Q_n, so P_n is true for all n. ■

Chapter 6

The Integers $\mathbb{Z}$ and the Rational Numbers $\mathbb{Q}$

We are still preparing to write down the axioms for $\mathbb{R}$. These axioms will include the existence of identity and inverse elements for two operations, addition and multiplication. So far we have considered the natural numbers $\mathbb{N}$. By considering identity and inverse elements for addition, we are led to the integers $\mathbb{Z}$. Then considering the identity and inverse elements for multiplication, we are led to the rational numbers $\mathbb{Q}$. Then we will see why we don't stop with the rational numbers: we will be missing numbers like $\sqrt{2}$. To include such "irrational" numbers, we will need to "complete" $\mathbb{Q}$ to get $\mathbb{R}$.

The integers $\mathbb{Z}$

In our assumptions for $\mathbb{R}$, we will assume that there is an operation, called *addition* and denoted "+." By operation, we mean that given two elements $x, y \in \mathbb{R}$, their addition, or sum, called $x + y$, is defined and is an element of $\mathbb{R}$. We will also assume that addition is commutative ($x + y = y + x$), associative ($(x + y) + z = x + (y + z)$), that there exists an identity element, called "0," so that $x + 0 = x$ for all $x \in \mathbb{R}$, and that every element $x \in \mathbb{R}$ has an additive inverse $-x$ such that $x + (-x) = 0$. (For those of you who have had some experience with abstract algebra, we can summarize these assumptions by saying that $(\mathbb{R}, +)$ is an abelian group.) Before considering $\mathbb{R}$, let's start with $\mathbb{N}$, which we have already considered, and see what these assumptions about addition will imply. First, in addition to $\mathbb{N} = \{1, 2, 3, \dots, \}$, we will have the additive identity element 0. Also, because of the assumption about inverses, for each $n \in \mathbb{N}$, we will have an element $-n$. Thus we will have at least the integers $\mathbb{Z}$, defined informally as

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$$

or more formally as $\mathbb{N} \cup \{0\} \cup \{-n : n \in \mathbb{N}\}$.

The rational numbers $\mathbb{Q}$

We will also assume the existence of an operation, called *multiplication*, denoted "$\cdot$," on $\mathbb{R}$, which is commutative ($x \cdot y = y \cdot x$), associative ($(x \cdot y) \cdot z = x \cdot (y \cdot z)$), that there exists an identity element, called "1," for multiplication, so that $x \cdot 1 = x$ for all $x \in \mathbb{R}$, and that every element $x \in \mathbb{R}$ such that $x \neq 0$ has a multiplicative inverse $\frac{1}{x}$ such that $x \cdot \frac{1}{x} = 1$. We usually omit the notation "$\cdot$" and write xy instead of $x \cdot y$. Since we have at least the integers contained in the real numbers, then for each $n, m \in \mathbb{Z}$ with $m \neq 0$, there exists the number $\frac{1}{m}$, and hence $n \cdot \frac{1}{m}$, which we write as $\frac{n}{m}$. Numbers of the form $\frac{n}{m}$ with $m \neq 0$ are called *rational* numbers (because they come from ratios of "whole" numbers, i.e., elements of $\mathbb{N}$); the form $\frac{n}{m}$ is called the *fraction* form of the number (there is also the decimal form, which we will generally not need to consider). The inverse is unique (if y and z are both inverses of x, then $yx = 1$ and $zx = 1$, so $y = y \cdot 1 = y(zx) = y(xz) = (yx)z = 1 \cdot z = z$). Then by commutativity and associativity, $\frac{1}{m\ell}$ and $\frac{1}{m} \cdot \frac{1}{\ell}$ are both inverses of $m\ell$, so $\frac{1}{m\ell} = \frac{1}{m} \cdot \frac{1}{\ell}$ when $m, \ell \neq 0$. Then the multiplication of rational numbers must take the form

$$\frac{n}{m} \cdot \frac{k}{\ell} = n \cdot \frac{1}{m} \cdot k \cdot \frac{1}{\ell} = (nk)\frac{1}{m\ell} = \frac{nk}{m\ell}.$$

Notice that the product of rational numbers is still a rational number. Also, by definition, $\frac{1}{1}$ is the multiplicative inverse of 1, but also 1 is the multiplicative inverse of 1 since $1 \cdot 1 = 1$, so by the uniqueness of inverses noted above, we have $\frac{1}{1} = 1$. Then more generally for $n \in \mathbb{Z}$ we have

$$n = n \cdot 1 = n \cdot \frac{1}{1} = \frac{n}{1},$$

so $\mathbb{Q}$ contains $\mathbb{Z}$ as a subset.

This characterization of multiplication leads to a tricky point: the same number can be represented in multiple ways. To see that this is true, consider a fraction of the form $\frac{kn}{km}$, where $k, n, m \in \mathbb{Z}$ and $k, m \neq 0$, so that the numerator and denominator have a common factor k. Then

$$\frac{kn}{km} = \frac{k}{k} \cdot \frac{n}{m} = k \cdot \frac{1}{k} \cdot \frac{n}{m} = 1 \cdot \frac{n}{m} = \frac{n}{m}.$$

That is, the fraction $\frac{kn}{km}$ is equal to the fraction $\frac{n}{m}$. This fact is commonly understood as saying that we can "cancel" the common term k from the numerator and denominator. When all common factors other than 1 have been cancelled, we say the fraction is in *reduced* form. This, although each rational number has infinitely many representations $\frac{kn}{km}$ for every $k \in \mathbb{N}$, there is one representation, the one in reduced form, which is simplest, and we often assume that we have selected that form.

Note that with $m_1, m_2 \neq 0$, we have that $\frac{n_1}{m_1} = \frac{n_2}{m_2}$ if and only if $n_1 m_2 = n_2 m_1$ (to prove this fact, multiply both sides of the equation $\frac{n_1}{m_1} = \frac{n_2}{m_2}$ by $m_1 m_2$ for the "only if" direction and multiply the equation $n_1 m_2 = n_2 m_2$ by $\frac{1}{m_1 m_2}$ for the "if" direction). Thus we define

$$\mathbb{Q} = \left\{ \frac{n}{m} : n, m \in \mathbb{Z}, m \neq 0, \right\}$$

with the understanding that $\frac{n_1}{m_1} = \frac{n_2}{m_2}$ if and only if $n_1 m_2 = n_2 m_1$. For those of you who have studied equivalence relations, they can be used to make this "understanding" mathematically precise. If we define the relation $\sim$ by defining $\frac{n_1}{m_1} \sim \frac{n_2}{m_2}$ if and only if $n_1 m_2 = n_2 m_1$, then $\sim$ forms an equivalence relation (we leave this exercise to the reader). Then $\mathbb{Q}$ consists of the equivalence classes under this equivalence relation.

We make one more algebraic assumption, the distributive property, about $\mathbb{R}$, that guarantees that addition and multiplication are linked in an appropriate way. The distributive property states that $a(b+c) = ab+ac$. This guarantees that the addition of rational numbers behaves the way we learned in grade school. First, if two rational numbers have the same denominator, we add them by adding the numerators, because, using the distributive property at the first step,

$$\frac{p}{m} + \frac{q}{m} = \frac{1}{m}(p+q) = \frac{p+q}{m}.$$

More generally, if we want to add two fractions that don't have the same denominator, we can first find alternate realizations of these numbers that do have the same denominator and then add them:

$$\frac{p}{n} + \frac{q}{m} = \frac{pm}{nm} + \frac{qn}{nm} = \frac{pm+qn}{nm},$$

as we learned in elementary school. Thus addition of rational numbers is defined, and always yields another rational number.

Why we need irrational numbers

The rational numbers provide a number system on which the operations of addition and multiplication are defined and satisfy all of the desired properties (commutativity, associativity, the distributive law, existence of identity and inverse elements). Why not stop with the rational numbers? Why do we need to go further and develop the real numbers? Why should there be any numbers other than the rational numbers? The ancient Greeks found ratios of integers to be pleasing, and assumed that all numbers were rational. The origin of the word " rational" with the meaning of "making sense" comes from the Greek idea that ratios are meaningful, clear concepts. To some degree, ratios of integers took on a kind of mystical or religious significance to the ancient Greeks. Therefore (according to legend, which is probably not true), Pythagoras,

who was working with right triangles, asked the question: if a right triangle has legs of length 1, what ratio of integers gives the number that represents the length ℓ of the hypoteneuse? From the Pythagorean theorem, he deduced that $\ell^2 = 1^2 + 1^2 = 2$, so he wanted to find $n, m \in \mathbb{N}$ (since $\ell > 0$, Pythagoras knew we could take both n and m to be positive) such that $\ell = \frac{n}{m}$. Much to his alarm, his reasoning after that showed that there are no such $n, m \in \mathbb{N}$. This was a kind of sacrilege to the ideas that all numbers should be ratios of integers, and legend has it that Pythagoras sacrificed a goat to atone for the sin of his discovery. Here was Pythagoras' reasoning. It is the first result of this course that is of sufficient depth that we call it a theorem. By "depth" we mean something that is not on the surface, i.e., one wouldn't naturally think it is true. One has to dig a little to find this truth. Pythagoras' argument is the most classic example of proof by contradiction.

Theorem 6.0.1 *(Pythagoras) There are no $n, m \in \mathbb{N}$ such that $\left(\frac{n}{m}\right)^2 = 2$.*

Proof. By way of contradiction, suppose $n, m \in \mathbb{N}$ satisfy $\left(\frac{n}{m}\right)^2 = 2$. If n and m have any common factors, we can cancel them to obtain $\frac{n}{m} = \frac{p}{q}$ with $p, q \in \mathbb{N}$ and $q \neq 0$, such that p and q have no common factors. Therefore

$$\left(\frac{p}{q}\right)^2 = 2, \text{ or } \frac{p^2}{q^2} = 2, \text{ or } p^2 = 2q^2.$$

Therefore p is even (since, if p is odd, then p^2 is odd, but $p^2 = 2q^2$ so p^2 is even). So $p = 2k$, for some $k \in \mathbb{N}$. Substituting $2k$ for p in the equation $p^2 = 2q^2$ gives

$$4k^2 = (2k)^2 = p^2 = 2q^2.$$

Dividing by 2 (or multiplying by $\frac{1}{2}$) yields $q^2 = 2p^2$. Therefore q^2 is even, so, by the same argument as for p above, it follows that q is even. But then p and q are both even, which means that they have the common factor 2. This conclusion contradicts our condition on p and q, and hence contradicts our assumption that $\left(\frac{n}{m}\right)^2 = 2$. Hence there are no $n, m \in \mathbb{N}$ such that $\left(\frac{n}{m}\right)^2 = 2$. ■

Pythagoras' result means that if we want to use numbers to measure lengths, we must use numbers that are not rational. This was a revolution in mathematical thinking that paves the way for the modern understanding of the real numbers.

The proof above is so important that it is worth understanding it in more detail. Somehow reaching the contradiction that p and q are both even seems odd; one might think that we just didn't cancel enough factors of 2 from m and n, so we should just cancel one more and start over. But that is incorrect, because we could then re-apply the same proof and still reach that p and q are both even. We could do this no matter how many factors of 2 we cancel, and we can't have an unlimited number of factors of 2 in n and m. If you are familiar with prime factorization, we know that an even number p can be written $p = 2^j \ell$, where ℓ is odd. That is, there is some fixed number j of factors of 2 in p. Then $p^2 = (2^j \ell) = 2^{2j} \ell^2$, and ℓ^2 is odd since ℓ is odd, so p^2 has exactly $2j$ factors of 2. In particular, p^2 has an even number of factors of 2. But the same is true for q^2: it has an even number of factors of 2. But the equation $p^2 = 2q^2$ is no longer possible: if q^2 has an even number of factors of 2, then $2q^2$ has one more factor of 2, hence an odd number of factors of 2. But then we can't have $p^2 = 2q^2$ since we know p^2 has an even number of factors of 2.

The proof we have written seems to depend on the properties of even and odd numbers, such as that the square of an odd number is odd. But if we think of even numbers just as numbers that have 2 as a factor, the same reasoning can be applied to numbers that have 3 as a factor, or 5 as a factor, etc. In particular, we can then show that there is no rational number r such that $r^2 = 3$, by the following approach. For $n, m \in \mathbb{N}$, let's use the notation $n|m$, spoken as "n divides m," to mean that n divides evenly into m, which, more precisely, means that there exists $k \in \mathbb{N}$ such that $m = kn$ (or $\frac{m}{n} \in \mathbb{N}$).

Now let's make an observation about divisibility. Suppose $p \in \mathbb{N}$ and $3|p^2$. Does it follow that $3|p$? The answer is yes, but let's see why. There is a general result about prime numbers that if $s \in \mathbb{N}$ is prime and $s|(nm)$, for $n, m \in \mathbb{N}$, then either $s|n$ or $s|m$ (i.e., either n or m must have s as part of it's prime factorization, since s is part of the prime factorization of nm). This fact would imply that $3|p$, by letting

$n = m = p$. However, we don't want to take the time to study prime numbers to the point that we can prove that general result. Instead, with a little extra work, we can prove in an elementary manner the part of that conclusion that we need. First we need a lemma that shows that we can divide one natural number into another, getting a natural number or 0, plus a remainder. This fact, called the *division algorithm*, may seem too obvious to prove, but we include the proof for completeness, and to show that it can be proved based on what we know.

Lemma 6.0.2 *Suppose $m, n \in \mathbb{N}$. Then we can write $n = km+r$, where $k \in \mathbb{N}\cup\{0\}$ and $r \in \{0, 1, 2, \ldots, m-1\}$*

In simple language, if we divide m into n, then m goes into n some non-negative integer number k times, with an integer remainder of r, where $0 \leq r < m$.

Proof. If $m = 1$ and $n \in \mathbb{N}$, we can write $n = n \cdot 1$, so the result holds with $k = n$ and $r = 0 < 1$, as required.

Now suppose $m > 1$. Fix the positive integer m. The proof is by induction on n. Let P_n be the statement that there exist non-negative integers k and r with $r < m$ such that $n = km + r$.

(i) P_1 is true: since $1 = 0 \cdot m + 1$, the result holds with $k = 0$ and $r = 1 < m$.

(ii) $P_n \Longrightarrow P_{n+1}$: Suppose P_n holds, so that $n = \ell m + s$, where $\ell \in \mathbb{N} \cup \{0\}$ and $s \in \{0, 1, \ldots, m-1\}$. If $s \neq m-1$, then $s+1 \in \{1, 2, \ldots, m-1\}$ and $n+1 = \ell m+s+1$ and the required conclusion holds with $k = \ell$ and $r = s+1$. The remaining possibility is that $s = m-1$. In that case, $n+1 = \ell m+s+1 = \ell m+m = (\ell+1)m$, so the required conclusion holds with $k = \ell + 1 \in \mathbb{N}$ and $r = 0$.

By induction, P_n holds for all $n \in \mathbb{N}$. ■

Let's suppose $3|p^2$, say $p^2 = 3\ell$ with $\ell \in \mathbb{N}$. We want to prove that $p = 3k$ for some $k \in \mathbb{N}$, which means that there is no remainder when we divide 3 into p, but we don't know that yet. But by Lemma 6.0.2, we can divide 3 into p, to write $p = 3k + j$ where $k \in \mathbb{N} \cup 0$ and $j \in \{0, 1, 2\}$; i.e, j is the remainder when we divide 3 into p. Because $p^2 = 3\ell$ and $p = 3k + j$, we have

$$3\ell = p^2 = (3k + j)^2 = 9k^2 + 6kj + j^2.$$

Then $j^2 = 3\ell - 9k^2 - 6kj = 3(\ell - 3k^2 - 2kj)$. Let $m = \ell - 3k^2 - 2kj$ and note that $m \in \mathbb{Z}$. So we have $j^2 = 3m$. Recall that $j \in \{0, 1, 2\}$. If $j = 1$, we have $1 = 3m$, or $m = \frac{1}{3}$, which is impossible since $m \in \mathbb{Z}$. If $j = 2$ then we have $4 = j^2 = 3m$, or $m = \frac{4}{3}$, which is impossible for the same reason. So we must have $j = 0$, so $p = 3k + 0 = 3k$, so $3|p$.

Equipped with the fact that $3|p^2 \Longrightarrow 3|p$, we can show that there is no rational number r satisfying $r^2 = 3$ by virtually the same argument as for Theorem 6.0.1.

Theorem 6.0.3 *There are no $n, m \in \mathbb{N}$ such that $\left(\frac{n}{m}\right)^2 = 3$.*

Proof. By way of contradiction, if such n, m exist, then we can find $p, q \in \mathbb{N}$ satisfying $\left(\frac{p}{q}\right)^2 = 3$ such that p and q have no common factors. Then

$$\left(\frac{p}{q}\right)^2 = 3, \text{ or } p^2 = 3q^2.$$

Therefore $3|p^2$. By our observation just preceding this proof, it follows that $3|p$. That is, $p = 3k$, for some $k \in \mathbb{N}$. Substituting $3k$ for p in the equation $p^2 = 3q^2$ gives

$$9k^2 = (3k)^2 = p^2 = 3q^2.$$

Multiplying by $\frac{1}{3}$ yields $q^2 = 3p^2$. Therefore $3|q^2$, so $3|q$. But then p and q have the common factor 3. This conclusion contradicts our condition on p and q, and hence contradicts our assumption that $\left(\frac{n}{m}\right)^2 = 3$. Hence there are no $n, m \in \mathbb{N}$ such that $\left(\frac{n}{m}\right)^2 = 3$. ■

It may seem like the argument above will work with any number in place of 2 or 3. Why doesn't the same argument prove that there is no rational number r satisfying $r^2 = 4$ (which is obviously false)? Where does the argument break down? Suppose $\left(\frac{p}{q}\right)^2 = 4$ where p and q have no common factors. Then $p^2 = 4q^2$. Thus $4|p^2$. Does it follow that $4|p$? No, and this is where the argument fails. For example, if $p = 6$, then $p^2 = 36$ so $4|p^2$. But 4 does not divide evenly into 6. Going even further, why doesn't the approach above show that if $4|p^2$, then $4|p$? Let's try to follow that argument. We can write $p = 4k + j$ where $j \in \{0, 1, 2, 3\}$. Then $p^2 = 4\ell$ for some $\ell \in \mathbb{N}$, so we have

$$4\ell = p^2 = (4k + j)^2 = 16k^2 + 8kj + j^2.$$

Hence $j^2 = 4\ell - 16k^2 - 8kj = 4(\ell - 4k^2 - 2kj) = 4m$ for $m = \ell - 4k^2 - 2kj \in \mathbb{Z}$. We then consider the possible values $j = 0, 1, 2, 3$, trying to rule out $j = 1, 2, 3$ to force $j = 0$. We can rule out $j = 1$ and $j = 3$ this way, but we cannot rule out $j = 2$, because when $j = 2$ we get $4 = j^2 = 4m$, which holds for $m = 1 \in \mathbb{Z}$. So that part of the argument breaks down.

We went through the explanation above because sometimes to get a full understanding of a proof, it helps to see its limitations and/or restrictions.

However, let's not lose sight of our main point, which is that the rational numbers are missing some values that we would really like to have, like $\sqrt{2}$. So we need to construct a "complete" number system which contains $\mathbb{Q}$. This leads us to the real numbers $\mathbb{R}$.

Chapter 7

Fields and the Algebraic Properties of $\mathbb{R}$

We will define the real numbers $\mathbb{R}$ to be a "complete ordered field," and any complete ordered field turns out to be "equivalent" to $\mathbb{R}$. We begin by defining a field; the field axioms describe the algebraic structure of $\mathbb{R}$. Then we describe the order properties of $\mathbb{R}$. Finally we will describe the completeness property of $\mathbb{R}$, which is more subtle.

Fields

Definition 7.0.1 *A field F is a set with operations $+$ and $\cdot$ (which means that for all $a, b \in F$, the sum $a + b$ and the product $a \cdot b$ are defined and are elements of F) satisfying:*

A1) $a + (b + c) = (a + b) + c$ for all $a, b, c \in F$ (associativity of addition);
A2) $a + b = b + a$ for all $a, b \in F$ (commutativity of addition);
A3) there exists $0 \in F$ such that $a + 0 = a$ for all $a \in F$ (existence of additive identity);
A4) for all $a \in F$, there exists an element $-a \in F$ such that $a + (-a) = 0$ (existence of additive inverses);

M1) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in F$ (associativity of multiplication);
M2) $a \cdot b = b \cdot a$ for all $a, b \in F$ (commutativity of multiplication);
M3) there exists $1 \in F$ such that $1 \neq 0$ and $a \cdot 1 = a$ for all $a \in F$ (existence of multiplicative identity);
M4) for all $a \in F$ such that $a \neq 0$, there exists an element $a^{-1} \in F$ such that $a \cdot a^{-1} = 1$ (existence of multiplicative inverses);

and

D) $a \cdot (b + c) = a \cdot b + a \cdot c$ for all $a, b, c \in F$ (distributive law).

Note that $A1 - A4$ describe the properties of addition (i.e., that $(F, +)$ is an "abelian group"), whereas $M1 - M4$ are the corresponding properties of multiplication (i.e., that $(F \setminus \{0\}, \cdot)$ is an abelian group). The distributive property D links addition and multiplication, making them compatible. We make the remark that a mathematical equation such as $ab = ba$ is interpreted as "ab and ba are the same element." In other words, "$=$" stands for identity; the two sides of an equation are the same object. In particular, the statement "$x = y$" has the same meaning as the statement "$y = x$."

We will usually write $\frac{1}{a}$ instead of a^{-1} and ab instead of $a \cdot b$. Also we follow the standard order of operations, in which multiplications are performed first, then additions, so that $ab + c$ means $(ab) + c$, not $a(b + c)$.

The assumptions in Definition 7.0.1 fit what we believe for $\mathbb{R}$, so they are somewhat natural. However, $\mathbb{R}$ is not the only field; for example, the rational numbers $\mathbb{Q}$ and the complex numbers $\mathbb{C}$ are fields. There are even finite fields, which you may or may not have studied: $\mathbb{Z}_p$, the integers modulo p, where p is a prime.

The goal of having axioms is to clarify exactly what is assumed; everything else needs to be proved, based on the axioms. In choosing the axioms, one hopes that they will all be reasonable, and one wants to have

as few as possible. The axioms for a field were developed over a long time, and found to be the minimum necessary to obtain all of the standard properties of addition and multiplication. The next proposition lists several standard properties; the proof shows that they are all derived from the field axioms in Definition 7.0.1.

Proposition 7.0.2 *Let F be a field. Then for any $a, b, c \in F$,*

(i) If $a + c = b + c$ then $a = b$,
(ii) $a \cdot 0 = 0$,
(iii) $(-a) \cdot b = -(a \cdot b)$,
(iv) $(-a) \cdot (-b) = a \cdot b$,
(v) if $a \cdot c = b \cdot c$ and $c \neq 0$ then $a = b$,
(vi) if $ab = 0$ then either $a = 0$ or $b = 0$;
(vii) $-(-a) = a$; and
(viii) $-0 = 0$.

Proof. (i) Let's write the proof two ways. For the first way, we have

$$a + c = b + c$$

by assumption. We can add $-c$ (which exists, by A4) to both sides, since $a + c$ and $b + c$ are the same element of F:

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

By A1, applied on each side of the equation, we get

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

By A4, $c + -c = 0$, so

$$a + 0 = b + 0.$$

But $a + 0 = a$ and $b + 0 = b$, by A3. So we have

$$a = b.$$

This way of writing the proof has the advantage of starting with the assumption, and proceeding step-by-step until the desired conclusion is obtained. However, it is somewhat long-winded. A more concise way to write the same argument is:

$$a \overset{A3}{=} a + 0 \overset{A4}{=} a + (c + -c) \overset{A1}{=} (a + c) + -c \overset{a+c=b+c}{=} (b + c) + -c \overset{A1}{=} b + (c + -c) \overset{A3}{=} b + 0 \overset{A4}{=} b.$$

For the remaining parts, we use the second, shorter way of writing the proofs.

(ii) We have

$$0 + a \cdot 0 \overset{A2}{=} a \cdot 0 + 0 \overset{A3}{=} a \cdot 0 \overset{A3:0=0+0}{=} a \cdot (0 + 0) \overset{D}{=} a \cdot 0 + a \cdot 0.$$

Hence, by (i), we conclude that $0 = a \cdot 0$.

(iii) We have

$$(-a)b + ab \overset{M2}{=} b(-a) + ba \overset{D}{=} b(-a + a) \overset{A2}{=} b(a + -a) \overset{A4}{=} b \cdot 0 \overset{(ii)}{=} 0 \overset{A4}{=} ab + -(ab) \overset{A2}{=} -(ab) + ab.$$

Since we have $(-a)b + ab = -(ab) + ab$, then, by (i), $(-a)b = -(ab)$.

(iv) We have

$$(-a)(-b) + (-a)b \overset{D}{=} -a(-b + b) \overset{A2}{=} -a(b + -b) \overset{A4}{=} -a(0) \overset{(ii)}{=} 0 \overset{A4}{=} ab + -(ab) \overset{(iii)}{=} ab + -(a)b.$$

Therefore we have $(-a)(-b) + (-a)b = ab + (-a)b$. By (i), we conclude that $(-a)(-b) = ab$.

(v) Since $c \neq 0$, by M4 there exists $c^{-1} \in F$ such that $c \cdot c^{-1} = 1$. Hence

$$a \overset{M3}{=} a \cdot 1 \overset{M4}{=} a \cdot (c \cdot c^{-1}) \overset{M1}{=} (a \cdot c) \cdot c^{-1} \overset{a \cdot c = b \cdot c}{=} (b \cdot c) \cdot c^{-1} \overset{M1}{=} b \cdot (c \cdot c^{-1}) \overset{M4}{=} b \cdot 1 \overset{M3}{=} b.$$

(vi) If $b = 0$ then the conclusion holds. If $b \neq 0$ then by M4 there exists $b^{-1} \in F$ such that $b \cdot b^{-1} = 1$. Hence

$$a \overset{M3}{=} a \cdot 1 \overset{M4}{=} a \cdot (b \cdot b^{-1}) \overset{M1}{=} (a \cdot b) \cdot b^{-1} \overset{a \cdot b = 0}{=} 0 \cdot b^{-1} \overset{M1}{=} b^{-1} \cdot 0 \overset{(ii)}{=} 0.$$

(vii) $-(-a) + -a \overset{A2}{=} -a + -(-a) \overset{A4}{=} 0 \overset{A4}{=} a + -a$, so by (i), $-(-a) = a$.

(viii) $-0 + 0 \overset{A2}{=} 0 + -0 \overset{A4}{=} 0 \overset{A3}{=} 0 + 0$, so by (i), $-0 = 0$.

■

It is sometimes useful to note that $-b = (-1) \cdot b$, which follows by letting $a = 1$ in (iii). Then, for example, $-(a + b) = (-1) \cdot (a + b) = (-1) \cdot a + (-1) \cdot b = -a + -b$ follows from the distributive property D (of course there are other ways to prove this fact).

We have presented the details of these proofs to exhibit how all of the usual properties of addition and multiplication can be derived from the field axioms. Justifying further algebraic properties (such as the rules for exponents) is time-consuming and more in the provence of abstract algebra than analysis, so we will not carry out all of these proofs. We hope the reader is convinced that it is possible to prove all of the standard facts of algebra starting only with the field axioms. From now on, we will assume that is the case, and we will use standard algebraic facts without further comment.

Chapter 8

Ordered Fields and the Order Properties of $\mathbb{R}$

Ordered Fields

In addition to being a field, the real numbers are an ordered field, which we will soon define. An "order" on a set S is an example of a *relation*, so we begin by defining a relation. A relation R on a set S is a set of ordered pairs (x, y) with $x, y \in S$; i.e., R is a subset of $S \times S$. We say that x is related to y if $(x, y) \in R$. The concept of a relation is very abstract, but for us saying that $<$ is a relation on a field F just means that for two elements x and y of F, it may be the case that $x < y$. The relation R describes the set of x and y which are related to each other. An example of a relation on the set of human beings is motherhood, where we say that person x is related by motherhood to person y if x is the (biological) mother of y (let's ignore surragate mothers for the sake of this example). Given two people x and y, they are either related by motherhood or they are not, which is all that is needed for motherhood to define a relation. It is important to note that relations involve only 2 elements; there is no such thing as a relation that relates, say, 3 elements of a set.

Definition 8.0.1 *An ordered field $(F, <)$ is a field with a relation "$<$," called an order, which satisfies*

O1) For all $a, b \in F$, one and only one of the following holds: $a < b, a = b$, or $b < a$;
O2) If $a, b, c \in F$, $a < b$, and $b < c$, then $a < c$;
O3) If $a, b, c \in F$ and $b < c$, then $a + b < a + c$;
O4) If $a, b, c \in F$, $b < c$, and $0 < a$, then $ab < ac$.

We read $a < b$ as "a is less than b" or "a is smaller than b." We also define the relation "$>$," read as "greater than" or "larger than" or "bigger than" as follows: we say $b > a$ if and only if $a < b$. In O1, the statement $a = b$ again means that a and b are the same element. That is why statements of the form $a < b$ or $a > b$ are called "inequalities:" by O1 $a < b$ or $a > b$ rules out the possibility of the equality $a = b$. Moreover O1 states that there is always a comparison: if a and b are not the same element, then one of them is larger than the other. (There is a different concept of a "partial ordering" in which this comparison principle does not hold; an example of a partial inclusion is the subset relation, where a set S is related to a set T if $S \subseteq T$.) If $0 < a$, we say a is *positive*. If $a < 0$ we say that a is *negative*.

O2 is the natural transitivity property of the relation $<$. O3 guarantees that the order relation is consistent with the addition operation (adding the same thing to both sides of an inequality preserves the inequality), and O4 shows consistency with multiplication (multiplying an inequality by a positive quantity preserves the inequality). The real numbers $\mathbb{R}$ and the rational numbers Q are examples of ordered fields.

As for the field axioms, we claim that all of the familiar properties of inequalities follow from the axioms O1-O4 and the field axioms. The next Proposition gives credence to this claim.

Proposition 8.0.2 *Let $(F, <)$ be an ordered field. Then for any $a, b, c \in F$,*

(i) if $a < b$ then $-b < -a$;
(ii) if $a < b$ and $c < 0$, then $bc < ac$;

(iii) if $0 < a$ *and* $0 < b$ *then* $0 < ab$*;*
(iv) if $a \neq 0$ *then* $0 < a^2$*;*
(v) $0 < 1$*;*
(vi) if $0 < a$ *then* $0 < a^{-1}$*;*
and
(vii) if $0 < a < b$ *then* $0 < b^{-1} < a^{-1}$*.*

PROOF. In this proof, we will utilize the associativity and commutativity properties (A1, A2, M1, and M2 from Section 7) of F repeatedly without spelling out their use.

(i) We use O3 to add $-b - a$ to both sides of the inequality $a < b$, obtaining

$$-b = -b + 0 = -b - a + a < -b - a + b = -a + -b + b = -a + 0 = -a.$$

(ii) Since $c < 0$, we have $0 < -c$ by (i). Hence by O4 and our assumption $a < b$ we obtain $a(-c) < b(-c)$, or, by Proposition 7.0.2 (iii), $-(ac) < -(bc)$. By (i) again, then, we have $--(bc) < --(ac)$. Using Proposition 7.0.2 (vii), we have $bc < ac$.

(iii) Multiplying the assumed inequality $0 < a$ on both sides by b, which is assumed to be positive, gives (using Proposition 7.0.2 (ii)) that $0 = 0b < ab$ by O4.

(iv) By O1, either $0 < a$ or $a < 0$, since the assumption is that $a \neq 0$. If $0 < a$ then $a^2 = a \cdot a > 0$ by (iii). If $a < 0$, then $0 < -a$ by (i), so $a^2 = a \cdot a = (-a) \cdot (-a) = (-a)^2 > 0$ by the case we just proved, applied to $-a > 0$.

(v) $0 < 1^2 = 1$, by (iv), since $1 \neq 0$ by M3 of Definition 7.0.1.

(vi) Suppose $a^{-1} = 0$. Then $1 = a \cdot a^{-1} = a \cdot 0 = 0$, which contradicts (v). Now suppose $a^{-1} < 0$. Since $0 < a$ by assumption, multiplying the inequality $a^{-1} < 0$ by a preserves the inequality, so $1 = a \cdot a^{-1} < a \cdot 0 = 0$, which again contradicts (v). Hence, by O1, the only possibility is that $0 < a^{-1}$.

(vii) Since $0 < a$ we have $0 < a^{-1}$ by O2. Also, from $0 < a < b$ we deduce $0 < b$ by O2, and hence $0 < b^{-1}$ by (vi). Since $0 < b^{-1}$, multiplying the equation $a < b$ by b^{-1} preserves the inequality, so $ab^{-1} < bb^{-1} = 1$. Since $0 < a^{-1}$ multiplying the equation $ab^{-1} < 1$ by a^{-1} preserves the inequality, so $a^{-1}ab^{-1} < a^{-1} \cdot 1$, hence $b^{-1} < a^{-1}$. ■

Result (vii) shows us, for example, that $3 < 4$ implies that $\dfrac{1}{4} < \dfrac{1}{3}$.

We also define a relation "$\leq$" on an ordered field as follows: $a \leq b$ if either $a < b$ or $a = b$. By O1, then, the negation of the statement $a \leq b$ is the statement $b < a$. Then the following analogues of the statements above hold, for all $a, b, c \in F$:

$O1'$) : either $a \leq b$ or $b < a$;
$O2'$) : if $a \leq b$ and $b \leq c$ then $a \leq c$;
$O3'$) : if $b \leq c$ then $a + b \leq a + c$;
$O4'$) : if $b \leq c$ and $a \geq 0$, then $ab \leq ac$;
$(i)'$: if $a \leq b$ then $-b \leq -a$;
$(ii)'$: if $a \leq b$ and $c \leq 0$ then $bc \leq ac$;
$(iii)'$: if $0 \leq a$ and $0 \leq b$ then $0 \leq ab$;
$(iv)'$: $0 \leq a^2$.

There is no version of (v) that involves equality, and we don't have analogues of (vi) or (vii) because a^{-1} is not defined if $a = 0$. Proving all of these statements involving $\leq$ just requires using the result for $<$ and then also considering the result if we allow equality; e.g., to prove $O3'$: we know the result if $b < c$ by O3, whereas if $b = c$ then $a + b = a + c$. We leave checking all of the details of the rest to the reader. We also define the relation $\geq$ by $a \geq b$ if $a > b$ or $a = b$; then $a \geq b$ if and only if $b \leq a$.

The $\leq$ relation is of great importance in analysis, because many of the key results in analysis involve estimating one quantity by another. If an analyst is asked to prove an identity $a = b$, there is a good chance he or she will proceed by proving $a \leq b$ and $b \leq a$.

Just as for the algebraic properties of $\mathbb{R}$, we have presented the axioms for the ordering $<$ just to make clear that there are axioms, and to convince the reader that all of the standard properties of the ordering of $\mathbb{R}$ (such as: if you multiply an iequality by a negative number, you reverse the inequality) follow from the axioms. From now on we will assume standard facts about the ordering of $\mathbb{R}$. However, there are some key facts that may not be so obvious, which we present now. These facts involve the *magnitude* or *absolute value* $|a|$ of an element a of F, defined as follows.

Absolute Values and the Triangle Inequality

Definition 8.0.3 *Let $(F, <)$ be an ordered field. For $a \in F$, define $|a| \in F$ by*

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

The basic facts about the absolute value are contained in the next result.

Proposition 8.0.4 *Let $(F, <)$ be an ordered field. Then for all $a, b \in \mathbb{F}$,*

(i) $|a| \ge 0, a \le |a|$, and $-a \le |a|$;
(ii) $|ab| = |a||b|$;
and
(iii) $|a + b| \le |a| + |b|$ (triangle inequality).

Proof. (i) First suppose $0 \le a$. In that case, $|a| = a$, so $|a| = a \ge 0$ holds, and $a \le |a|$ holds because $|a| = a$. Since $0 \le a$, we have $-a \le 0$ by $(i)'$, and hence $-a \le 0 \le a = |a|$ implies $-a \le |a|$ by $O2'$.

Now suppose $a < 0$. Then $|a| = -a$ by definition, and $-a > 0$ by Proposition 8.0.2 (i). Hence $|a| = -a \ge 0$. Also $a < 0 < -a = |a|$, so $a \le |a|$. Finally $-a = |a|$ so $-a \le |a|$.

(ii) Again we consider all possible cases. First suppose $a \ge 0$ and $b \ge 0$. Then $|a| = a$ and $|b| = b$ by definition. By $(iii)'$, we have $ab \ge 0$, so $|ab| = ab$. Hence $|ab| = ab = |a||b|$.

Second, suppose $a \ge 0$ and $b < 0$. Then $|a| = a$ and $|b| = -b$. By $(ii)'$ (with a replaced by 0, b replaced by a, and c replaced by b), we have $ab \le 0$. Then $|ab| = -ab$ (this fact follows from the definition of absolute value if $ab < 0$, but it also holds in case $ab = 0$ because both sides are 0). Hence $|ab| = -ab = a(-b) = |a||b|$.

Third, suppose $a < 0$ and $b \ge 0$. By reversing the roles of a and b, the previous case shows that $|ab| = |a||b|$ in this case also.

Fourth, suppose $a < 0$ and $b < 0$. Then $|a| = -a$ and $|b| = -b$. Also $ab = (-a)(-b) > 0$ by Proposition 8.0.2 applied to $-a$ and $-b$, so $|ab| = ab = (-a)(-b) = |a||b|$.

(iii) First suppose $a + b \ge 0$. In that case, $|a + b| = a + b$, by definition of absolute value. Notice that since $a \le |a|$ by (i), we get $a + b \le |a| + b$ by $O3'$. Similarly, since $b \le |b|$, we can add $|a|$ to both sides of the inequality $b \le |b|$ to get $|a| + b \le |a| + |b|$. These two inequalities give $a + b \le |a| + b \le |a| + |b|$, so transitivity (i.e., $O2'$) gives

$$|a + b| = a + b \le |a| + |b|,$$

Now suppose $a+b < 0$. Then $|a+b| = -(a+b) = -a+-b$. Then, similarly to the first case, we use $-a \le |a|$ and $-b \le |b|$ (by (i)) to obtain

$$|a + b| = -a + -b \le |a| + -b \le |a| + |b|.$$

Hence we have $|a + b| \le |a| + |b|$ in both possible cases. ■

The triangle inequality is used often in analysis to make estimates.

We can use the absolute value to define the distance $d(a, b)$ between two points $a, b \in F$, as follows.

Definition 8.0.5 *Let $(F, <)$ be an ordered field. For $a, b \in F$, define $d(a, b) = |b - a|$.*

This quantity d has the natural properties we would think a "distance" should have, as follows.

Proposition 8.0.6 *The distance d on an ordered field $(F, <)$ satisfies: for all $a, b, c \in F$,*

(i) $d(a, b) \geq 0$;
(ii) $d(a, b) = 0$ if and only if $a = b$;
(iii) $d(a, b) = d(b, a)$;
(iv) $d(a, c) \leq d(a, b) + d(b, c)$ (general triangle inequality).

Proof.
(i) We have $d(a, b) = |b - a| \geq 0$ by Proposition 8.0.4 (i).

(ii) If $a = b$, then $b - a = 0$, so $d(a, b) = |b - a| = |0| = 0$ by definition of absolute value. Conversely, if $d(a, b) = 0$, then $|b - a| = 0$, which can only happen if $b - a = 0$ (from the defintion of absolute value: if $c > 0$ then $|c| = c \neq 0$, and if $c < 0$ then $|c| = -c \neq 0$, so the only way $|c| = 0$ is if $c = 0$), hence $b = a$.

(iii) $d(a, b) = |b - a| = |(-1)(a - b)| = |-1||a - b| = 1 \cdot |a - b| = |a - b| = d(b, a)$, where we used Proposition 8.0.4 (ii) to obtain $|(-1)(a - b)| = |-1||a - b|$.

(iv) $d(a, c) = |c - a| = |c - b + b - a| \leq |c - b| + |b - a| = d(b, c) + d(a, b) = d(a, b) + d(b, c)$, where we used the triangle inquality applied to the elements $c - b$ and $b - a$ to obtain $|c - b + b - a| \leq |c - b| + |b - a|$. ■

After understanding $\mathbb{R}$, the next setting one should consider in $\mathbb{R}^n$, or n-dimensional Euclidean space. There is a natural notion of distance d on $\mathbb{R}^n$, also defined by $d(a, b) = |b - a|$, where here a and b are vectors (e.g., $a = (a_1, a_2, \ldots, a_n)$ where each $a_i \in \mathbb{R}$) and $|b - a|$ is the length of the vector from a to b. Then Property (iv) of Proposition 8.0.6 is called the triangle inequality because it says that the straight line distance between a and c is less than or equal to the straight line distance from a to b plus the straight line distance from b to c; that is, it is shorter to go directly along the base of the triangle from a to c than to go along one leg from a to b and then along the third leg of the triangle from b to c. We call the inequality $|a + b| \leq |a| + |b|$ in Proposition 8.0.4 (iii) the triangle inequality because it is equivalent to $d(a, c) \leq d(a, b) + d(b, c)$ in $\mathbb{R}$, even though there aren't any nondegenerate triangles in $\mathbb{R}$.

More generally, if we have any set S and a distance function d defined on $S \times S$ satisfying the properties (i)-(iv) of Proposition 8.0.6, we say that (S, d) is a *metric space*. Thus $\mathbb{R}$ and $\mathbb{R}^n$ are examples of metric spaces. Nearly all of the spaces considered in analysis are metric spaces, including the spaces of functions mentioned in Examples 2 and 3 of Chapter 1. Much of what we will consider (open sets, closed sets, continuity, etc.) in this course can be considered on a general metric space, often with virtually the same proofs. Nevertheless, it is best to understand these concepts in the familiar context of $\mathbb{R}$ before exporting them to more exotic circumstances.

As noted before, $\mathbb{Q}$ and $\mathbb{R}$ are both ordered fields. What distinguishes $\mathbb{R}$ is the property of completeness (as motivated in Section 6), which we discuss next.

Chapter 9

The Completeness Axiom and the Definition of $\mathbb{R}$

The Least Upper Bound, or Supremum, of a Set

We have one more step left to describe $\mathbb{R}$ axiomatically. The algebraic properties of $\mathbb{R}$ are fully described by saying that $\mathbb{R}$ is a field, as defined in Chapter 7. However, there are lots of fields. In the last chapter, we considered ordered fields $(F, <)$, and it will turn out that $\mathbb{R}$ is an ordered field. Most fields do not have orders; in particular the complex numbers $\mathbb{C}$ do not have an ordering (the proof of that fact is left as an exercise). However, the rational numbers $\mathbb{Q}$ also form an ordered field. We argued in Section 6 that $\mathbb{Q}$ is "missing" some values, like $\sqrt{2}$, that we want to have in our number system. To guarantee that no numbers are missing in $\mathbb{R}$, we will include a completeness axiom for $\mathbb{R}$. To atate the completeness property, we need some preliminary defintions.

Definition 9.0.1 *Let $(F, <)$ be an ordered field, and let $A \subseteq F$.*

(i) A is bounded above if there exists some $b \in F$ such that $a \leq b$ for all $a \in A$. If such an element b exists, we call b an upper bound for A.

(ii) A is bounded below if there exists some $c \in F$ such that $c \leq a$ for all $a \in A$. If such an element c exists, we call c an lower bound for A.

(iii) A is bounded if A is both bounded above and bounded below.

Let's consider some examples in the real numbers. We will use the standard notation for intervals from here on, as follows. Let $-\infty \leq a < b \leq +\infty$. Then

$$(a, b) = \{x \in \mathbb{R} : a < x < b\}.$$

If $a \neq -\infty$, define

$$[a, b) = \{x \in \mathbb{R} : a \leq x < b\}.$$

If $b \neq +\infty$ (but possibly $a = -\infty$), let

$$(a, b] = \{x \in \mathbb{R} : a < x \leq b\}.$$

If $a \neq -\infty$ and $b \neq +\infty$, let

$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}.$$

Example 9.0.2 *Let $F = \mathbb{R}$ with the usual ordering $<$. Let $A = (-2, 3)$. Then A is bounded above, with examples of upper bounds including $3, 4, 27$, and $10^{1,000}$. Notice that no upper bound that we can find for A belongs to A itself. The definition of an upper bound only requires the upper bound to belong to F.*

A is also bounded below, with, for example, lower bound -6. Note that -1 is not a lower bound because there are elements of A, for example -1.5, which are less than -1.

Since A is bounded above and below, A is bounded.

Example 9.0.3 *Let $F = \mathbb{R}$ with the usual ordering $<$. Let $\mathbb{N} = \{1, 2, 3, \ldots\}$ denote the natural numbers. Then (intuitively, and we will prove this fact later!) $\mathbb{N}$ is not bounded above, and hence is not bounded. However, $\mathbb{N}$ is bounded below, for example by 0 or by 1. Note that in this case, there is a lower bound, namely 1, which belongs to the set $\mathbb{N}$.*

Definition 9.0.4 *Suppose F is an ordered field and $A \subseteq F$.*

(i) If there exists $M \in A$ such that M is an upper bound for A, we call M the maximum of A. In this case we write $M = \max A$.

(ii) If there exists $m \in A$ such that m is a lower bound for A, we call m the minimum of A. In this case we write $m = \min A$.

There can be at most one maximum of a set A, since if M_1 and M_2 are both maxima of A, then $M_1 \leq M_2$ (since $M_1 \in A$ and M_2 is an upper bound for A), and similarly $M_2 \leq M_1$, so $M_1 = M_2$. It is important to observe that although some sets have a maximum (e.g., 1 is the maximum of the interval $(0, 1]$), some sets do not have maxima (e.g., $(0, 1)$ does not have a maximum since a maximum must be an upper bound that is in the set). Similarly, some sets have a minimum (e.g., $[0, 1)$ but some do not (e.g., $(0, 1)$).

We leave it as an exercise in induction for the reader to prove that any finite set of real numbers has a maximum and a minimum.

In cases where a set is bounded above but does not have a maximum, it seems natural to look for the "best" upper bound. There are similar remarks for lower bounds, but to avoid jumping back and forth between upper and lower bounds, let's just think about upper bounds for the time being. Once upper bounds are understood it will be easy to consider lower bounds similarly.

Definition 9.0.5 *Let $(F, <)$ be an ordered field and suppose $A \subseteq F$. We say that $s \in F$ is the least upper bound, or supremum, of A, if*

(i) s is an upper bound for A,
and
(ii) if $t \in F$ is another upper bound for A, then $s \leq t$.

If s is the least upper bound for A, we write $s = \sup A$.

It should be clear that we can talk about "the" least upper bound of A, because if there were two least upper bounds s_1 and s_2 for A, then by (i), s_1 and s_2 are both upper bounds for S, and so by (ii), we would have $s_1 \leq s_2$ and $s_2 \leq s_1$, so $s_1 = s_2$.

The word "supremum" may be less descriptive than "least upper bound" because it does not as clearly imply that the supremum is the smallest value that is a an upper bound. The word supremum is used just because it is more elegant to abbreviate: $\sup A$ looks much nicer than "lub A."

For the next two examples, we assume standard facts about $\mathbb{R}$, although we have not defined $\mathbb{R}$ formally yet.

Example 9.0.6 *Let $F = \mathbb{R}$ with the usual ordering $<$. Let $A = [-1, 1]$. We claim that $\sup A = 1$. This fact may seem obvious, but the main point is to see how the definition of supremum applies. To show that $\sup A = 1$, we need to show the two properties (i) and (ii) in Definition 9.0.5.*

(i) By definition, if $a \in [-1, 1]$, then $a \leq 1$, so 1 is an upper bound for A.

(ii) Suppose t is an upper bound for A. Since $1 \in [0, 1]$, then $1 \leq t$. Hence property (ii) holds for $s = 1$.

Hence $1 = \sup A$.

Example 9.0.7 *Let $F = \mathbb{R}$ with the usual ordering $<$. Let $A = (0, 1)$. We claim that $\sup A = 1$. Property (i) is easy, because $a \in (0, 1)$ implies that $a < 1$, so 1 is an upper bound for A. To prove (ii), suppose t is an upper bound for $(0, 1)$. We need to show that $t \geq 1$. We prove that $t \geq 1$ by contradiction. Suppose $t < 1$. Since $\frac{1}{2} \in (0, 1)$, we know $t \geq \frac{1}{2} > 0$. So $0 < t < 1$. Adding one to all terms of these inequalities gives $1 < t + 1 < 2$. Dividing by 2 gives*

$$\frac{1}{2} < \frac{t+1}{2} < 1,$$

so, in particular, $\frac{t+1}{2} \in (0,1)$*. However, taking the inquality* $t < 1$ *and adding* t *to both sides gives* $2t < t+1$*, and then dividing by* 2 *gives* $t < \frac{t+1}{2} \in (0,1)$*. Hence* t *is not an upper bound for* $(0,1)$*, contradicting our assumption. Therefore we have proved that* $t \geq 1$*, or* $1 \leq t$*. Since this conclusion holds for any upper bound* t*, we conclude property (ii) in the definition of the supremum. Since properties (i) and (ii) hold,* 1 *is the least upper bound, or supremum, of* A*.*

Notice that in the last example, $\sup A \notin A$, whereas in the previous example we had $\sup A \in A$. It is important to remember that a set's supremum may or may not be in the set.

When a set A has a maximum M, then M is the supremum of A as well, because M is an upper bound for A and if there is any other upper bound t for A, then $M \leq t$ since $M \in A$.

Example 9.0.8 *Let* $F = \mathbb{R}$*, with the usual ordering* $<$*. Let*

$$A = \{x \in \mathbb{R} : x^2 < 2\}.$$

We expect that A *is the interval* $(-\sqrt{2}, +\sqrt{2})$*, and that* $\sup A = \sqrt{2}$*. This expectation will turn out to be correct, but we are not yet in a position to prove it, because all we know about the real numbers so far is that they form an ordered field. We do not yet know that there is a real number* $\sqrt{2}$*, and the ordered field axioms are not enough guarantee the existence of* $\sqrt{2}$*, as we see from the next example.*

Example 9.0.9 *Let* $F = \mathbb{Q}$*, the rational numbers, with the usual ordering* $<$*. (Recall that the rational numbers form an ordered field.) Let*

$$A = \{x \in \mathbb{Q} : x^2 < 2\}.$$

The set A *is bounded above, since, for example, if* $x > 2$ *then* $x^2 > 4$*, so* $x \notin A$*; hence* 2 *is an upper bound for* A*. We know that there is no rational number* r *such that* $r^2 = 2$*; that is,* $\sqrt{2} \notin \mathbb{Q}$*. So if we ask what the supremum of* A *is, within the ordered field* $\mathbb{Q}$*, we reach the conclusion that there is none. If we choose a rational number* r *less than* $\sqrt{2}$*, then it is not an upper bound for* A*, because there will be rational numbers between* r *and* $\sqrt{2}$ *(we will prove that fact later). If we choose a rational number* r *greater than* $\sqrt{2}$*, then* r *will be an upper bound for* A*, but it won't be a least upper bound for* A *because there will be rational numbers between* $\sqrt{2}$ *and* r *which will be smaller upper bounds for* A *than* r*. So* A *has no supremum in* $\mathbb{Q}$*.*

Completeness, and the Axioms for $\mathbb{R}$

The reason behind the last example is that the rational numbers are missing points that, in some sense, they should have, like $\sqrt{2}$. The property that distinguishes the real numbers from the rational numbers is that all non-empty, bounded above subsets of real numbers have least upper bounds. This means that the real numbers have no "gaps," no points missing. This will be an axiom for us, which we call the *completeness* axiom for $\mathbb{R}$.

Definition 9.0.10 *An ordered field* $(F, <)$ *is complete if every non-empty, bounded above subset of* F *has a least upper bound.*

Definition 9.0.11 $\mathbb{R}$ *is a complete ordered field. In more detail, we assume that there exists a set* $\mathbb{R}$ *with operations* $+, \cdot$ *and an ordering* $<$ *such that*

(i) $(\mathbb{R}, +, \cdot)$ *forms a field, i.e., axioms* $A1 - A4, M1 - M4$*, and* D *from Chapter 7 hold;*

(ii) $(\mathbb{R}, <)$ *forms and ordered field, i.e., axioms O1-O4 from Chapter 8 hold;*

and

(iii) $\mathbb{R}$ *is complete in the ordering* $<$*.*

These axioms for $\mathbb{R}$ are called the *Peano* axioms, named for the mathematician who formulated them. The purpose of these axioms is to have a clear set of assumptions, which serve as our starting point. Everything we prove should ultimately follow from the Peano axioms. If there is ever a dispute about a mathematical statement about $\mathbb{R}$, we can resolve it, in principle, by determining whether the statement follows from the Peano axioms. The Peano axioms put our mathematical work in a clear logical framework, as it should be.

However, we are going to depart a little from the full rigor of this plan, in order to deal with the natural numbers $\mathbb{N}$. It is possible to prove, based on Definition 9.0.11, that $\mathbb{R}$ contains a subset $\mathbb{N}$ which satisfies the axioms for $\mathbb{N}$ in Section 5. (See, for example, Munkres, *Topology: A First Course*, Ch. 1.4.) However, this approach is highly abstract (the natural numbers are defined as the intersection of all "inductive" sets containing 1), and tedious to check in detail. We will just assume that $\mathbb{R}$ contains the natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$ and that $\mathbb{N}$ satisfies N1-N5 from Section 5. In fact, we know that $1 \in \mathbb{R}$, since $\mathbb{R}$ is a field, hence $1 + 1 = 2$, $2 + 1 = 3$, etc., all belong to $\mathbb{R}$ by the properties of addition. The axiomatic properties N1-N5 of $\{1, 2, 3, \ldots\}$ are so reasonable that little rigor is lost in making these assumptions.

Although this approach works amazingly well, there are several issues to address at this time. In general, in mathematics, one wants to choose the simplest, smallest, and most reasonable set of axioms that one can. The field axioms seem natural, as do the order axioms O1-O4. The completeness axiom is less clear. It may seem like a big assumption, perhaps not believable. Perhaps it is assuming too much; how do we even know that the completeness axiom doesn't contradict the other axioms? Intuitively, one can think that if a non-empty set is bounded above and does not have a supremem in $\mathbb{R}$, that just means that a point is missing, and we should just add it to $\mathbb{R}$. If we do that with all possible missing points, we would call the result $\mathbb{R}$. (It's not really that simple, in fact, because once you add all of these points, you have a lot more bounded above subsets which must have suprema, etc.) It turns out that there is a way to construct a set, starting with the rational numbers, that forms a complete ordered field, and we then call that set $\mathbb{R}$. That construction is somewhat advanced, and we won't cover it here. (There are two well-known methods, one involving "Dedekind cuts," and one involving Cauchy sequences. In fact, one can go further, and using only the axioms of set theory, one can construct the natural numbers, then use them to construct the rationals, and then construct the real numbers. So if the axioms of the real numbers are contradictory, then the axioms of set theory must be contradictory; if that were the case, all of mathematics would be invalidated.) We will just assume the existence of the set $\mathbb{R}$ with the complete ordered field properties, having $\mathbb{N}$, satisfying N1-N5, as a subset.

The other question is whether we have assumed enough in the axioms for $\mathbb{R}$. In particular, could there be more than one complete ordered field? If so, we can't talk about "the" real numbers, and the answer to some questions might be different depending on which choice you make. It turns out, although we won't go into this point in detail, that there is only one complete ordered field up to an equivalence that means that if you have two complete ordered fields, one is the same as the other just with the names of the numbers changed. For example, if we use the numbers $0', 1', (3/4)', \pi' \ldots$ etc., instead of $0, 1, 3/4, \pi \ldots$, we haven't really created a different complete ordered field. The precise way to make this notion of equivalence precise is to use the notion of "isomorphism" from abstract algebra. If there are two complete ordered fields, say $\mathbb{R}_1$ (with additive and multiplicative identity elements 0_1 and 1_1 respectively) and $\mathbb{R}_2$ (with additive and multiplicative identity elements 0_2 and 1_2), then there is a $1-1$, onto map $T : \mathbb{R}_1 \to \mathbb{R}_2$ satisfying $T(0_1) = 0_2, T(1_1) = 1_2$ which preserves the operations (that is, $T(x) + T(y) = T(x + y)$ and $T(x)T(y) = T(xy)$), which means that corresponding elements add and multiply the same way, and T preserves the order: $x < y$ implies $T(x) < T(y)$). This means that $\mathbb{R}_2$ is the "same" in structure as $\mathbb{R}_1$ but with names x for points in $\mathbb{R}_1$ replaced with $T(x)$ for points in $\mathbb{R}_2$. Anything you know about $\mathbb{R}_1$ translates to the corresponding fact for $\mathbb{R}_2$, and vice versa, so $\mathbb{R}_1$ and $\mathbb{R}_2$ have equivalent structure. One way to think of this is that we have to have the numbers 0 and 1 in any field, by definition; then we must have $2 = 1 + 1, 3 = 2 + 1$, etc., by the existence of addition; thus we obtain $\mathbb{N}$. We can show that $\mathbb{N}$ satisfies the axioms for $\mathbb{N}$ from Section 5, thus validating proof by induction. Then the multiplication axioms force us to have $\mathbb{Q}$, and finally the completeness axiom forces us to fill in all the "gaps," giving us the irrational numbers as well. Then the order assumptions force us to stop with the rational and irrational numbers, rather than going to a larger set, like the complex numbers. So the axioms force $\mathbb{R}$ to be the real numbers we have always believed in, and nothing else. We won't prove any of these facts, because they take too long and aren't needed for what we do later; we just need to have the axioms, which we assume. Having the axioms, however, is essential for having a clear starting point for all of our future logical deductions.

The completeness axiom is quite subtle. Next we will consider some of its implications. Ultimately there is a logical straight line leading from the completeness axiom to the existence of limits of increasing, bounded above sequences, to compactness, to the existence of the Riemann integral of a continuous function on a closed interval, to the fundamental theorem of calculus. It took mathematicians many years after the development of calculus by Newton and Liebniz to fill in all of the gaps needed to make all of this

logically rigorous. Having that rigorous foundation then allows us to answer further, deeper questions and have confidence in our answers. The astounding number of unexpected discoveries and the resolution of longstanding open problems (often for problems where our intuition doesn't give any sense of what should be correct) in modern mathematics is a testament to the success of this procedure.

Chapter 10

Suprema and Infima

Based on the discussion so far, we have agreed to accept the existence of the real number system $\mathbb{R}$, which is characterized by being a complete ordered field, and we have assumed that $\mathbb{R}$ contains the set $\mathbb{N}$ satisfying the axioms for the natural numbers. The field axioms give the algebraic structure of $\mathbb{R}$, and the order axioms describe how the "$<$" relation behaves. The completeness axiom is more subtle; it says that any non-empty, bounded above subset of $\mathbb{R}$ has a least upper bound, or supremum (which is an element of $\mathbb{R}$, but not necessarily an element of the subset). Essentially, the completeness axiom guarantees that there are no points missing from the real line. For example, in this section we will show that there is a real number x satisfying $x^2 = 2$; so we are justified in talking about the number $\sqrt{2}$. Up until now our discussion has sometimes been intuitive, assuming things we believe to be true based on experience. But from now on, since we have explicit axioms, everything we do should follow logically from the axioms for $\mathbb{R}$ and $\mathbb{N}$.

The Greatest Lower Bound, or Infimum, of a Set

Recall that in Section 9, we defined what it means for a set to be bounded above or bounded below, but then we only discussed least upper bounds, for simplicity. It is also useful to consider the *greatest lower bound*, or *infimum*, of a set which is bounded below, as follows.

Definition 10.0.1 *Let $(F, <)$ be an ordered field and suppose $A \subseteq F$. We say that $s \in F$ is the greatest lower bound, or infimum, of A, if*

(i) s is a lower bound for A,
and
(ii) if $t \in F$ is another lower bound for A, then $s \geq t$.

If s is the greatest lower bound for A, we write $s = \inf A$.

The term "greatest lower bound" does not abbreviate (i.e., glb) any better than "least upper bound," which is the reason for the terminology "supremum" and "infimum."

If a set A has a greatest lower bound, then it is unique: if there were two greatest lower bounds s_1 and s_2 for A, then by (i), s_1 and s_2 are both lower bounds for S, and so by (ii), we would have $s_1 \geq s_2$ and $s_2 \geq s_1$, so $s_1 = s_2$.

The existence of the supremum of a non-empty bounded above set is part of our axioms for $\mathbb{R}$. Do we need to make a similar assumption that non-empty bounded below sets have a greatest lower bound? Fortunately, we do not need a separate assumption to guarantee the existence of infima; it follows from the existence of suprema, as the next result shows.

Lemma 10.0.2 *Suppose $A \subset \mathbb{R}$, $A \neq \emptyset$, and A is bounded below. Then a point s exists in $\mathbb{R}$ which is the greatest lower bound for A.*

Proof. Let

$$B = \{b \in \mathbb{R} : b \text{ is a lower bound for } A\}.$$

Then $B \neq \emptyset$ since A has a lower bound, by assumption. We claim that B is bounded above, in fact by any element $a \in A$ (at least one $a \in A$ exists because we assumed $A \neq \emptyset$). To verify this statement, suppose $a \in A$. If $b \in B$, then b is a lower bound for A, so $b \leq a$. This inequality $b \leq a$ holds for an arbitrary $b \in B$, hence for all $b \in B$, so a is an upper bound for B.

Since B is non-empty and bounded above, B has a least upper bound s, by the completeness axiom for $\mathbb{R}$. We claim that $s = \inf A$. To show that $s = \inf A$, we need to demonstrate properties (i) and (ii) in Definition 10.0.1.

(i) Let $a \in A$. We have just shown that a is an upper bound for B. But s is the least upper bound of B, so $s \leq a$ by property (ii) in the Definition 9.2 of the supremum. Since the inequality $s \leq a$ holds for an arbitrary $a \in A$ (hence for all $a \in A$), s is a lower bound for A.

(ii) Let t be a lower bound for a. Then $t \in B$. Since $s = \sup B$, s is an upper bound for B, so $t \leq s$. ■

So now we know that every non-empty bounded below subset of $\mathbb{R}$ has a greatest lower bound.

Recall (Definition 9.0.4) that a set $A \subseteq \mathbb{R}$ has a *minimum* if there exists $s \in A$ such that s is a lower bound for A. In this case, $s = \inf A$, because (1) s is a lower bound for A and (2) any other lower bound t for A must satisfy $t \leq s$ since $s \in A$.

The following lemma may seem so obvious that it doesn't need proof. But it turns out not to be obvious how to prove this lemma from the axioms we have. Since it is a statement about $\mathbb{N}$, one might expect to use induction. But there isn't any quantity n in the statement to induct on. The trick of the proof is to introduce an appropriate n and use generalized induction on that n.

Lemma 10.0.3 *Let A be a non-empty subset of $\mathbb{N}$. Then A has a minimum element $m \in A$. Hence $\inf A \in A$.*

Proof. For $n \in \mathbb{N}$, let P_n be the statement: If $A \subseteq \mathbb{N}$ and $n \in A$, then A has a minimum element $m \in A$. We prove P_n for all $n \in \mathbb{N}$ by induction. We first prove P_1. Assume $1 \in A$. Then 1 is a minimum element in A, since $A \subseteq \mathbb{N}$ so every element $k \in A$ satisfies $1 \leq k$. So P_1 holds.

Now suppose $n > 1$ and P_k is true for $1 \leq k \leq n$. We claim that P_{n+1} is true. Suppose $A \subseteq \mathbb{N}$ and $n + 1 \in A$. We consider two possibilities.

(i) If $A \cap \{1, 2, \ldots, n\} = \emptyset$, then $m = n + 1 \in A$ is the minimum element in A, since, if $k \in A$, then $k \geq n + 1$ since $k \neq 1, 2, \ldots, n$.

(ii) If $A \cap \{1, 2, \ldots, n\} \neq \emptyset$, let $k \in A \cap \{1, 2, \ldots, n\}$. Then $k \in A$ and $k \leq n$. By P_k, which holds by the induction assumption, A has a minimum element $m \in A$.

Hence in all possible cases, we have shown that P_{n+1} holds. Hence, by the generalized induction principle (Lemma 5.0.4), P_n holds for all $n \in \mathbb{N}$.

For the statement of the Lemma, if $A \subseteq \mathbb{N}$ and $A \neq \emptyset$, then there exists some $n \in \mathbb{N}$ such that $n \in A$. Then by P_n, A has a minimum element m. Then $m$must be the infimum of A, as noted above. In particular, $\inf A \in A$. ■

Lemma 10.0.3 is so natural that it is often used without comment in the following way. Suppose P_n is some property that depends on n, for $n \in \mathbb{N}$, and may or may not hold for each n (for example, P_n could be the statement that $n^2 > 17$). If we know that it holds for some $m \in \mathbb{N}$, we can select the smallest $k \in \mathbb{N}$ for which it holds. Formally, to apply Lemma 10.0.3, one should define $A = \{n \in \mathbb{N} : P_n \text{ holds}\}$; then $A \neq \emptyset$ since $m \in \mathbb{N}$, and hence A has a minimum. But in practice this part will be understood and we will just say: let n be the smallest natural number such that P_n holds.

A result similar to Lemma 10.0.3 holds for the supremum of subsets of $\mathbb{N}$ which are bounded above.

Lemma 10.0.4 *Let A be a non-empty subset of $\mathbb{N}$ such that A is bounded above. Then A has a maximum element $m \in A$. Hence $\sup A \in A$.*

Proof. Let $B = \{n \in \mathbb{N}$ such that n is an upper bound for $A\}$. Then $B \subseteq \mathbb{N}$ by definition and $B \neq \emptyset$ since B is bounded above. By Lemma 10.0.3, B has a minimum element m. We claim that m is the

maximum of A. Since $m \in B$, m is an upper bound for A; that is, we have $a \leq m$ for all $a \in A$. Since m is the minimum element of B, we know $m-1 \notin B$. Thus $m-1$ is not an upper bound for A. Therefore there exists $a \in A$ such that $m-1 < a$. So $m-1 < a \leq m$, and $a \in A \subseteq \mathbb{N}$. But the only natural number a satisfying $m-1 < a \leq m$ is $a = m$. So $m = a \in A$, and m is an upper bound for A, so m is the maximum of A. Therefore $m = \sup A \in A$. ■

Further Remarks on Sups and Infs

The proof of Lemma 10.0.2 (i) used an argument that is not difficult, but which will be used so often that it will be useful to formulate it as a general principle.

Remark 10.0.5 *Suppose $A \subseteq \mathbb{R}$ and $A \neq \emptyset$.*

(i) Suppose $a \leq b$ for all $a \in A$. Then $\sup A \leq b$. (That is, you can pass from an upper bound for every element in a set to the same upper bound for the supremum of the set in a $\leq$ inequality.)

(ii) Suppose $b \leq a$ for all $a \in A$. Then $b \leq \inf A$. (You can pass from a lower bound for every element of a set to the same lower bound for the infimum of the set in a $\leq$ inequality.)

The proof of (i) is just to note that b is an upper bound for A, so the least upper bound $\sup A$ (which exists because we just saw that A is bounded above, by b) satisfies $\sup A \leq b$ since $\sup A$ is the least upper bound (i.e., by (ii) of Definition 9.2). Similarly for (ii), where we have assumed that b is a lower bound for A, so we have that $\inf A$ satisfies $b \leq \inf A$, since $\inf A$ is the greatest lower bound of A (i.e., by (ii) in Definition 10.0.1).

In simplest terms, we can pass from the set to the supremum or infimum in a $\leq$ inequality. This fact is worth noting because it can be used operationally when working with inequalities in an efficient way, without repeating the steps of the proof every time. For example, in proving (i) in Lemma 10.0.2, we could now say: "we just showed that for $a \in A$, we have $b \leq a$ for every $b \in B$. Then by Remark 10.0.5 (i), $\sup B \leq a$. Therefore $s = \sup B$ is a lower bound for A."

The similar statement for a " $<$ " inequality is not true: you can't pass to the supremum or infimum in a $<$ inequality. For example, for the set $(-1, 1)$, we have $-1 < a < 1$ for all $a \in (-1, 1)$, but we do not have $\sup A < 1$ or $-1 < \inf A$, because $\sup A = 1$ and $\inf A = -1$.

Here is a typical exercise involving suprema; it says that the supremum of a subset is smaller than or equal to the supremum of any containing set..

Example 10.0.6 *Suppose $A \subseteq B \subseteq \mathbb{R}$ and $A \neq \emptyset$. If B is bounded above, prove that A is bounded above and $\sup A \leq \sup B$.*

Proof. Let $a \in A$. Note $\sup B$ exists because B is bounded above by assumption and $B \neq \emptyset$ since $A \neq \emptyset$ and $A \subseteq B$. Since $A \subseteq B$, and $\sup B$ is an upper bound for B, we have $a \leq \sup B$. Thus A is bounded above. Since $a \leq \sup B$ holds for an arbitrary $a \in A$, it holds for all $a \in A$. So we can use Remark 10.0.5 to pass to the supremum on the left side, concluding that $\sup A \leq \sup B$. ■

The following result gives a useful characterization of the supremum and infimum.

Lemma 10.0.7 *Suppose $A \subseteq \mathbb{R}, A \neq \emptyset$, and $s \in \mathbb{R}$.*

(i) Then $s = \sup A$ if and only if s satisfies both (1): s is an upper bound for A, and (2): for all $\epsilon > 0$, there exists $a \in A$ (a depending on ϵ) such that $a > s - \epsilon$,

(ii) Then $s = \inf A$ if and only if s satisfies both (1) s is a lower bound for A, and (2) for all $\epsilon > 0$, there exists $a \in A$ such that $a < s + \epsilon$.

Proof. We prove (i), leaving the proof of (ii), which is similar, as an exercise. To prove (i), first suppose $s = \sup A$. Then (1) holds by definition of the supremum of A. To prove (2), suppose $\epsilon > 0$. Since $s - \epsilon < s$,

then $s - \epsilon$ is not an upper bound for A (since s is $\leq$ any upper bound of A, by definition of the supremum). Therefore there exists $a \in A$ such that $a > s - \epsilon$.

To prove the other direction, suppose (1) and (2) holds. By (1) s is an upper bound for A. This is the first property in the definition of $\sup A$ in Definition 9.2. To prove the second property, suppose $t < s$. Let $\epsilon = s - t > 0$. Then $t = s - \epsilon$. By assumption (2), there exists $a \in A$ such that $a > s - \epsilon = t$, so t is not an upper bound for A. Hence all upper bounds for A must be greater than or equal to s, which is the seoond property in the definition of the supremum. Thus $s = \sup A$.

■

Lemma 10.0.7 says that there must be elements of the set A arbitrarily (that is, within ϵ for every $\epsilon > 0$, no matter how small) close to the least upper bound $\sup A$; otherwise, there would be a smaller upper bound than the least upper bound. Similarly, there must be an element of A arbitrarily close to the infimum of A.

Chapter 11

Consequences of the Completeness Axiom

The completeness axiom for $\mathbb{R}$ has surprisingly far-reaching consequences. In this section, we will consider several of these.

The Archimedean Property and the Density of the Rational Numbers

We begin by proving a statement that may seem so obvious as to not require a proof. However, we are committed to proving everything from just the axioms we assumed for the real numbers (i.e., that the real numbers form a complete ordered field). It is curious to see how the completeness axiom is used to prove the following fact.

Lemma 11.0.1 *The set $\mathbb{N}$ (the natural numbers) is not bounded above.*

PROOF. The proof is by contradiction. Suppose $\mathbb{N}$ is bounded above. By the completeness axiom, then, $\mathbb{N}$ has a least upper bound in $\mathbb{R}$, which we call s. Then s is an upper bound for $\mathbb{N}$, so $n \leq s$ for all $n \in \mathbb{N}$. But for $n \in \mathbb{N}$, we also have $n+1 \in \mathbb{N}$, so $n+1 \leq s$, again for all $n \in \mathbb{N}$. That is, $n \leq s-1$, for all $n \in \mathbb{N}$. Thus $s-1$ is an upper bound for $\mathbb{N}$, and $s-1 < s = \sup \mathbb{N}$, which is impossible because s is the least upper bound for $\mathbb{N}$. This contradiction shows that $\mathbb{N}$ is not bounded above. ■

The next consequence says that you can get a big thing from enough small things; for example, it is possible to get rich if you save enough pennies.

Corollary 11.0.2 *(Archimedean Property) Suppose $x, y \in \mathbb{R}$ with $x > 0$. Then there exists $n \in \mathbb{N}$ such that $nx > y$.*

PROOF. Since $\frac{y}{x}$ is not an upper bound for $\mathbb{N}$ (since there isn't one, by the previous lemma), there exists $n \in \mathbb{N}$ such that $n > \frac{y}{x}$. Since $x > 0$, multiplying both sides by x preserves the inequality, so $nx > y$. ■

The next fact also seems obvious, but we need a proof based on the axioms for $\mathbb{R}$. It says that if two real numbers are a distance greater than one apart, then there is an integer in-between them.

Lemma 11.0.3 *Suppose $a, b \in \mathbb{R}$ with $a < b$ and $b - a > 1$. Then there exists $m \in \mathbb{Z}$ such that $a < m < b$.*

PROOF. First suppose $a \geq 0$. Since a is not an upper bound for $\mathbb{N}$ (by Lemma 11.0.1, there is no upper bound for $\mathbb{N}$), there are natural numbers greater than a. Then by Lemma 10.0.3, we can choose the smallest natural number such that $m > a$. Then $m - 1 \leq a$, since m is the smallest natural number greater than a. Hence $m \leq a + 1 < b$ (the last inequallity is by the assumption $b - a > 1$). So $a < m < b$, with $m \in \mathbb{N} \subseteq \mathbb{Z}$.

Now suppose $a < 0$ and $b \leq 0$. Then $0 \leq -b < -a$ and $-a - (-b) = b - a > 1$. We apply the first case with a replaced by $-b$ and b replaced by $-a$, to obtain $k \in \mathbb{N}$ such that $-b < k < -a$. Then $a < -k < b$, so the required conclusion holds for $m = -k \in \mathbb{Z}$.

The only other possibility is that $a < 0$ and $b > 0$, or $a < 0 < b$, in which case $m = 0$ satisfies the requirements. ■

As simple as these last two results may seem, they lead to the next fact which is not so trivial. It says that between any two distinct real numbers, there is a rational number.

Theorem 11.0.4 *(Density of $\mathbb{Q}$ in $\mathbb{R}$) Suppose $x, y \in \mathbb{R}$ with $x < y$. Then there exists $r \in \mathbb{Q}$ such that $x < r < y$.*

PROOF. Since $y - x > 0$, we can apply the Archimedean Property, i.e., Corollary 11.0.2, with x replaced by $y - x$ and y replaced by 1, to obtain that there exists $n \in \mathbb{N}$ such that $n(y - x) > 1$, or $ny > 1 + nx$. By Lemma 11.0.3, there exists $m \in \mathbb{Z}$ such that $nx < m < ny$. Since $n \in \mathbb{N}$, we can divide by n and maintain the inequality, yielding $x < \frac{m}{n} < y$. Since $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, we have $r = \frac{m}{n} \in \mathbb{Q}$ and $x < r < y$. ■

In other words, between any two distinct real numbers, there is a rational number. Later we will see that there is also an irrational number between any two distinct real numbers.

The Existence of $\sqrt{2}$ and the Density of the Irrational Numbers

Next we resolve the issue raised in Section 6, where we showed that there is no rational number r satisfying $r^2 = 2$. We hoped that by replacing $\mathbb{Q}$ by a complete ordered field, namely $\mathbb{R}$, we would obtain a real number s such that $s^2 = 2$. We are ready to verify that hope, after a very simple lemma.

Lemma 11.0.5 *Suppose $a, b \in \mathbb{R}$ with $a > 0$ and $b > 0$. Then $a < b$ if and only if $a^2 < b^2$.*

PROOF. First suppose $a < b$. Multiplying the inequality $a < b$ by a (which is positive, by assumption, so the inequality is preserved), we get $a^2 < ab$. Similarly, multiplying the inequality $a < b$ by b gives $ab < b^2$. Using the transitive property (Definition 8.1 (ii))of the order, we obtain $a^2 < b^2$.

The converse implication is that $a^2 < b^2 \Longrightarrow a < b$. The contrapositive of that implication (which is equivalent) is that $b \leq a \Longrightarrow b^2 \leq a^2$. To prove this contrapositive statement, suppose $b \leq a$. If $b = a$, then $b^2 = a^2$ so $b^2 \leq a^2$. If $b < a$, then by the first direction with a and b interchanged, we have $b^2 < a^2$, so $b^2 \leq a^2$. ■

Theorem 11.0.6 *There exists $s \in \mathbb{R}$ such that $s^2 = 2$.*

PROOF. Let $A = \{x \in \mathbb{R} : x^2 < 2\}$. Then $A \neq \emptyset$ since, for example, $0 \in A$. Also, A is bounded above, for example, by 3: if $x \in A$ satisfies $x \leq 0$, then $x < 3$, and if $x \in A$ satisfies $x > 0$ then $x^2 < 2 < 9 = 3^2$, so by the previous lemma, $x < 3$. Since A is non-empty and bounded above, A has a supremum $s \in \mathbb{R}$. We claim that $s^2 = 2$. To prove that $s^2 = 2$, we will consider the possibilities $s^2 > 2$ and $s^2 < 2$ and reach a contradiction in each case. By the trichotemy property (O1 in Definiton 8.0.1), we conclude that $s^2 = 2$.

First suppose that $s^2 > 2$. The idea of the proof is to show that for some sufficiently large $n \in \mathbb{N}$, the number $s - \frac{1}{n}$ is an upper bound for A, whihc contradicts the fact that s is the least upper bound. Since $s^2 > 2$ by assumption, the number $\frac{2s}{s^2-2}$ is defined and positive. Since it is not an upper bound for $\mathbb{N}$ (since there isn't any such thing), we can find $n \in \mathbb{N}$ such that $n > \frac{2s}{s^2-2}$. Then $s^2 - 2 > \frac{2s}{n}$, hence $s^2 - \frac{2s}{n} > 2$. Therefore

$$\left(s - \frac{1}{n}\right)^2 = s^2 - \frac{2s}{n} + \frac{1}{n^2} > s^2 - \frac{2s}{n} > 2.$$

Then for all $x \in A$, we have $x^2 < 2 < (s - \frac{1}{n})^2$. Note that since $1.1 \in A$ and s is an upper bound for A. It follows that $s > 1$, hence $s - \frac{1}{n} > 0$. Thus by Lemma 11.0.5, we get $x < s - \frac{1}{n}$ (if $x \leq 0$, we automatically have $x \leq 0 < s - \frac{1}{n}$, and if $x > 0$ we apply Lemma 11.0.5). Hence $s - \frac{1}{n}$ is an upper bound for A which is smaller than s, the least upper bound for A, which is a contradiction. Thus $s^2 > 2$ is impossible.

Now suppose $s^2 < 2$. Let $n \in \mathbb{N}$ satisfy $n > \frac{4s}{2-s^2}$ (note $\frac{4s}{2-s^2} > 0$ since $s^2 < 2$) and $n > \frac{1}{2s}$ (such n exists because $\max(\frac{4s}{2-s^2}, \frac{1}{2s})$ is not an upper bound for $\mathbb{N}$). Then $\frac{4s}{n} < 2 - s^2$, since $n > \frac{4s}{2-s^2}$. Also $\frac{1}{n} < 2s$ so $\frac{1}{n^2} < \frac{2s}{n}$, since $n > \frac{1}{2s}$. Therefore

$$\left(s + \frac{1}{n}\right)^2 = s^2 + \frac{2s}{n} + \frac{1}{n^2} < s^2 + \frac{2s}{n} + \frac{2s}{n} = s^2 + \frac{4s}{n} < 2.$$

Hence $s + \frac{1}{n} \in A$, which is impossible because s is an upper bound for A. So $s^2 < 2$ is impossible.

Hence $s^2 = 2$. ■

The number s obtained in the proof of Theorem 11.0.6 is positive. We call it $\sqrt{2}$. Note that $-\sqrt{2}$ is another real number whose square is 2.

The last proof can be modified to show the existence $a^{1/n}$ for any $a > 0$ and $n \in \mathbb{N}$, but the computations become more awkward as n increases. Once we have discussed limits of sequences, we can provide a more elegant solution for the existence of arbitrary n^{th} roors of positive numbers. Our point now is to demonstrate how the least upper bound axiom guarantees that a number like $\sqrt{2}$ exists in $\mathbb{R}$; without $\sqrt{2}$, $\mathbb{R}$ would not satisfy the completeness axiom.

We can now give a simple proof of the following, which is complementary to Theorem 11.0.4.

Theorem 11.0.7 *(Density of the irrational numbers in $\mathbb{R}$) Suppose $x, y \in \mathbb{R}$ with $x < y$. Then there exists $t \in \mathbb{R} \setminus \mathbb{Q}$ such that $x < t < y$.*

Proof. Since $x < y$, we have $x - \sqrt{2} < y - \sqrt{2}$. By Theorem 11.0.4, there exists $r \in \mathbb{Q}$ such that $x - \sqrt{2} < r < y - \sqrt{2}$. Hence $x < r + \sqrt{2} < y$. Let $t = r + \sqrt{2}$. If $t \in \mathbb{Q}$, then $\sqrt{2} = t - r \in \mathbb{Q}$, which we know to be false by Theorem 6.3. Therefore $t \in \mathbb{R} \setminus \mathbb{Q}$ and $x < t < y$. ■

The Nested Interval Property

We now investigate another consequence of completeness, the *Nested Interval Property*. Its importance is not clear now, but, for example, later it will be used to show that $\mathbb{R}$ is 'uncountable," a concept that will be defined in the next section.

Theorem 11.0.8 *(Nested Interval Property) Suppose that for each $n \in \mathbb{N}$, we have $a_n, b_n \in \mathbb{R}$ with $a_n \leq b_n$. Let I_n be the interval $I_n = [a_n, b_n]$ (if $a_n = b_n$ then $I_n = \{a_n\}$). Suppose that $I_{n+1} \subseteq I_n$ for each n (or, equivalently, $a_n \leq a_{n+1}$ and $b_{n+1} \leq b_n$). Then $\cap_{n=1}^{\infty} I_n \neq \emptyset$.*

Proof. Let $A = \{a_n : n \in \mathbb{N}\} = \{a_1, a_2, a_3, \ldots\}$. Since $a_n \leq a_{n+1}$, the terms a_n are non-decreasing. Also, since $b_{n+1} \leq b_n$, the terms b_n are non-increasing. Suppose $k, n \in \mathbb{N}$. If $k \leq n$, then $a_k \leq a_n \leq b_n$. Also, if $k > n$, then $a_k \leq b_k \leq b_n$. Hence in all cases we obtain

$$a_k \leq b_n,$$

for all $k, n \in \mathbb{N}$. In particular, A is bounded above (by any b_n) and non-empty, so $s = \sup A$ exists. Then for each $n \in \mathbb{N}$, we have

(i) $a_n \leq s$, since s is an upper bound for A, and

(ii) $s \leq b_n$, since we can pass to the supremum over k in the left side of the inequality $a_k \leq b_n$ (by Remark 10.5 (i)).

By (i) and (ii), we have $s \in [a_n, b_n] = I_n$, for all $n \in \mathbb{N}$. Hence $s \in \cap_{i=1}^{\infty} I_n$, so in particular $\cap_{i=1}^{\infty} I_n \neq \emptyset$.
■

The intervals I_n in Theorem 11.0.8 are called *nested* because we have $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$. Thus the intervals I_n may shrink down. The Nested Interval Property says that they can not shrink down to the empty set; their intersection is non-empty. A similar statement for open intervals (a, b) is false; for example, $\cap_{n=1}^{\infty}(0, \frac{1}{n}) = \emptyset$.

Chapter 12

Cardinality, Countable and Uncountable Sets

Cardinality of Sets

For finite sets A and B, the statement "A has more elements than B" is perfectly clear. For infinite sets, this issue is much more subtle, as we can see from the following example.

Example 12.0.1 *Let* $2\mathbb{N}$ *denote the set* $\{2n : n \in \mathbb{N}\} = \{2, 4, 6, 8, \ldots\}$. *Consider the question: does* $2\mathbb{N}$ *have more or less elements than the set* $\mathbb{N} = \{1, 2, 3, \ldots\}$*?*

At first consideration, the answer seems obvious: since $2\mathbb{N} \subseteq \mathbb{N}$, we think that $\mathbb{N}$ has more elements than $2\mathbb{N}$; in fact, strictly more because there are elements such as $1, 3, 5$, etc., which belong to $\mathbb{N}$ but not to $2\mathbb{N}$. However, let's change the notation for $\mathbb{N}$ and reconsider the question. Let's rename the element 1 as $\dot{4}$. Let's rename 2 as $\dot{8}$, and so on: 3 becomes $1\dot{2}$, and in general, n becomes $4\dot{n}$. With this renaming, we have

$$\mathbb{N} = \{\dot{4}, \dot{8}, 1\dot{2}, \ldots.\}$$

Now the set $\mathbb{N}$ looks essentially the same the set $4\mathbb{N} = \{4n : n \in \mathbb{N}\} = \{4, 8, 12, \ldots\}$. But since $4\mathbb{N} \subseteq 2\mathbb{N}$ and $2\mathbb{N}$ has elements like $2, 6$, etc., the same reasoning that convinced us that $\mathbb{N}$ has more elements than $\mathbb{N}$ now convinces us that $4\mathbb{N}$, which is effectively the same as $\mathbb{N}$ under a renaming, has less elements than $2\mathbb{N}$. But just changing the names of the elements shouldn't cause a set to go from having more elements than another set to having less elements.

What is going on? Example 12.0.1 shows us that the concept of "having more elements," or being "bigger," is not a clear or precise concept when it comes to infinite sets. However, there is another idea which is mathematically precise, which can be applied to infinite sets, namely the concept of whether a $1-1$ correspondence, or bijection, exists between the sets.

Definition 12.0.2 *For two sets* A *and* B*, we say* $A \approx B$ *if there exists a bijection* $f : A \to B$*. If* $A \approx B$*, we say that* A *and* B *have the same cardinality.*

This definition may seem tricky because it says $A \approx B$ if a bijection exists: you may wonder how you can ever tell if a bijection exists. If you can define a map and prove that it is $1-1$ and onto, then you have shown that the sets have the same cardinality. But it may seem impossible to ever show that two sets don't have the same cardinality. How can you ever show that no bijection exists? But we will see that it is possible to show in some cases that no bijection exists.

Example 12.0.3 $\mathbb{N} \approx 2\mathbb{N}$

Proof. Define $f : \mathbb{N} \to 2\mathbb{N}$ by $f(n) = 2n$. First note that for $n \in \mathbb{N}$, we have $f(n) = 2n \in 2\mathbb{N}$, so f really is a map from $\mathbb{N}$ to $2\mathbb{N}$. To show that f is $1-1$, suppose $n_1, n_2 \in \mathbb{N}$ and $f(n_1) = f(n_2)$. That means that $2n_1 = 2n_2$, so dividing by 2 implies that $n_1 = n_2$. Hence f is $1-1$. (See Section 4C for a reminder on

how to show a function is $1-1$ or onto.) To show that f is onto, let $y \in 2\mathbb{N}$. By definition of $2\mathbb{N}$, that means $y = 2k$ for some $k \in \mathbb{N}$. So $f(k) = 2k = y$. Since $y \in 2\mathbb{N}$ was arbitrary, f is onto. Hence f is a bijection, so $\mathbb{N} \approx 2\mathbb{N}$. ■

The next example shows that all intervals $(a, b) \subseteq \mathbb{R}$ have the same cardinality.

Example 12.0.4 *Let $a, b \in \mathbb{R}$ with $a < b$. Then $(0, 1) \approx (a, b)$.*

Proof. Define $f : (0, 1) \to (a, b)$ by $f(x) = a + x(b - a)$. Note that since $b - a > 0$ and $x > 0$, we have $a + x(b - a) > a$. Also since $x < 1$, we have $x(b - a) < b - a$, so $a + x(b - a) < a + (b - a) = b$. Thus for $x \in (0, 1)$, we have $a < f(x) < b$, so f does map $(0, 1)$ into (a, b). To prove that f is $1 - 1$, suppose $x_1, x_2 \in (0, 1)$ and $f(x_1) = f(x_2)$. Then $a + x_1(b - a) = a + x_2(b - a)$. Hence $x_1(b - a) = x_2(b - a)$, and therefore (since $b - a \neq 0$), $x_1 = x_2$. Therefore f is $1 - 1$. To show f is onto, suppose $y \in (a, b)$. Then $a < y < b$, so, subtracting a from each term, $0 < y - a < b - a$. Hence $0 < \frac{y-a}{b-a} < 1$. So $x = \frac{y-a}{b-a} \in (0, 1)$ and

$$f(x) = a + x(b - a) = a + \frac{y - a}{b - a} \cdot (b - a) = a + y - a = y.$$

Since $y \in (a, b)$ was arbitrary, f is onto. ■

This last result may seem counter-intuitive, since, for example, intervals of length larger than 1 seem bigger than intervals of length 1. But the proof shows that the concept of length can not be captured just using the notion of bijection. The proof can be thought of as taking a rubber band strip of length 1, which at rest would exactly cover the interval $(0, 1)$, and moving the rubber band strip so that its left end sits fixed at a, and then stretching the rubber band until the right end hits b (in the case $b - a > 1$). Intuitively, the stretching doesn't change the number of points in the strip.

The following proposition means that the relation "$\approx$" on sets is an equivalence relation, if you are familiar with that term.

Proposition 12.0.5 *For any sets A, B, and C,*

(i) $A \approx A$,
(ii) if $A \approx B$ then $B \approx A$,
(iii) if $A \approx B$ and $B \approx C$, then $A \approx C$.

The proof only uses the properties of $1-1$ and onto functions, and is left as an exercise. We will regularly use these properties, for example not distinguishing between $A \approx B$ and $B \approx A$, since they are equivalent.

The notion of cardinality allows us to divide sets into certain classes, as follows.

Definition 12.0.6 *Let A be a set. We say*

(i) A is finite if $A = \emptyset$ or $A \approx \{1, 2, 3, \ldots, n\}$ for some $n \in \mathbb{N}$;
(ii) A is countably infinite if $A \approx \mathbb{N}$;
(iii) A is countable if A is finite or countably infinite;
(iv) A is uncountable if A is not countable.

At the moment, we have no reason to believe that any uncountable sets exist, but soon we will see that they do. In case (i), if $A \approx \{1, 2, 3, \ldots, n\}$, we say that A has n elements.

Let's think intuitively about countable sets a bit more. If X is countably infinite, then there exists a bijection $f : \mathbb{N} \to X$. Let $x_1 = f(1), x_2 = f(2), \ldots$, etc., with $x_k = f(k)$, for each $k \in \mathbb{N}$. Note that if $k, j \in \mathbb{N}$ with $j \neq k$, then $x_j = f(j) \neq f(k) = x_k$, so the elements x_k of X are all distinct. Also, since f is onto, every element of X is $f(k) = x_k$ for some $k \in \mathbb{N}$. Altogether then, we see that

$$X = \{x_1, x_2, x_3, \ldots, \} = \{x_k\}_{k=1}^{\infty}.$$

On the other hand, if X is finite and non-empty, then there is a bijection between X and $\{1, 2, 3, \ldots, n\}$ for some $n \in \mathbb{N}$, so by the same process we can write

$$X = \{x_1, x_2, \ldots, x_n\}.$$

Thus countable sets are sets whose elements can be listed, in such a way that each element is reached at some point in the list. Sometimes countable sets are called *denumerable*, because the elements of the set can be enumerated in this way.

You might think that any set is countable: you just start with one element, call it x_1, then another element, call it x_2, and continue, either exhausting the set in finitely many steps if the set is finite, or continuing, getting a infinite listing of the set, so that the set is countably infinite. However, if you just take elements at random, one by one, there is no guarantee that every element of the set will eventually appear in the list. In fact, below we will give an example showing that a certain set is uncountable.

In general, for a set A, we denote by $\mathcal{P}(A)$ the set of all subsets of A. For example, if $A = \{1, 2, 3\}$, then

$$\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

We will show that $\mathcal{P}(\mathbb{N})$, the set of all subsets of $\mathbb{N}$, is uncountable. The proof shows more generally that any function $f : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ is not onto. Although the proof can be stated briefly, it is difficult to understand the proof without first considering an example, which we do as follows. Suppose $f : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ satisfies

$f(1) = A_1 = \{1, 2, 6, 24\}$,
$f(2) = A_2 = \{3, 5, 7, 9, 11, \ldots\}$,
$f(3) = A_3 = \{3, 9, 27, 81, \ldots\}$,
$f(4) = A_4 = \{1, 7\}$,

and so on, defining $A_k = f(k)$ for each $k \in \mathbb{N}$. We claim that such a function can't really be onto, that it must miss some elements of $\mathcal{P}(\mathbb{N})$. How can we show that, without knowing f in general? We will exhibit a set $A \subseteq \mathbb{N}$ which cannot be $A_k = f(k)$ for any $k \in \mathbb{N}$. We will construct the set A in the following manner. We don't want A to be A_1, so we will make sure that A disagrees with A_1 with regard to the element 1. That is, in our example, since $1 \in A_1$, we decide that $1 \notin A$, which is already enough to guarantee $A \neq A_1$. Next, to make sure $A \neq A_2$, we choose A so that A and A_2 disagree with regard to the element 2. In this example, $2 \notin A_2$, so we select $2 \in A$. We keep going in this way: we choose $3 \notin A$ because $3 \in A_3$, which guarantees that $A \neq A_3$. We choose $4 \in A$ since $4 \notin A_4$. We proceed through all $k \in \mathbb{N}$, so that $A \neq A_k$, for all k. In our example, $A = \{2, 4, \ldots\}$. Our principle for constructing A is that $k \in A$ if and only if $k \notin A_k$. That is, we define

$$A = \{k \in \mathbb{N} : k \notin A_k\}.$$

For each $k \in \mathbb{N}$, we have $A \neq A_k$, since if $k \in A_k$, then $k \notin A$, by definition of A, and if $k \notin A_k$, then $k \in A$, again by definition of A.

This argument may seem vulnerable, because it only exhibits one set A which is not in the range of f. One might try to correct that problem by defining a new map $g : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ by letting $g(1) = A$ and then $g(k) = f(k - 1)$ for $k \geq 2$. In effect we just add A to the list by putting it at the top. But then we could run our argument above with g in place of f, to find another subset of $\mathbb{N}$ which is not in the range of g. We could repeat this process, and obtain yet another missing set. In particular, then, there must be a lot of subsets of $\mathbb{N}$ which are not in the range of any $f : \mathbb{N} \to \mathcal{P}(\mathbb{N})$.

Here is the formal statement and proof.

Proposition 12.0.7 *$\mathcal{P}(\mathbb{N})$ is uncountable.*

Proof. We will show that if $f : \{1, 2, 3, \ldots, n\} \to \mathcal{P}(\mathbb{N})$ or $f : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ is a function, then f is not onto. If so, then there is no bijection $f : \{1, 2, 3, \ldots, n\} \to \mathcal{P}(\mathbb{N})$ or $f : \mathbb{N} \to \mathcal{P}(\mathbb{N})$, so $\mathcal{P}(\mathbb{N})$ is not finite (since $\mathcal{P}(\mathbb{N})$ is not empty) or countably infinite. Hence $\mathcal{P}(\mathbb{N})$ is uncountable.

Suppose first that $f : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ is a function. Let $A_k = f(k) \in \mathcal{P}(\mathbb{N})$, for each $k \in \mathbb{N}$. Let $A = \{k \in \mathbb{N} : k \notin A_k\}$. Let $k \in \mathbb{N}$. If $k \in A_k$, then by definition $k \notin A$, so $A \neq A_k$. On the other hand, if $k \notin A_k$, then by definition $k \in A$, so again $A \neq A_k$. Thus for each $k \in \mathbb{N}$, we have shown $A \neq A_k = f(k)$. Hence f is not onto.

Now suppose $f : \{1, 2, 3, \ldots, n\} \to \mathcal{P}(\mathbb{N})$ for some $k \in \mathbb{N}$. Define $g : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ by letting $g(k) = f(k)$ for $1 \leq k \leq n$ and $g(k) = f(1)$ for all $k > n$. Note that the range of g (i.e., $g(\mathbb{N})$) is the same as the range of f (i.e., $f(\{1, 2, \ldots, n\})$). By the previous observation, g is not onto, so f is not onto either. ∎

We could have handled the second part (about $\{1, 2, 3, \ldots, n\}$) by the same argument as for $\mathbb{N}$, but it seemed easier to argue as we did. The previous proposition can be generalized to show that for any set A, there is no onto function $f : A \to \mathcal{P}(A)$, by essentially the same argument (which we leave to the reader). So uncountable sets exist. This raises some natural questions based on our earlier work: (i) is $\mathbb{Q}$ countable or uncountable? and (ii) is $\mathbb{R}$ countable of uncountable? We could answer the second question with what we know now, but for the first we need to learn a bit more about countable sets.

Countable Sets

It seems reasonable that a countable set cannot have an uncountable subset.

Theorem 12.0.8 *Suppose X is countable and $A \subseteq X$. Then A is countable.*

Proof. If $A = \emptyset$ or A is finite, then A is countable. Now suppose A is infinite. Since $A \subseteq X$, it follows that X is infinite (i.e., if X is finite, say X has n elements, one can prove by induction on n that any subset of X has at most m elements for some $m \leq n$, and so can't be infinite). Since X is countably infinite, there exists a bijection $f : \mathbb{N} \to X$. Letting $x_k = f(k)$, then since f is onto,

$$X = \{x_1, x_2, x_3, \ldots, \} = \{x_k\}_{k\in\mathbb{N}}.$$

(This observation is just the statement, made earlier, that X can be listed.) Note that $x_{j_1} \neq x_{j_2}$ if $j_1 \neq j_2$ since f is $1-1$.

The idea of this proof is to make a sublist of the elements of A and enumerate them by their position on the sublist. Let

$$C = \{k \in \mathbb{N} : x_k \in A\}.$$

Note that C is infinite since A is assumed to be infinite. Using the fact that any nonempty subset of $\mathbb{N}$ has a smallest element (Lemma 10.0.3), let

$$k_1 = \min C, \text{ and let } a_1 = x_{k_1}.$$

Since C is infinite, there must be elements $k \in C$ satisfying $k > k_1$. Let

$$k_2 = \min\{k \in C : k > k_1\}, \text{and let } a_2 = x_{k_2}.$$

Continue in this way, inductively, to define a sequence $\{k_j\}_{j\in\mathbb{N}}$ of natural numbers and a sequence $\{a_j\}_{j\in\mathbb{N}}$ of elements of A such that

$$k_j = \min\{k \in C : k > k_{j-1}\}, \text{and } a_j = x_{k_j} \tag{12.1}$$

for all $j \in \mathbb{N}$ (where we set $k_0 = 0$ to handle the case $j = 1$). To verify the induction, given k_j, note that because C is infinite, the set $\{k \in C : k > k_j\}$ is non-empty, and hence has a smallest element. Define

$$k_{j+1} = \min\{k \in C : k > k_j\}, \text{and let } a_{j+1} = x_{k_{j+1}}.$$

By our definition, then, at the j^{th} stage, we have $k_{j+1} > k_j$, and $k_{j+1} \in C$, which means that $a_{j+1} = x_{k_{j+1}} \in A$. Hence (12.1) holds for all j by induction. Since $a_1 \in A$, we conclude also by induction that $a_j \in A$ for all $j \in \mathbb{N}$. We have obtained

$$(a_1, a_2, \ldots, a_j, \ldots) = (a_j)_{j\in\mathbb{N}},$$

with $a_j \in A$ for all $j \in \mathbb{N}$. Define $g : \mathbb{N} \to A$ by

$$g(j) = a_j.$$

We claim that g is a bijection. To prove that g is $1-1$, suppose $j_1, j_2 \in \mathbb{N}$ and $g(j_1) = g(j_2)$. By definition of g, that means that $a_{j_1} = a_{j_2}$, which, by definition of the a_j's, means that $x_{j_1} = x_{j_2}$. We noted earlier that if $j_1 \neq j_2$, then $x_{j_1} \neq x_{j_2}$, so we conclude $j_1 = j_2$. Hence g is $1-1$.

Before proving that g is onto, we observe that $k_j \geq j$ for each $j \in \mathbb{N}$. We prove this claim inductively. It holds for $j = 1$ because $k_1 \in \mathbb{N}$, hence $k_1 \geq 1$. Now suppose $k_j \geq j$. Then $k_{j+1} > k_j \geq j$, and $k_{j+1} \in \mathbb{N}$, so $k_{j+1} \geq j + 1$. This establishes the inductive step and hence the calim.

To prove g is onto, let $a \in A$. Since $A \subseteq X$, and $f : \mathbb{N} \to X$ defined above is onto, we conclude that $a = f(\ell) = x_\ell$, for some $\ell \in \mathbb{N}$. Let

$$m = \min\{j \in \mathbb{N} : k_j \geq \ell\}.$$

The set $\{j \in \mathbb{N}; k_j \geq \ell\}$ has a minimum since it is a subset of $\mathbb{N}$ which is nonempty since, for $j > \ell$, we have $k_j \geq j > \ell$. Since a subset of $\mathbb{N}$ has a minimum, which is in the set by the definition of minimum, we have $k_m \geq \ell$, and since $m - 1$ is not in the set, we have $k_{m-1} < \ell$. In particular, then, since also $x_\ell \in A$, we have

$$\ell \in \{k \in C : k > k_{m-1}\}.$$

Recall the the definition of k_m is $k_m = \min\{k \in C : k > k_{m-1}\}$. Therefore $k_m \leq \ell$, since the minimum of a set is less than or equal to any element of the set. Hence we have both $k_m \geq \ell$ and $k_m \leq \ell$. Therefore $k_m = \ell$, so

$$a = x_\ell = x_{k_m} = a_m = g(m).$$

Therfore g is onto. Thus $g : \mathbb{N} \to A$ is a bijection, so A is countable. ■

Theorem 12.0.8 has the following useful consequence.

Corollary 12.0.9 *Suppose A is a non-empty set. Then A is countable if and only if there exists a $1-1$ function $f : A \to \mathbb{N}$.*

Proof. First suppose that A is countable. By definition, either $A \approx \{1, 2, \ldots, n\}$ or $A \approx \mathbb{N}$. If $A \approx \mathbb{N}$, then there exists a bijection, and in particular a $1-1$ map, from A to $\mathbb{N}$. If $A \approx \{1, 2, \ldots, n\}$, then there is a bijection $g : A \to \{1, 2, \ldots, n\}$. Define $h : \{1, 2, \ldots, n\} \to \mathbb{N}$ by $h(k) = k$ for $1 \leq k \leq n$. Then h is $1-1$. Hence $f = h \circ g : A \to \mathbb{N}$ is $1-1$, since f is the compostion of two 1-1 functions (Proposition 4.3 (i)).

Now suppose that there exists a $1-1$ function $f : A \to \mathbb{N}$. Let $B = f(A) \subseteq \mathbb{N}$. By Theorem 12.0.8, B is countable. Define a function $g : A \to B = f(A)$ by $g(a) = f(a)$ for all $a \in A$. Then g is $1-1$ since f is assumed to be $1-1$, and g is onto since $B = f(A)$ by definition. Therefore $A \approx B$. So A is countable since B is countable. (At the end we are using the properties of $\approx$ from Proposition 12.0.5: if $B \approx \{1, 2, \ldots, n\}$ then since $A \approx B$, then we obtain $A \approx \{1, 2, \ldots, n\}$ and so A is finite. Similarly if $B \approx \mathbb{N}$ then $A \approx \mathbb{N}$ and A is countably infinite.) ■

Of course the empty set is countable, but we excluded that case from the last corollary because it is not immediately clear what a function defined on the empty set would be.

The last corollary makes it easier to show that a set A is countable. To do so using the definition of countability (Definition 12.0.6) we would need to construct a bijection $f : A \to \{1, 2, \ldots, n\}$ or $f : A \to \mathbb{N}$. But by Corollary 12.0.9, we only need to find $f : A \to \mathbb{N}$ which is $1-1$, not necessarily onto. This fact is particular to countable sets; for uncountable sets B and C, to show $B \approx C$ you still need to construct a bijection $f : B \to C$. We exploit the advantage given by Corollary 12.0.9 in the next result.

Theorem 12.0.10 $\mathbb{N} \times \mathbb{N}$ *is countably infinite.*

Proof. Define $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by $f(n, m) = 2^n 3^m$. We claim that f is $1-1$. To prove this claim, suppose $n_1, m_1, n_1, m_1 \in \mathbb{N}$ and $f(n_1, m_1) = f(n_2, m_2)$. By definition of f, that means that $2^{n_1} 3^{m_1} = 2^{n_2} 3^{m_2}$. Let's consider 3 cases:

(i) Suppose $n_1 > n_2$. Then dividing the equation $2^{n_1} 3^{m_1} = 2^{n_2} 3^{m_2}$ on both sides by $2^{n_2} 3^{m_1}$ gives $2^{n_1 - n_2} = 3^{m_2 - m_1}$. Since $n_1 > n_2$, we have $2^{n_1 - n_2}$ is a positive integer, hence so is $3^{m_2 - m_1}$. Then $2^{n_1 - n_2}$ is even, but $3^{m_2 - m_1}$ is odd, contradicting their equality. Thus $n_1 > n_2$ is not possible.

(ii) Suppose $n_2 > n_1$. By reversing the roles of n_1 and n_2 in the argument for (i), we see that this case is not possible either.

(iii) Thus $n_1 = n_2$. Then $3^{m_2 - m_1} = 2^{n_1 - n_2} = 1$, so $m_2 = m_1$.

Since only case (iii) is possible, we obtain that $n_1 = n_2$ and $m_1 = m_2$, hence $(n_1, m_1) = (n_2, m_2)$. Therefore f is $1-1$. By Corollary 12.0.9, $\mathbb{N} \times \mathbb{N}$ is countable. However, $\mathbb{N} \times \mathbb{N}$ contains the infinite subset $\{(n, 1) : n \in \mathbb{N}\}$ and hence is not finite. So $\mathbb{N} \times \mathbb{N}$ is countably infinite. ■

One can construct a bijection from $\mathbb{N} \times \mathbb{N}$ by writing the elements of $\mathbb{N} \times \mathbb{N}$ in a rectangular array and listing them by counting them off along the diagonals in a sequential manner, but it is technical to write that map down explicitly.

A slightly more general, and useful, version of the previous result is as follows.

Corollary 12.0.11 *Suppose A and B are countable sets. Then $A \times B$ is countable.*

PROOF. If either A or B is empty, then so is $A \times B$, hence $A \times B$ is countable. Now assume $A \neq \emptyset$ and $B \neq \emptyset$. Since A and B are countable, by Corollary 12.0.9 there exist $1-1$ maps $f : A \to \mathbb{N}$ and $g : B \to \mathbb{N}$. Since $\mathbb{N} \times \mathbb{N}$ is countable, there exists s $1-1$ map $h : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. Define $\varphi : A \times B \to \mathbb{N}$ by

$$\varphi(a, b) = h((f(a), g(b))),$$

for $a \in A$ and $b \in B$. We claim that φ is $1-1$. To prove the claim, suppose that $a_1, a_2 \in A, b_1, b_2 \in B$, and $\varphi((a_1, a_2)) = \varphi((a_2, b_2))$. That means that $h((f(a_1), g(b_1)) = h((f(a_2), g(b_2))$. Since h is $1-1$, we conclude that $(f(a_1), g(b_1)) = (f(a_2), g(b_2))$. By the definition of ordered pairs, we gave $f(a_1) = f(a_2)$ and $g(b_1) = g(b_2)$. Since f and g are $1-1$, we have $a_1 = a_2$ and $b_1 = b_2$, so $(a_1, b_1) = (a_2, b_2)$. Therefore $\varphi : A \times B \to \mathbb{N}$ is $1-1$. By Corollary 12.0.9, $A \times B$ is countable. ■

Looking at the product of two sets is natural enough, but we will use our last result to consider unions of countable sets. One natural question is whether the union of two (or more generally, finitely many) countable sets is countable. It will turn out that not only is that true, but a much stronger statement is true: the union of countably many countable sets is countable.

Theorem 12.0.12 *Suppose that A_k is a countable set, either for each k in $\{1, 2, \ldots, n\}$ or for each $k \in \mathbb{N}$. Then $\cup_k A_k$ is countable.*

PROOF. First suppose A_k is countable for each $k \in \mathbb{N}$. If $\cup_{k=1}^{\infty} A_k = \emptyset$, then $\cup_{k=1}^{\infty} A_k$ is countable, so assume $\cup_{k=1}^{\infty} A_k \neq \emptyset$. Since each A_k is countable, each A_k can be listed, i.e., we can write

$$A_1 = \{a_{1,1}, a_{1,2}, a_{1,3}, \ldots\} = \{a_{1,j}\}_j,$$

$$A_2 = \{a_{2,1}, a_{2,2}, a_{2,3}, \ldots\} = \{a_{2,j}\}_j,$$

and, more generally,

$$A_k = \{a_{k,1}, a_{k,2}, a_{k,3}, \ldots\} = \{a_{k,j}\}_j,$$

for each $k \in \mathbb{N}$. The index j runs over a countable set for each k, which may be either finite or countably infinite depending on j. It may even be that $A_k = \emptyset$ for some $k \in \mathbb{N}$, in which case there are no $a_{k,j}$. For each fixed k, we have $a_{k,j} \neq a_{k,\ell}$ if $j \neq \ell$ (i.e., there are no repetitions in the listing for each A_k). The difficulty to overcome is that the same element may belong to different A_k and hence be repeated in the listings for different k; for example, it may be that $a_{1,3} = a_{2,4}$. To deal with this problem, for each $a \in \cup_k A_k$, define

$$k(a) = \min\{k \in \mathbb{N} : a \in A_k\}.$$

(The set $\{k \in \mathbb{N} : a \in A_k\}$ is a subset of $\mathbb{N}$ which is non-empty since $a \in \cup_k A_k$, so it has a minimum which is in the set, by Lemma 10.0.3.) Then for each $a \in \cup_k A_k$, there exist unique $k(a) \in \mathbb{N}$ and $j(a) \in \mathbb{N}$ such that $a \in A_{k(a)}$ and $a = a_{k(a),j(a)}$.

Define $f : \cup_{k=1}^{\infty} A_k \to \mathbb{N} \times \mathbb{N}$ by

$$f(a) = (k(a), j(a)).$$

This map is well-defined (that is, defined unambiguously) because $k(a)$ and $j(a)$ are uniquely determined by a. We claim that f is $1-1$. To prove this claim, suppose $a, b \in \cup_k A_k$ and $f(a) = f(b)$. By definition, this means that $(k(a), j(a)) = (k(b), j(b))$. Hence $k(a) = k(b)$ and $j(a) = j(b)$. Since $k(a) = k(b)$, we have $a, b \in A_{k(a)} = A_{k(b)}$. Since elements are not repeated in the listing for $A_{k(a)} = A_{k(b)}$, and we have $j(a) = j(b)$, we deduce that $a = b$. Hence f is $1-1$.

Since $\mathbb{N} \times \mathbb{N}$ is countably infinite (by Theorem 12.0.10), there exists a $1-1$ map $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. Then $g \circ f : \cup_k A_k \to \mathbb{N}$ is 1-1 (since f and g are $1-1$, using Proposition 4.0.8). Hence $\cup_k A_k$ is countable, by Corollary 12.0.9.

Now suppose we have only finitely many countable sets $A_1, A_2, \ldots, A_n$. Define $A_{n+1} = \emptyset, A_{n+2} = \emptyset$, etc., i.e., $A_k = \emptyset$ for all $k > n$. Then all A_k are countable, for $k \in \mathbb{N}$. By the first part of this proof, $\cup_{k=1}^{\infty} A_k$ is countable. But since $A_k = \emptyset$ for all $k > n$, we have $\cup_{k=1}^{n} A_k = \cup_{k=1}^{\infty} A_k$, which we just noted is countable. ■

The last result states that a countable union of countable sets is countable.

We now have enough background to apply our work on countable and uncountable sets to examples relating to the real numbers. We first resolve the question of the countability or uncountability of the rational numbers $\mathbb{Q}$.

Proposition 12.0.13 *$\mathbb{Q}$ is countable.*

PROOF. We first show that $\mathbb{Q}_+ = \{r \in \mathbb{Q} : r > 0\}$ is countable. For each $r \in \mathbb{Q}_+$, we can write $r = \frac{p}{q}$, with $p, q \in \mathbb{N}$, such that p and q have no common factors (i.e., $\frac{p}{q}$ is the representation of r in lowest terms). For each $r \in \mathbb{Q}_+$, these $p, q \in \mathbb{N}$ are unique. Therefore we can define $f : \mathbb{Q}_+ \to \mathbb{N} \times \mathbb{N}$ by

$$f(r) = (p, q), \text{ where } r = \frac{p}{q} \text{ and } p \text{ and } q \text{ have no common factors.}$$

Then f is $1-1$: if $r_1, r_2 \in \mathbb{Q}_+$ and $f(r_1) = f(r_2)$, then we have $f(r_1) = (p_1, q_1)$ where $r_1 = \frac{p_1}{q_1}$, and $f(r_2) = (p_2, q_2)$ where $r_2 = \frac{p_2}{q_2}$. Hence $(p_1, q_1) = f(r_1) = f(r_2) = (p_2, q_2)$, so $p_1 = p_2$ and $q_1 = q_2$. Hence $r_1 = \frac{p_1}{q_1} = \frac{p_2}{q_2} = r_2$. Therefore f is $1-1$.

Since $\mathbb{N} \times \mathbb{N}$ is countably infinite (Theorem 12.0.10), there exists a $1-1$ map $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. Then $g \circ f : \mathbb{Q}_+ \to \mathbb{N}$ is $1-1$ (since f and g are $1-1$). Therefore $\mathbb{Q}_+$ is countable, by Corollary 12.0.9.

Let $\mathbb{Q}_- = \{r \in \mathbb{Q} : r < 0\}$. Then the map $h : \mathbb{Q}_+ \to \mathbb{Q}_-$ defined by $h(r) = -r$ is easily seen to be a bijection. Therefore $\mathbb{Q}_-$ is countable.

Hence $\mathbb{Q}$ is countable, by Theorem 12.0.12, because $\mathbb{Q} = \mathbb{Q}_+ \cup \mathbb{Q}_- \cup \{0\}$ is the union of three countable sets. ■

It is not so easy to enumerate the rationals explicitly, that is, to specify the values r_i for all i such that

$$\mathbb{Q} = \{r_1, r_2, \ldots, \} = \{r_k\}_{k=1}^{\infty},$$

but we now know that such an enumeration of $\mathbb{Q}$ exists. For many purposes, an exact formula for the enumeration is not needed; we just need to know that an enumeration exists.

Cardinality of $\mathbb{R}$

The results of the previous section show that countability is a relatively stable property: products of countable sets are countable, and countable unions of countable sets are countable. Moreover the rational numbers are countable, and we know that the rational numbers are dense in $\mathbb{R}$ in the sense of Theorem 11.4. So perhaps the next result comes as a surprise.

Theorem 12.0.14 *$\mathbb{R}$ is uncountable.*

PROOF. We claim that the interval $[0, 1]$ is uncountable. If so, then $\mathbb{R}$ can not be countable, since a countable set cannot have an uncountable subset (Theorem 12.0.8). To show that $[0, 1]$ is uncountable, we argue by contradiction. If $[0, 1]$ is countable, we can write

$$[0, 1] = \{x_1, x_2, x_3, \ldots, \} = \{x_k\}_{k=1}^{\infty}.$$

We claim that we can find a sequence $\{I_k\}_{k=1}^{\infty}$ such that each I_k is an interval $I_k = [a_k, b_k]$ with $a_k < b_k$ (so I_k is non-empty), such that $I_{k+1} \subseteq I_k$ for all $k \geq 1$, and $x_k \notin I_k$ for each $k \in \mathbb{N}$. To prove the claim, we argue by induction.

To define I_1, we note that $[0, 1]$ is the union of the three intervals $[0, \frac{1}{3}], [\frac{1}{3}, \frac{2}{3}]$, and $[\frac{2}{3}, 1]$. Since these intervals overlap at some of the endpoints, it is possible for a point to belong to two of these intervals. But is not possible for a point to belong to all 3 of these intervals. Let I_1 be one of these intervals that does not contain x_1 (if there are 2 such intervals, choose one).

We now divide I_1 into three equal length closed sub-intervals (i.e., intervals containing the endpoints) which intersect at most at the endpoints. Then x_2 can not belong to all 3 of these sub-intervals (in fact, x_2 might not belong to any of these three sub-intervals, because x_2 may not belong to I_1). Let I_2 be any of these sub-intervals that does not contain x_2. Then $I_2 \subseteq I_1$ and $x_2 \notin I_2$.

We continue in this way by induction. That is, suppose I_k is a non-empty closed interval contained in I_{k-1} with $x_k \notin I_k$. Then I_k is the union of 3 equal length sub-intervals which overlap at most at the endpoints. Let I_{k+1} be one of these sub-intervals that does not contain x_{k+1} (there may be only one, or there may be more than one, but there is at least one). Then $I_{k+1} \subseteq I_k$ and $x_{k+1} \notin I_{k+1}$. This completes the induction step and thus by induction we have the existence of the sequence $\{I_k\}_{k=1}^{\infty}$ satisfying the stated conditions.

By the Nested Interval Property (Theorem 11.0.8), $\cap_{k=1}^{\infty} I_k \neq \emptyset$. Let $x \in \cap_{k=1}^{\infty} I_k$. Then $x \in I_1 \subseteq [0,1]$, so $x \in [0,1]$. Note that for each $n \in \mathbb{N}$, we have $x_n \notin I_n$, but $x \in I_n$ (since $x \in \cap_{k=1}^{\infty} I_k$). Therefore $x \neq x_n$, for each $n \in \mathbb{N}$. Therefore $x \notin \{x_n\}_{n=1}^{\infty} = [0,1]$, a contradiction. Hence $[0,1]$ is uncountable. ■

The fact that $\mathbb{R}$ is uncountable but $\mathbb{Q}$ is countable means that there are a lot of irrational numbers (in some sense, "most" or even "nearly all" real numbers are irrational, which would horrify Pythagoras). What examples do we have of numbers that we know are irrational? We know, for example, that $\sqrt{2}$ is an irrational number. One might speculate that most irrational numbers are obtained by taking roots of rational numbers; that is, that most irrational numbers are of the form $\sqrt[n]{\frac{p}{q}}$ for some $n, p, q \in \mathbb{N}$ (of course, some of these are rational). However, that is not the case either. More generally, we say that a real number x is *algebraic* if it satisfies

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

for some $n \in \mathbb{N}$ and some integers $a_0, a_1, \ldots, a_n$. In other words, x is algebraic if x is the root of some polynomial with integer coefficients. For example $x = \sqrt[n]{\frac{p}{q}}$ is algebraic because x satisfies $x^n - \frac{p}{q} = 0$, or equivalently $qx^n - p = 0$. However, it is not terribly difficult to show that the set of algebraic numbers is countable. A real number that is not algebraic is called *transcendental* (because, in abstract algebraic terms, it cannot be obtained as an element of a finite degree field extension of $\mathbb{Q}$, if those terms are familiar to you). Thus "most" real numbers are transcendental. Even though we know this to be true, there are very few explicit numbers that are known to be transcendental. I believe that e and π are known to be transcendental, but I've heard (but can not verify) that it is unknown whether $e + \pi$ is algebraic or transcendental.

One approach to exhibiting irrational numbers explicitly is through the decimal expansion of a number. We have not utilized decimal expansions for several reasons: there are different expansions depending on what base one chooses, it takes quite a bit of work to demonstrate the existence and properties of decimal expansions (which are really defined via infinite series, which we have not considered), and decimal expansions are not necessarily unique (for example, $1 = .99999\cdot = .\overline{9}$, where all digits to the right of the decimal point are 9). Still, it is true that a real number is rational if and only if its decimal expansion eventually repeats. But (I assume) there is no known characterization of algebraic numbers in terms of their decimal expansion.

We have seen that "most" numbers are quite unlike the numbers we usually work with, like integers, rational numbers, or their roots. This fact means that most real numbers are somewhat mysterious, and that the real number system is more complex and subtle than we would have assumed naively.

Comments on Cardinality of Sets

The theory of cardinality is largely due to Georg Cantor (1845-1928), published in the period 1874-1884, long after the development of calculus. Cantor's theory was revolutionary and rejected by some of the prominent mathematicians of his day, including Poincaré and Kronecker. Poincaré called Cantor's work a "grave disease," and Kronecker objected to Cantor's proof that the transcendental numbers are uncountable, calling Cantor a "charlatan" and a "corrupter of youth." Cantor's work attracted the attention of the philosopher Wittgenstein, who described it as "wrong" and "utter nonsense." Nevertheless, an understanding of countable and uncountable sets is crucial to the theory of Lebesgue integration, developed around 1900, which supercedes Riemann integration and is foundational to the modern study of analysis and differential equations. Cantor's work also included the construction of the *Cantor set*, an uncountable set of real numbers which has length 0 in an appropriate sense that we won't discuss here (the right concept is that of the *measure* of a set, which is part of the Lebesgue theory of integration).

One result that Cantor tried for years to prove but failed, is what is now called the Schroeder-Bernstein theorem. It states that if A and B are sets and there exist $1-1$ maps $f : A \to B$ and $g : B \to A$, then $A \approx B$, i.e., there exists a bijection $h : A \to B$. This result is not easy, but can be proved in about than an hour's lecture. Assuming that result, it is reasonable to say that if there is a $1-1$ map $f : A \to B$, then card $A \leq$ card B, read as "the cardinality of A is less than or equal to the cardinality of B." If $A \approx B$, we say that card $A =$ card B. Thus cardinality is a notion of the size of a set which makes sense for infinite sets, which is what we started this section considering. In these terms, the Schroeder-Bernstein theorem states that if card $A \leq$ card B and card $B \leq$ card A, then card $A =$ card B, which means that the notation "$\leq$" in this context behaves as we think it should. If card $A \leq$ card B but it is not true that $A \approx B$, then we say card $A <$ card B. This allows us to consider a hierarchy of cardinalities.

The smallest cardinality possible for an infinite set is the cardinality of $\mathbb{N}$. To see this fact, suppose A is an infinite set. Then $A \neq \emptyset$, so we can select an element $a_1 \in A$ (using the axiom of choice, which we assume along with the axioms of set theory). Since A is infinite, $A \setminus \{a_1\} \neq \emptyset$, so we can select $a_2 \in A \setminus \{a_1\}$. We continue inductively, selecting $a_k \in A \setminus \{a_1, a_2, \ldots, a_{k-1}\}$, for each $k \in \mathbb{N}$. Hence we obtain distinct elements $\{a_k\}_{k=1}^{\infty}$ with each a_k belonging to A. Then the map $f : \mathbb{N} \to A$ defined by $f(k) = a_k$, is a $1-1$ map, so card $A \leq$ card $\mathbb{N}$.

A natural question now is: "what is the next cardinality?" We have seen two examples of uncountable sets so far, namely $\mathcal{P}(\mathbb{N})$ and $\mathbb{R}$. With some work involving decimal expansions, one can show that these have the same cardinality, i.e., $\mathcal{P}(\mathbb{N}) \approx \mathbb{R}$. So one question is whether there are any cardinalities strictly in-between the cardinalities of $\mathbb{N}$ and $\mathbb{R}$. That is, does there exist a set A with card $\mathbb{N} <$ card $A <$ card $\mathbb{R}$. The conjecture that there is no such in-between cardinality is known as the *continuum hypothesis*, because it states that the cardinality of $\mathbb{R}$ ($\mathbb{R}$ is referred to as the *continuum*) is the next cardinality after the cardinality of $\mathbb{N}$. This problem obsessed Cantor for many years, but he was not able to solve it. It turns out to be a very subtle problem. Suppose we make the axioms of Zermelo-Fraenkel set theory plus the axiom of choice (togehter referred to as "ZFC") our starting point, which is standard in modern mathematics. In 1940 the famous logician Kurt Gödel showed that the continuum hypothesis is consistent with ZFC in the sense that if ZFC plus the continuum hypothesis result in a contradiction, then there is already a contradiction in ZFC. (Nobody knows for certain whether there are any contradictions resulting from ZFC, but if there are, then all of mathematics is invalidated, so it is widely assumed, but not proved, that ZFC is consistent.) It might seem that Gödel's result proves the continuum hypothesis, but that is not the case. Gödel's result just says that the continuum hypothesis is consistent with ZFC, not that it follows from ZFC. The world of logic and mathematics was staggered in 1963 when Paul Cohen proved that the negation of the continuum hypothesis is also consistent with ZFC, in the same sense. Thus the continuum hypothesis can not be resolved within ZFC. Either the continuum hypothesis or its negation has to be a separate axiom. The continuum hypothesis is what is now called "undecidable." The *generalized continuum hypothesis*, which states that for any set A, there is no cardinality strictly between card A and card $\mathcal{P}(A)$, is also undecidable (it is an exercise that card $A <$ card $\mathcal{P}(A)$, for any set A). Gödel proved that in any axiomatic system meeting certain minimum requirements, there will arise questions which are undecidable within that system. This result, known as Gödel's *incompleteness* theorem, has generated a great deal of speculation about the limitations of human (or even non-human) knowledge. In practice, however, it has turned out that questions in mathematics that have any chance of being applicable to real-world problems never turn out to be undecidable.

Some logicians study what is called "large cardinals," i.e., sets of very large cardinality. For example, one could let $A_1 = \mathbb{N}, A_2 = \mathcal{P}(\mathbb{N}), A_3 = \mathcal{P}(\mathcal{P}(A))$, and, more generally, define inductively $A_{k+1} = \mathcal{P}(A_k)$ for each $k \in \mathbb{N}$. Then one could let $B = \cup_{k=1}^{\infty} A_k$. Then one could start the process over, letting $B_1 = B$ and $B_{k+1} = \mathcal{P}(B_k)$, and then form $\cup_{k=1}^{\infty} B_k$, and so on. However, within applicable mathematics, one usually does not need to consider sets with cardinality greater than $\mathbb{R}$, and almost never does one need to consider sets with cardinality greater than $\mathcal{P}(\mathbb{R})$.

There can't be a largest cardinality, because for any set A, the set $\mathcal{P}(A)$ has a larger cardinality than A. However, this implies that one has to be somewhat careful about what one regards as a set. For example, let's ask the question: can we define the union of all sets, i.e., the set

$$A = \cup\{B : B \text{ is a set}\}?$$

Then A would contain every set as a subset, including $\mathcal{P}(A)$. But for any two sets B and C, if $B \subseteq C$, then card $B \leq$ card C, because the identity map $i : B \to C$, defined by $i(b) = b$ for all $b \in B$, is $1-1$. Thus we

would have card $\mathcal{P}(A) \leq$ card A, which we know to be false. The resolution of this apparent contradiction is that not everything that we imagine to be a set is really a set. One has to be careful about what things are really sets. In particular, there is no set which is the union of all sets. However, although this point is somewhat disconcerting, it turns out that set theory does allow for the existence of all of the sets we need and use in analysis and differential equations, so in practice we don't need to pay a great deal of attention to the issue of what defines a set.

A related and famous conundrum is *Russell's paradox*, after the philospher Bertrand Russell who tried, but eventually gave up on, putting philosophy on a firm foundation by basing it on set theory. Russell considered forming a set A consisting of all sets. That is, the elements of A would themselves be sets, and all sets would be elements of A. So A would be the set of all sets. Then A would have to be an element of A, i.e., $A \in A$. The possibility of allowing a set to be an element of itself then allows one to consider the set of all sets which are not an element of themselves, i.e.,

$$B = \{C : C \text{ is a set and } C \notin C\}.$$

Russell then asked the thorny question: is $B \in B$? If $B \in B$, then by definition of B, B does not satisfy the criterion for being an element of B, so $B \notin B$, contradicting the assumption that $B \in B$. Alternatively, if $B \notin B$, then B does satisfy the criterion for being an element of B, so $B \in B$, which contradicts the assumption $B \notin B$. The resolution of Russell's paradox is that A and B above do not exist as sets. Instead, one has to define more carefully what sets exist.

The philosopher Gottlob Frege (1848-1925), described by Wikipedia as "largely ignored during his lifetime," wrote a philosophical text which claimed, among other things, that the notion of a set was sufficiently simple that dealing with it naively could not lead to any contradictions. As his text neared completion, he received a short letter from Russell laying out Russell's paradox, which demonstrates that one can not just deal naively with sets. Frege had the intellectual honesty to recognize the validity of Russell's point, and wrote what Wikipedia calls "the exceptionally honest comment:"

"Hardly anything more unfortunate can befall a scientific writer than to have one of the foundations of his edifice shaken after the work is finished. This was the position I was placed in by a letter of Mr. Bertrand Russell, just when the printing of this volume was nearing its completion."

Chapter 13

Sequences and Limits

Sequences

Basic logic, set theory, functions, cardinality, and the fundamental properties of $\mathbb{R}$ are part of the foundations of mathematics. Analysis really starts with the notion of limits. We are (finally) ready to start analysis. We begin by considering limits in the context of sequences of real numbers. We start with the formal definition of such a sequence.

Definition 13.0.1 *A sequence of real numbers is a function $s : \mathbb{N} \to \mathbb{R}$.*

We will just use the term "sequence" rather than "sequence of real numbers" when the context is clear. From the definition, a sequence s is determined by its values $s(n)$ for $n \in \mathbb{N}$. Generally we change notation and write s_n instead of $s(n)$, and instead of thinking of s as a function, we think of it as an infinite ordered list of real numbers:

$$s = (s_1, s_2, s_3, \ldots, s_n, \ldots) = (s_n)_{n=1}^{\infty}.$$

Note that n in the notation "(s_n)" is a "dummy index," because it takes the values $1, 2, 3, \ldots$. It doesn't matter what letter we use for the index; the sequence (s_n) is the same as the sequence (s_k) or (s_i). The index just stands for the non-negative integer values it takes.

The values s_n in the sequence are sometimes called the *entries* in the sequence. For concision, we usually use the notation (s_n) to denote the sequence $(s_n)_{n=1}^{\infty}$. Sometimes it is convenient to modify notation and allow the indexing of a sequence to start at a different value, such as $(s_0, s_1, s_2, \ldots)$ or $(s_3, s_4, s_5, \ldots)$. That variation is not a serious violation of the definition, because we could make such a list start at index 1 by renaming. For example, if we let $t_n = s_{n-1}$ then $(s_0, s_1, s_2, \ldots) = (t_1, t_2, t_3, \ldots)$, and (t_n) satisfies the usual indexing for a sequence.

Sometimes a sequence is given by a simple formula, such as $s_n = n^2$, which defines the sequence

$$(s_n) = (1, 4, 9, 16, 25, \ldots).$$

Sometimes a sequence is given by a pattern that is not easy to express in a formula, such as

$$(s_n) = \left(1, \frac{1}{2}, 1, \frac{1}{2}, \frac{1}{3}, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, 1, \ldots\right).$$

Or there may not be any clear pattern; for example we know that we can enumerate $\mathbb{Q}$ (since $\mathbb{Q}$ is countable, by Proposition 12.15). So there exists a sequence (r_n) such that

$$\mathbb{Q} = \{r_1, r_2, r_3, \ldots\}.$$

(Note however that the set $\{r_1, r_2, \ldots\}$ is different from the sequence $(r_1, r_2, \ldots)$ because the ordering of the elements of a set doesn't matter, but the ordering of the entries of a sequence does matter: the sequences $(1, 0, 1, 0, \ldots)$ and $(0, 1, 0, 1, 0, \ldots)$ are different but the sets $\{0, 1\}$ and $\{1, 0\}$ are the same.) We can imagine a sequence geometrically as an infinite number of hash-marks on the real line, chosen sequentially. There

doesn't have to be any clear process for choosing the values in the sequence; any values can be used to form a sequence. Sequences can be defined recursively, such as the sequence in Section 1 defined by $s_1 = 2$ and, for each $n \geq 1$, $s_{n+1} = \frac{1}{2}\left(s_n + \frac{C}{s_n}\right)$. Often we are interested in the long-term behavior of a sequence. That is, we ask what happens to s_n as n gets very large. This raises the question of convergence.

Definition of Convergence for Sequences

For the sequence defined by $s_n = \frac{1}{n}$, i.e.,

$$(s_n) = \left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\right),$$

it is intuitively clear that as n gets large, the values s_n get small. We are inclined to say that s_n *converges* to 0. However, there are other sequences for which the question of convergence is not obvious. For example, consider the sequence

$$(s_n) = (1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, \ldots). \tag{13.1}$$

As you go out further in the sequence, you will find 1,000,000 0's followed by a single 1, then 1,000,001 0's followed by a single 1, and so on. Does (s_n) converge to 0? It certainly becomes more and more like the sequence of all 0's as you go out, and the sequence goes to 0 in some average sense. But, no matter how far out you go in the sequence, there will still be entries later that are 1. So the values do not go uniformly to 0. The point is that we cannot decide whether a sequence converges until we have a clear definition of convergence. Defining convergence turns out to be much more difficult than one might think, and the definition looks unnecessarily complicated. However, years of consideration in the 1800s eventually convinced mathematicians that no simpler defintion will do.

To motivate the definition which we will give shortly, let's try to rigorously formulate what it should mean for a sequence (s_n) to converge to a real number ℓ (convergence to $+\infty$ or $-\infty$ requires a slightly different definition, so for now we don't consider those possibilities). We use the notation $s_n \to \ell$ to mean that s_n converges (as $n \to \infty$, but that is understood) to ℓ. At the most basic level, $s_n \to \ell$ should mean that the values s_n get close to ℓ eventually. So our first approximation to the definition of $s_n \to \ell$ is:

(i) $s_n \to \ell$ if s_n gets close to ℓ eventually.

This definition is highly imprecise, but it is a starting point that we can try to make more exact. Perhaps the first question we should ask is: how do we measure closeness? That is fairly simple to answer: the distance between s_n and ℓ, which is $|s_n - \ell|$, should be small. So our second approximation to the definition of convergence is

(ii) $s_n \to \ell$ if $|s_n - \ell|$ gets small eventually.

Nest, let's work on making the phrase "gets small" more precise. Whether a number is small or not depends on your perspective, or, in some sense, your choice of units for $\mathbb{R}$. Is it good enough it $|s_n - \ell| < .001$ eventually, because .001 is generally thought to be pretty small? Well, no, because no one would think that the constant sequence $(10^{-6}, 10^{-6}, 10^{-6}, \ldots)$ is converging to 0, even though all of its values are within .001 of 0. We might say that s_n gets closer and closer to ℓ (i.e., $|s_n - \ell|$ gets smaller and smaller) as n increases, but that phrasing can be misleading because it seems to imply that $|s_n - \ell|$ is monotonically decreasing, i.e., that $|s_{n+1} - \ell| \leq |s_n - \ell|$ for each n, which we do not want to be part of the definition. For example, the sequence with $s_n = \frac{1}{n}$ for n odd and $s_n = \frac{1}{n^2}$ for n even, i.e.,

$$(s_n) = \left(1, \frac{1}{4}, \frac{1}{3}, \frac{1}{16}, \frac{1}{5}, \frac{1}{36}, \ldots\right)$$

certainly seems to be converging to 0, because all of the terms get small if we go out far enough in the sequence. But they don't approach 0 monotonically; e.g., $\frac{1}{16}$ comes earlier in the sequence than $\frac{1}{5}$, but $\frac{1}{16}$ is closer to 0 than $\frac{1}{5}$. What we really mean is that for any degree of closeness we want, the terms in the sequence are within that level of closeness. For s_n to converge to ℓ, it should be the case that eventually $|s_n - \ell| < .1$, but then even further out ("more eventually") perhaps, we should have $|s_n - \ell| < .01$, then

even further, $|s_n - \ell| < .001$, and so on. The way to say that mathematically is to say that for any small positive number, eventually $|s_n - \ell|$ that is less than that number. It is standard to use the letter ϵ to denote a number that is positive but can be arbitrarily small. We can then say that for any $\epsilon > 0$, we must have $|s_n - \epsilon| < \epsilon$ eventually. So our third approximation to the definition of convergence is

(ii) $s_n \to \ell$ if, for any $\epsilon > 0$, we have $|s_n - \ell| < \epsilon$ eventually.

This approximation to the definition is still vague. First, in view of the sequence in (13.1), we have to decide whether we mean $|s_n - \ell| < \epsilon$ on average, or uniformly from some point on. Although convergence in some average sense is sometimes considered, it turns out that the most useful concept is convergence in the uniform sense, which means that $|s_n - \ell|$ is required to be smaller than ϵ for all values of n, eventually.

Next we need to clarify what we mean by "eventually." Of course there is no prescribed rate that defines "eventually." That is, we don't say that we must have $|s_n - \ell| < .01$ for $n > 10,000$, say, because the sequence that starts with $10,001$ entries that are 1 and has all remaining entries equal to 0 still converges to 0. What we mean is that there exists some point such that, past that point, we have $|s_n - \epsilon| < .01$, for example. We can make this precise by saying that there exists $N \in \mathbb{N}$ such that $|s_n - \epsilon| < .01$ for all $n > N$. However, we also need to have $|s_n - \ell| < .001$ for all n greater than some N, but that value of N may be larger than the N that implies $|s_n - \epsilon| < .01$ for all $n > N$. In other words, for each $\epsilon > 0$, there must exist an $N \in \mathbb{N}$ such that $|s_n - \ell| < \epsilon$ for all $n > N$, but it is important to understand that the required N depends on ϵ. As we vary ϵ, we will have to choose different N. Logically, however, the fact that N may depend on ϵ is implicit in the fact that we said that for each $\epsilon > 0$ there exists an N. In general, when we say that under certain conditions, something else exists, we are implying that that something else is determined by, and hence varies with, the assumed conditions. If we want to emphasize that N may depend on ϵ we can write $N = N(\epsilon)$, even though striclty speaking, it is not logically necessary. So our fourth attempt at the definition is as follows (and is the ultimate definition we need).

Definition 13.0.2 *Let (s_n) be a sequence of real numbers, and let $\ell \in \mathbb{R}$. Then $s_n \to \ell$ (also written $\lim_{n\to\infty} s_n = \ell$) if, for all $\epsilon > 0$, there exists $N = N(\epsilon) \in \mathbb{N}$ such that $|s_n - \ell| < \epsilon$ for all integers $n > N$.*

Using the quantifier notation $\forall$ ("for all") and $\exists$ ("there exists"), we can write this definition entirely in symbols:

$$s_n \to \ell \iff \forall \epsilon > 0, \exists N \in \mathbb{N} : n > N \Rightarrow |s_n - \ell| < \epsilon.$$

To those not mathematically trained, this formulation appears incomprehensible, like random symbols, or like the representations of obscene language in comic strips (% * # @ !!). But once one's understanding of mathematical language reaches the point that one can interpret this definition, it has the advantage that it is unambiguous. Whereas words often can be interpreted in different ways (that's part of the beauty of language, where, for example, one can give multiple interpretations of the same work of literature), mathematical statements must have only one interpretation. Taking out all of the words and expressing the statement entirely in symbols forces us to make the meaning precise.

Let's see how the definition of the convergence of a sequence can be used in examples.

Examples of Proving Sequence Convergence From the Definition

To prove convergence of a sequence (s_n) to a limit $\ell \in \mathbb{R}$, the definition tells us that we have to show that, for every $\epsilon > 0$, there is some $N \in \mathbb{N}$ such that the estimate $|s_n - \ell| < \epsilon$ holds for all $n > N$. We can regard the definition as a challenge: if you are given an $\epsilon > 0$, you have to show that you can find N. It is not enough to do it for a certain $\epsilon > 0$, such as $\epsilon = .01$; you have to show a procedure that will work no matter what positive ϵ you are given.

As our first example, suppose we want to show that $\frac{1}{n} \to 0$. Let's work through what we need before writing the formal proof. Suppose that we are given an $\epsilon > 0$. Here $s_n = \frac{1}{n}$ and $\ell = 0$. So the quantity we need to estimate is $|s_n - \ell| = |\frac{1}{n} - 0| = \frac{1}{n}$. That is, we need to show that eventually, $\frac{1}{n} < \epsilon$. When is that true? By multiplying the inequality by n and dividing by ϵ (neither of which changes the direction of the inequality, because both terms are positive), we see that $\frac{1}{n} < \epsilon$ holds when $n > \frac{1}{\epsilon}$. However, we need to find N so that the estimate holds for all $n > N$. To do that, we can just choose $N > \frac{1}{\epsilon}$, because then if $n > N$ then we have $n > N > \frac{1}{\epsilon}$, so $n > \frac{1}{\epsilon}$. There is one more point to be sure of: how do we know we can select an $N \in \mathbb{N}$ such that $N > \frac{1}{\epsilon}$? Well, if there were no such $N \in \mathbb{N}$, then we would have $N \leq \frac{1}{\epsilon}$ for all

$N \in \mathbb{N}$, which would mean that $\frac{1}{\epsilon}$ would be an upper bound for $\mathbb{N}$. But by Lemma 11.0.1, $\mathbb{N}$ is not bounded above. So we can select $N \in \mathbb{N}$ satisfying $N > \frac{1}{\epsilon}$. Once we know how it will go, the formal proof can be very concise, as follows.

Example 13.0.3 *Prove that* $\frac{1}{n} \to 0$.

Proof. Let $\epsilon > 0$. Select $N \in \mathbb{N}$ such that $N > \frac{1}{\epsilon}$. Then for all $n > N$, we have $n > N > \frac{1}{\epsilon}$, so $\frac{1}{n} < \epsilon$, and hence

$$\left|\frac{1}{n} - 0\right| = \frac{1}{n} < \epsilon.$$

■

Let's consider a more difficult example. Let $s_n = \frac{3n+2}{4n+9}$. Note that as $n \to \infty$, both the numerator and the denominator of the fraction $\frac{3n+2}{4n+1}$ go to $+\infty$. However, we expect that the ration converges to $\frac{3}{4}$ because, for n really large, $3n+2$ is essentially like $3n$, and $4n+9$ is essentially like $4n$, and $\frac{3n}{4n} = \frac{3}{4}$. Let's see if we can verify this intuition rigorously using the definition of convergence. As for the first example, it is helpful to do some work on the side before writing the proof. Here $s_n = \frac{3n+2}{4n+9}$ and $\ell = \frac{3}{4}$. So the quantity that we will need to estimate is

$$|s_n - \ell| = \left|\frac{3n+2}{4n+9} - \frac{3}{4}\right| = \left|\frac{4\cdot(3n+2) - 3\cdot(4n+9)}{4(4n+9)}\right| = \left|\frac{-19}{16n+36}\right| = \frac{19}{16n+36}.$$

Notice that we had to do some algebra to simplify the expression for $s_n - \ell$. We need to choose $N \in \mathbb{N}$ to make $\frac{19}{16n+36} < \epsilon$ for $n > N$. Multiplying the inequality $\frac{19}{16n+36} < \epsilon$ on both sides by $16n+36$ and dividing by ϵ gives $16n + 36 > \frac{19}{\epsilon}$ (note that since $16n + 36 > 0$ for all $n \in \mathbb{N}$, and $\epsilon > 0$, these operations do not change the sign of the inequality). Subtracting 36 and dividing by 16 gives that we need $n > \frac{1}{16}\left(\frac{19}{\epsilon} - 36\right)$. To make sure that this inequality holds for all $n > N$, we choose $N > \frac{1}{16}\left(\frac{19}{\epsilon} - 36\right)$. Once we have determined our choice of N, we can write the proof directly as follows.

Example 13.0.4 *Prove that* $\frac{3n+2}{4n+9} \to \frac{3}{4}$.

Proof. Let $\epsilon > 0$. Select $N \in \mathbb{N}$ such that $N > \frac{1}{16}\left(\frac{19}{\epsilon} - 36\right)$. Assume $n > N$. Then $n > N > \frac{1}{16}\left(\frac{19}{\epsilon} - 36\right)$. Therefore $16n + 36 > \frac{19}{\epsilon}$, and thus $\frac{19}{16n+36} < \epsilon$. Therefore for $n > N$, we have

$$\left|\frac{3n+2}{4n+9} - \frac{3}{4}\right| = \left|\frac{4\cdot(3n+2) - 3\cdot(4n+9)}{4(4n+9)}\right| = \left|\frac{-19}{16n+36}\right| = \frac{19}{16n+36} < \epsilon.$$

■

Alternatively, since we are not required to find the best (smallest) N that works, just some N, we can use a simple inequality to make things easier. Namely, since $16n + 36 > 16n$, we can say that $\frac{19}{16n+36} < \frac{19}{16n}$. Thus if we make $\frac{19}{16n} < \epsilon$, we also have $\frac{19}{16n+36} < \epsilon$. To make $\frac{19}{16n} < \epsilon$, we just make $n > \frac{19}{16\epsilon}$, which will hold for all $n > N$ if we make $N > \frac{19}{16\epsilon}$. Hence an alternate proof goes as follows.

Example 13.0.5 *Prove that* $\frac{3n+2}{4n+9} \to \frac{3}{4}$.

Proof. Let $\epsilon > 0$. Select $N \in \mathbb{N}$ such that $N > \frac{19}{16\epsilon}$. Then for $n > N$, we have $n > N > \frac{19}{16\epsilon}$, hence $\frac{19}{16n} < \epsilon$. Therefore

$$\left|\frac{3n+2}{4n+9} - \frac{3}{4}\right| = \left|\frac{4\cdot(3n+2) - 3\cdot(4n+9)}{4(4n+9)}\right| = \left|\frac{-19}{16n+36}\right| = \frac{19}{16n+36} < \frac{19}{16n} < \epsilon.$$

■

The advantage of this proof is that the algebra is a little simpler than for Example 13.0.4.

Let's consider $\lim_{n\to\infty} \frac{3n+2}{4n-9}$. This example may seem to be essentially the same as the last one, since the only difference is that the denominator has changed from $4n+9$ to $4n-9$. Our intuition is that this change

shouldn't be important, since $4n$ is the dominant term in the denominator. This intuition is correct, but the fact that the denominator $4n-9$ can be negative for some values of n adds some technical complications in dealing with the inequalities. Here $s_n = \frac{3n+2}{4n-9}$ and we expect $\ell = \frac{3}{4}$, so the quantity we will need to estimate is

$$|s_n - \ell| = \left|\frac{3n+2}{4n-9} - \frac{3}{4}\right| = \left|\frac{4\cdot(3n+2) - 3\cdot(4n-9)}{4(4n-9)}\right| = \left|\frac{35}{16n-36}\right|.$$

Already we have a slight difference with the previous example, because we cannot say in general that $\left|\frac{35}{16n-36}\right| = \frac{35}{16n-36}$, because for $n=1$ and $n=2$, the quantity $\frac{35}{16n-36}$ is negative. That fact should not be important, because we are interested in what happens as $n\to\infty$, and as long as $n>2$ we have $\frac{35}{16n-36} > 0$. We can avoid this issue if we guarantee that $n>2$, which will hold if $n>N$ and $N>2$. So we should incorporate the requirement $N>2$ when we choose N.

Assuming that $n>N>2$, we have $|s_n-\ell| = \frac{35}{16n-36}$. Then as before, we have $\frac{35}{16n-36} < \epsilon$ if $\frac{35}{\epsilon} < 16n-36$, or $n > \frac{1}{16}\left(\frac{35}{\epsilon}+36\right)$. To get this inequality to hold for all $n>N$, we will choose $N > \frac{1}{16}\left(\frac{35}{\epsilon}+36\right)$. At this point we notice that $\frac{1}{16}\left(\frac{35}{\epsilon}+36\right) > \frac{36}{16} > 2$, so if $N > \frac{1}{16}\left(\frac{35}{\epsilon}+36\right)$, then $N>2$ automatically. So the condition $N>2$ is taken care of by the choice $N > \frac{1}{16}\left(\frac{35}{\epsilon}+36\right)$. In Example 13.0.7 below, we will see that we can incorporate two requirements on N using the maximum of the two estimates, but we don't need to do that here. So our proof can be written as follows.

Example 13.0.6 *Prove that* $\frac{3n+2}{4n-9} \to \frac{3}{4}$.

Proof. Let $\epsilon > 0$. Select $N\in\mathbb{N}$ such that $N > \frac{1}{16}\left(\frac{35}{\epsilon}+36\right)$. Assume $n>N$. Then $n > N > \frac{1}{16}\left(\frac{35}{\epsilon}+36\right)$. In particular, $n > \frac{36}{16} > 2$, so $n\geq 3$, and hence $16n-36 > 16\cdot 3 - 36 = 12 > 0$. Since $n > \frac{1}{16}\left(\frac{35}{\epsilon}+36\right)$, we deduce $16n-36 > \frac{35}{\epsilon}$, and thus $\frac{35}{16n-36} < \epsilon$ (using the fact that $16n-36>0$, so the direction of the inequality is preserved when dividing by $16n-36$). Therefore for $n>N$, we have

$$\left|\frac{3n+2}{4n-9} - \frac{3}{4}\right| = \left|\frac{4\cdot(3n+2) - 3\cdot(4n-9)}{4(4n+9)}\right| = \left|\frac{35}{16n-36}\right| = \frac{35}{16n-36} < \epsilon.$$

■

If one tries to simplify this problem with an inequality in the same way as for Example 13.0.5, there is a problem. We cannot reduce the inequality $\frac{35}{16n-36} < \epsilon$ by dropping the term -36 in the denominator. It is not true that $\frac{35}{16n-36}$ is less than $\frac{35}{16n}$, because the denominator $16n-36$ in $\frac{35}{16n-36}$ is smaller than $16n$, making $\frac{35}{16n-36}$ larger than $\frac{35}{16n}$. So making $\frac{35}{16n} < \epsilon$ does not give us $\frac{35}{16n-36} < \epsilon$, which is what we need to complete the proof. In this case the easiest approach is to continue we did in the proof of Example 13.0.6.

What happens if we don't choose the right value for the limit? For example, what goes wrong if we try to prove, for the sequence in Example 13.0.3, that $\frac{3n+2}{4n+9} \to \frac{2}{3}$? We will need to estimate

$$\left|\frac{3n+2}{4n+9} - \frac{2}{3}\right| = \left|\frac{3\cdot(3n+2) - 2\cdot(4n+9)}{3(4n+9)}\right| = \left|\frac{n-12}{12n+27}\right|.$$

However, there is no way to make $\frac{n-12}{12n+27}$ small for all large enough n. As $n\to\infty$, the values of $\frac{n-12}{12n+27}$ approach $\frac{1}{12}$, so for $\epsilon < \frac{1}{12}$, we will not be able to find the required N.

Sometimes you have to play a little with inequalities to prove convergence. For our next example, let $s_n = \frac{n+4}{3n^2-4n-6}$. Since the power of n in the denominator is greater than in the numerator, we expect that $s_n \to 0$. So $\ell = 0$ and $|s_n - \ell| = |s_n|$. Hence we just need to estimate $|s_n|$. One issue is that s_n is not always positive (e.g., $s_1 = -\frac{5}{7}$), so we cannot just drop the absolute values in $|s_n|$. However, the numerator is positive, and in the denominator the dominant term is $3n^2$, which is positive, so we expect that for n sufficiently large, we will have $s_n > 0$. Let's come back to this point in a moment. First, let's think about estimating $\frac{n+4}{3n^2-4n-6}$. For n large we expect that the highest order terms in the numerator and denominator dominate, so that $\frac{n+4}{3n^2-4n-6}$ should be very close to $\frac{n}{3n^2} = \frac{1}{3n}$ for very large n. It would make things simple if we could say that $\frac{n+4}{3n^2-4n-6} < \frac{1}{3n}$, but that is not true for large n, because (assuming n is large enough so that $3n^2-4n-6>0$) the inequality $\frac{n+4}{3n^2-4n-6} < \frac{1}{3n}$ is equivalent (by cross-multiplying)

to $3n^2 + 12n < 3n^2 - 4n - 6$, or $12n < -4n - 6$, which is clearly false. There are many ways to proceed, however. One way is to try to show something that is still good enough, namely that

$$\frac{n+4}{3n^2-4n-6} < \frac{2}{3n}$$

eventually. When is this true? Assuming $3n^2 - 4n - 6 > 0$, we can cross-multiply to say that this inequality is equivalent to $3n(n+4) < 2 \cdot (3n^2 - 4n - 6)$, or $3n^2 + 12n < 6n^2 - 8n - 12$, or $3n^2 > 20n + 12$. There is no need to find the smallest n where the last inequality is true; we can just say (for example), that if $n > 10$ then $n^2 > 12$ and $2n^2 > 20n$, so adding these inequalities gives $3n^2 > 20n + 12$. Working backwards, we see that $3n^2 > 20n + 12$ implies $\frac{n+4}{3n^2-4n-6} < \frac{2}{3n}$, as long as we know that $3n^2 - 4n - 6 > 0$. We can deal with this issue now, because if we assume $n > 10$ then (as we just showed) $3n^2 > 20n + 12 > 4n + 6$, and so $3n^2 - 4n - 6 > 0$. Thus we will estimate $\frac{n+4}{3n^2-4n-6}$ by $\frac{2}{3n}$ as long as $n > 10$. Then to get $\frac{2}{3n} < \epsilon$ at the end, we need to have $n > \frac{2}{3\epsilon}$. We need these inequalities for all $n > N$, so we choose $N > 10$ and $N > \frac{2}{3\epsilon}$. How do we ensure both conditions? We write: let $N > \max\left(10, \frac{2}{3\epsilon}\right)$. (The notation $x = \max(a, b)$ just means that x is the maximum of a and b; i.e., $x = a$ if $a \geq b$ and $x = b$ if $b \geq a$.) Since we think of ϵ as small, the step of choosing the maximum may seem unnecessary, since for small enough ϵ we will have $\frac{2}{3\epsilon} > 10$, but we need a condition that works for any $\epsilon > 0$, even ϵ large. For the formal proof we have to write much of the inequality work in reverse order, as follows.

Example 13.0.7 *Prove that* $\frac{n+4}{3n^2-4n-6} \to 0$.

Proof. Let $\epsilon > 0$. Select $N \in \mathbb{N}$ such that $N > \max\left(10, \frac{2}{3\epsilon}\right)$. Then for $n > N$, we have $n > 10$ and $n > \frac{2}{3\epsilon}$. Since $n > 10$ we have $n^2 > 10n$, hence $2n^2 > 20n$, and we also have $n^2 > 12$, so $3n^2 > 20n + 12$. Thus $3n^2 + 12n < 6n^2 - 8n - 12$, so $3n(n+4) < 2 \cdot (3n^2 - 4n - 6)$. Noting that $3n^2 > 20n + 12 > 4n + 6$, we have $3n^2 - 4n - 6 > 0$, so from $3n(n+4) < 2 \cdot (3n^2 - 4n - 6)$ we obtain $\frac{n+4}{3n^2-4n-6} < \frac{2}{3n} < \epsilon$, since $n > \frac{2}{3\epsilon}$. Therefore, for $n > N$,

$$\left| \frac{n+4}{3n^2-4n-6} - 0 \right| = \frac{n+4}{3n^2-4n-6} < \epsilon.$$

■

As you can imagine, proofs of this sort become more and more difficult as the degree of the numerator and denominator increase. In the next section, we will prove properties of limits that will allow us to give much simpler arguments dealing with such cases.

To help us understand convergence better, consider what it means if (and how to prove that) s_n does not converge to ℓ, written $s_n \not\to \ell$. Recall that when you take the negation of a "for all" statement, you get a "there exists" statement, and if you negate a "there exists" statment, you get a "for all" statement. So the negation of the statement "for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $|s_n - \ell| < \epsilon$ for all $n > N$" is: "there exists $\epsilon > 0$ such that for all $N \in \mathbb{N}$, there exists $n > N$ such that $|s_n - \ell| \geq \epsilon$." In other words,

$$s_n \not\to \ell \iff \exists \epsilon > 0 : \forall N \in \mathbb{N}, \exists n > N : |s_n - \ell| \geq \epsilon.$$

Let's apply this definition to the sequence in (13.1) to show that this sequence does not converge to 0. If we let $\epsilon = 2$, then we will certainly have $|s_n - 0| < 2$, since the terms s_n are either 0 or 1. However, we just have to find one $\epsilon > 0$ where things fail. Since $|s_n - 0| = 1$ when $s_n = 1$, we take $\epsilon < 1$. Let's choose $\epsilon = \frac{1}{2}$ just to be explicit. Then we have to show that no matter how large you choose N, there will be values $n > N$ for which $|s_n - \ell| > \frac{1}{2}$. But that conclusion holds by the definition of the sequence, since no matter how far out you go, it has entries of 1 beyond that point.

Example 13.0.8 *Prove that the sequence*

$$(s_n) = (1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, \ldots)$$

does not converge to 0

Proof. Let $\epsilon = \frac{1}{2}$. Then for all $N \in \mathbb{N}$, there exist $n \in \mathbb{N}$ such that $n > N$ and $s_n = 1$. For such n, we have

$$|s_n - 0| = |1 - 0| = 1 > \frac{1}{2}.$$

Thus the condition in the definition of convergence fails for $\epsilon = \frac{1}{2}$. ■

Chapter 14

Properties of Limits of Sequences

The examples in the previous section in which the convergence of a sequence is proved directly from the definition are instructive in helping us understand the definition of convergence. However, as the examples become more complicated, it becomes more and more technically complicated to carry out the proof. Instead, it is easier to prove some general properties of limits and use them to verify convergence in more complex examples. In this section we state and prove some key properties of limits of sequences.

The first comment, which seems almost obvious, but should still be proved, is that a sequence can have at most one limit. A sequence may not have a limit, but if it has one, that limit is unique.

Proposition 14.0.1 *Suppose (s_n) is a sequence of real numbers, $s_n \to \ell_1$ and $s_n \to \ell_2$, where $\ell_1, \ell_2 \in \mathbb{R}$. Then $\ell_1 = \ell_2$.*

Proof. We argue by contradiction. Suppose $\ell_1 \neq \ell_2$. Then $|\ell_1 - \ell_2| > 0$. Let $\delta = |\ell_1 - \ell_2|$. Since $s_n \to \ell_1$, there exists $N_1 \in \mathbb{N}$ such that $|s_n - \ell_1| < \delta/2$ for all $n > N_1$. Since $s_n \to \ell_2$, there exists $N_2 \in \mathbb{N}$ such that $|s_n - \ell_2| < \delta/2$ for all $n > N_2$. Let $n > \max(N_1, N_2)$. Then by the triangle inequality (Proposition 8.4 (iii)),

$$\delta = |\ell_1 - \ell_2| = |\ell_1 - s_n + s_n - \ell_2| \leq |\ell_1 - s_n| + |s_n - \ell_2| < \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

Hence we have $\delta < \delta$, a contradiction. So $\ell_1 = \ell_2$. ■

There are a couple of useful takeaways from the last proof. The first is that we can apply the definition of sequence convergence with any positive quantity taking the role of ϵ; here we used $\delta/2$ in place of ϵ. The definition says that a certain thing holds for any $\epsilon > 0$, but it doesn't matter if that quantity is given a different name. In later proofs we will have an $\epsilon > 0$ given and we will use that if $s_n \to \ell$, we can obtain $|s_n - \ell| < \epsilon/2$ for n sufficiently large. Second, the part in the proof where we added and subtracted the quantity s_n and then applied the triangle inequality, is a common trick that we will use repeatedly.

It is useful to consider the notion of a sequence being bounded above, bounded below, or both.

Definition 14.0.2 *A sequence (s_n) is bounded above if there exists $M \in \mathbb{R}$ such that $s_n \leq M$ for all $n \in \mathbb{N}$. If so, we say M is an upper bound for (s_n). We say (s_n) is bounded below if there exists $m \in \mathbb{R}$ such that $m \leq s_n$ for all $n \in \mathbb{N}$; in this case, m is a lower bound for (s_n). We say (s_n) is bounded if (s_n) is both bounded above and bounded below.*

Equivalently, the sequence (s_n) is bounded (or bounded above, or bounded below, respectively) if and only if the set $\{s_n\}_{n=1}^{\infty}$ is bounded (or bounded above, or bounded below, respectively), as in Definition 9.0.1.

For example, the sequence $\left(\frac{1}{n}\right)$ is bounded above (by 1) and bounded below (by 0). The sequence (n) is bounded below (by 1) but not bounded above. An alternate way of saying that (s_n) is bounded is to say that there exists some $M \in \mathbb{R}$ such that $|s_n| \leq M$ for all $n \in \mathbb{N}$ (since if $|s_n| \leq M$ then $-M \leq s_n \leq M$ for all $n \in \mathbb{N}$, so (s_n) is bounded above and below; conversely, if (s_n) is bounded above and below, then there exists $m_1, m_2 \in \mathbb{R}$ such that $m_1 \leq s_n \leq m_2$ for all $n \in \mathbb{N}$, so $|s_n| \leq \max(|m_1|, |m_2|) = M$ for all $n \in \mathbb{N}$). In other words, (s_n) is bounded if and only if $(|s_n|)$ is bounded above.

A basic fact is that convergent sequences are bounded.

Proposition 14.0.3 *Suppose (s_n) is a sequence of real numbers and $s_n \to \ell$, for some $\ell \in \mathbb{R}$. Then (s_n) is bounded.*

PROOF. Let Because $s_n \to \ell$, there exists $N \in \mathbb{N}$ such that $|s_n - \ell| < 1$ for all $n > N$ (applying the definition of convergence with $\epsilon = 1$). Let

$$M = \max(|s_1|, |s_2|, \ldots, |s_N|, 1 + |\ell|).$$

(Note $M \in \mathbb{R}$ because the maximum of finitely many numbers is finite.) For $1 \leq n \leq N$, we have $|s_n| \leq M$ by the definition of M. For $n > N$, using the triangle inequality gives

$$|s_n| = |s_n - \ell + \ell| \leq |s_n - \ell| + |\ell| < 1 + |\ell| \leq M.$$

Thus $|s_n| \leq M$ for all $n \in \mathbb{N}$. ■

Computing limits in many examples is made much easier by the following general facts.

Theorem 14.0.4 *Suppose (s_n) and (t_n) are sequences such that $s_n \to s$ and $t_n \to t$, for $s, t \in \mathbb{R}$. Then*

(1) for $c \in \mathbb{R}$, we have $cs_n \to cs$;
(2) $s_n + t_n \to s + t$;
(3) $s_n t_n \to st$;
(4) if $t_n \neq 0$ for all $n \in \mathbb{N}$ and $t \neq 0$, then $\dfrac{s_n}{t_n} \to \dfrac{s}{t}$.

Before writing the proof, let's think about proving (1). We want to show $cs_n \to cs$, so we will have to estimate $|cs_n - cs| = |c(s_n - s)| = |c||s_n - s|$. By our assumption, $s_n \to s$, so we can make $|s_n - s|$ as small as we like by taking n sufficiently large. It is tempting to use the definition of the convergence of s_n to s to say that we can choose $N \in \mathbb{N}$ so that $|s_n - s| < \epsilon$ for $n > N$. But substituting that estimate above gives $|cs_n - cs| = |c||s_n - s| < c\epsilon$ for $n > N$. But that conclusion is not exactly what is called for in the definition; we want $|cs_n - cs| < \epsilon$. Our problem comes from using ϵ to denote 2 different things; the small number in the definition of $s_n \to \epsilon$ and the small number in the definition of $cs_n \to cs$. One could call them by different terms, say ϵ_1 and ϵ_2, but that can be avoided if we just realize that we can apply the definition of $s_n \to s$ with any small positive quantity. So we can make ϵ be the positive number we want in our conclusion, and apply the assumption with the small positive quantity $\frac{\epsilon}{|c|}$ (We have to use $|c|$ to make sure the quantity is positive). So here's a first try at the proof:

Let $\epsilon > 0$. Since $s_n \to s$, there exists $N \in \mathbb{N}$ such that $|s_n - s| < \frac{\epsilon}{|c|}$ for all $n > N$. Thus for all $n > N$, we have

$$|cs_n - cs| = |c||s_n - s| < |c| \cdot \frac{\epsilon}{|c|} = \epsilon.$$

This effort is almost correct. There is one trivial problem: it is not valid if $c = 0$ because then $\frac{\epsilon}{|c|}$ is not defined. But the conclusion is still correct if $c = 0$, because if $c = 0$ then $cs_n = 0$ for all n, and $cs = 0$, and it is pretty obvious (and easy to prove, although we haven't done it yet) that the sequence whose entries are all 0 converges to 0 (and more generally, any constant sequence converges to that constant). So one approach is to start by saying: suppose $c = 0$. Then write the proof that the 0 sequence converges to 0. Then say: now suppose $c \neq 0$, and then write the proof above. This gives a valid proof, but there is another approach that is a little shorter (mathematicians love concision) and doesn't require that we consider separate cases (mathematicians find breaking proofs into cases annoying and inelegant, although sometimes it is unavoidable). The approach is based on the following trick: in applying the assumption $s_n \to s$, apply the definition with the small quantity being $\frac{\epsilon}{1+|c|}$ instead of $\frac{\epsilon}{|c|}$ as above. Then at the key step we use that $\frac{|c|}{1+|c|} < 1$. We obtain the following proof.

PROOF OF (1) Let $\epsilon > 0$. Since $s_n \to s$, there exists $N \in \mathbb{N}$ such that $|s_n - s| < \frac{\epsilon}{1+|c|}$ for all $n > N$. Thus for all $n > N$, we have

$$|cs_n - cs| = |c||s_n - s| < |c| \cdot \frac{\epsilon}{1 + |c|} = \epsilon \cdot \frac{|c|}{1 + |c|} < \epsilon.$$

□

The choice between this proof and the one where the case $c = 0$ is handled separately is a matter of taste only. Both are fine mathematically. The disadvantage of the last proof is that the choice of $\frac{\epsilon}{1+|c|}$ seems to come out of nowhere, until one understands that the term $1 + |c|$ was just used as a convenience so that the case $c = 0$ can be included without a separate argument.

The most important thing to gather from this proof is that we can apply our convergence assumptions with other small quantities in place of ϵ. It seems most clear to let ϵ be the small quantity in the definition of what you are trying to prove, and then apply the assumptions with small quantities expressed in terms of ϵ appropriately chosen so that the thing we want comes out at the end to be ϵ. In other words, instead of starting out stating the definitions of our assumptions, which seems natural (i.e., gather what you know in order to see how to use it), start out trying to prove the conclusion you need. Then bring in the assumptions as needed to make the proof work. If you state the assumptions in terms of ϵ and then the thing you want to prove also in terms of ϵ, then everything is confused because in the rest of the proof it is never clear which one you mean when you write ϵ. My advice: in analysis, start by trying to prove what you want, and bring in the assumptions when needed. The logic will be much more clear. We follow this approach in the proofs of the remaining parts of Theorem 14.0.4. In each case, it was necessary to work backwards from the estimate we want in order to determine what estimates we will need to get from our assumptions.

Proof of (2) Let $\epsilon > 0$. Since $s_n \to s$, there exists $N_1 \in \mathbb{N}$ such that $|s_n - s| < \frac{\epsilon}{2}$ for all $n > N_1$. Since $t_n \to t$, there exists $N_2 \in \mathbb{N}$ such that $|t_n - t| < \frac{\epsilon}{2}$ for all $n > N_2$. Let $N = \max(N_1, N_2)$. Then for all $n > N$, we have $n > N_1$ and $n > N_2$, so

$$|s_n + t_n - (s + t)| = |s_n - s + t_n - t| \leq |s_n - s| + |t_n - t| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

□

Proof of (3) Let $\epsilon > 0$. Since $s_n \to s$, there exists $N_1 \in \mathbb{N}$ such that $|s_n - s| < \frac{\epsilon/2}{1+|t|}$ for all $n > N_1$. Also, by Proposition 14.0.3, there exists $M \in \mathbb{R}$ such that $|s_n| \leq M$ for all $n \in \mathbb{N}$. Note that the inequality $|s_1| \leq M$ implies that $M \geq 0$. Since $t_n \to t$, there exists $N_2 \in \mathbb{N}$ such that $|t_n - t| < \frac{\epsilon/2}{1+M}$ for all $n > N_2$. Let $N = \max(N_1, N_2)$. Then for all $n > N$, we have $n > N_1$ and $n > N_2$, so

$$|s_n t_n - st| = |s_n t_n - s_n t + s_n t - st| = |s_n(t_n - t) + (s_n - s)t| \leq |s_n(t_n - t)| + |(s_n - s)t|$$

$$= |s_n| \cdot |(t_n - t)| + |s_n - s| \cdot |t| < M \cdot \frac{\epsilon/2}{1+M} + \frac{\epsilon/2}{1+|t|} \cdot |t| = \frac{M}{1+M} \cdot \frac{\epsilon}{2} + \frac{t}{1+t} \cdot \frac{\epsilon}{2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

□

The reason we added and subtracted $s_n t$ in the last proof is that it seems to be something related to both $s_n t_n$ and st; we could have used st_n, which leads to an analogous proof. Notice that we used $1 + M$ and $1 + |t|$ in the denominators just to avoid handling the cases $M = 0$ or $t = 0$ separately, as in the proof of (1).

For part (4), we will have to be concerned with how small a certain denominator can be. To deal with this issue, we first need to prove the following lemma. It says that if a sequence converges to a non-zero number, the sequence terms are eventually bounded away from 0.

Lemma 14.0.5 *Suppose (t_n) is a sequence of real numbers such that $t_n \to t$, where $t \in \mathbb{R}$ and $t \neq 0$. Then there exists $N \in \mathbb{N}$ such that $|t_n| > \frac{|t|}{2}$ for all $n > N$.*

Proof. Since $t_n \to 0$, there exists $N \in \mathbb{N}$ such that $|t_n - t| < \frac{|t|}{2}$ for all $n > N$ (because $|t|/2 > 0$ so it can be used as ϵ in the definition of convergence). By the triangle inequality, if $n > N$ we have

$$|t| = |t - t_n + t_n| \leq |t - t_n| + |t_n| < \frac{|t|}{2} + |t_n|,$$

so subtracting $\frac{|t|}{2}$ from both sides, we get that $|t_n| > \frac{|t|}{2}$ for all $n > N$. ■

Note how we used the triangle inequality in the last proof: we wanted to bound $|t_n|$ below by $\frac{|t|}{2}$, so we started with $|t|$ and used the triangle inequality to make a lower bound for $|t|$ that involved $|t_n|$.

PROOF OF (4) We will first show that $\frac{1}{t_n} \to \frac{1}{t}$. Let $\epsilon > 0$. Since $t_n \to t$ and $t \neq 0$, by Lemma 14.0.5, there exists N_1 such that $|t_n| > \frac{|t|}{2}$ for all $n > N_1$. Then $\frac{1}{|t_n|} < \frac{2}{|t|}$ for $n > N_1$. Also there exists $N_2 \in \mathbb{N}$ such that $|t_n - t| < \frac{\epsilon t^2}{2}$ for $n > N_2$ (because $\frac{\epsilon t^2}{2} > 0$ since $t \neq 0$). Let $N = \max(N_1, N_2)$. Then for $n > N$, we have $n > N_1$ and $n > N_2$, so

$$\left|\frac{1}{t_n} - \frac{1}{t}\right| = \left|\frac{t - t_n}{tt_n}\right| = \frac{1}{|t|} \cdot \frac{1}{|t_n|} \cdot |t_n - t| < \frac{1}{|t|} \cdot \frac{2}{|t|} \cdot \frac{\epsilon t^2}{2} = \epsilon.$$

Now to prove $\frac{s_n}{t_n} \to \frac{s}{t}$, we apply part (3) with t_n replaced by $\frac{1}{t_n}$, which we just showed converges to $\frac{1}{t}$. We obtain

$$\frac{s_n}{t_n} = s_n \cdot \frac{1}{t_n} \to s \cdot \frac{1}{t} = \frac{s}{t}.$$

□

Using these properties of limits makes proving certain limits much easier. For comparison, imagine the difficulty of doing the next example using only the definition of convergence, as in Examples 13.0.3 - 13.0.7.

Example 14.0.6 *Prove that* $\lim_{n\to\infty} \frac{3n^2-4n+5}{5n^2+6n-12} = \frac{3}{5}$.

SOLUTION. We begin an algebraic step: multiplying the numerator and denominator both by $\frac{1}{n^2}$ we get

$$\frac{3n^2 - 4n + 5}{5n^2 + 6n - 12} = \frac{(3n^2 - 4n + 5) \cdot \frac{1}{n^2}}{(5n^2 + 6n - 12) \cdot \frac{1}{n^2}} = \frac{3 - \frac{4}{n} + \frac{5}{n^2}}{5 + \frac{6}{n} - \frac{12}{n^2}}.$$

We calculate the limits of the numerator and denominator in the last expression. Using Theorem 14.0.4 (2),

$$\lim_{n\to\infty} \left(3 - \frac{4}{n} + \frac{5}{n^2}\right) = \lim_{n\to\infty} 3 + \lim_{n\to\infty} \left(-\frac{4}{n}\right) + \lim_{n\to\infty} \left(\frac{5}{n^2}\right).$$

We leave as an exercise to show (using the definition of convergence) that for any constant sequence defined by $s_n = c$ for all n, we have $\lim_{n\to\infty} s_n = c$. So $\lim_{n\to\infty} 3 = 3$. By Theorem 14.0.4 (1), $\lim_{n\to\infty} \left(-\frac{4}{n}\right) = -4 \lim_{n\to\infty} \frac{1}{n} = (-4) \cdot 0 = 0$, using $\lim_{n\to\infty} \frac{1}{n} = 0$ (Example 13.0.3). Using Theorem 14.0.4 (1) and (3),

$$\lim_{n\to\infty} \left(\frac{5}{n^2}\right) = 5 \left(\lim_{n\to\infty} \frac{1}{n}\right) \cdot \left(\lim_{n\to\infty} \frac{1}{n}\right) = 5 \cdot 0 \cdot 0 = 0.$$

Hence, $\lim_{n\to\infty}(3 - \frac{4}{n} + \frac{5}{n^2}) = 3$. In exactly the same way, $\lim_{n\to\infty}(5 + \frac{6}{n} - \frac{12}{n^2}) = 5$. Finally, using Theorem 14.0.4 (4),

$$\lim_{n\to\infty} \frac{3 - \frac{4}{n} + \frac{5}{n^2}}{5 + \frac{6}{n} - \frac{12}{n^2}} = \frac{\lim_{n\to\infty}(3 - \frac{4}{n} + \frac{5}{n^2})}{\lim_{n\to\infty}(5 + \frac{6}{n} - \frac{12}{n^2})} = \frac{3}{5}.$$

■

This method shows that in general, for a rational function $r(n) = \frac{p(n)}{q(n)}$ where p and q are polynomials, the limiting behavior (as $n \to \infty$) of r is determined entirely by the highest order terms in the numerator and denominator.

The basic properties of limits in Theorem 14.0.4 follow pretty directly from the definition of limits. The next result is a little deeper, because it depends on the completeness property of the real numbers, as discussed in Chapters 9-11. First we need a definition.

Definition 14.0.7 *A sequence* (s_n) *is*

(i) increasing if $s_n \leq s_{n+1}$ *for all* $n \in \mathbb{N}$*;*
(ii) strictly increasing if $s_n < s_{n+1}$ *for all* $n \in \mathbb{N}$*;*
(iii) decreasing if $s_{n+1} \leq s_n$ *for all* $n \in \mathbb{N}$*;*
and
(iv) strictly decreasing if $s_{n+1} < s_n$ *for all* $n \in \mathbb{N}$*.*

Some texts use the term "increasing" to mean "strictly increasing" and "non-decreasing" for what we are calling "increasing," but we will stick with the terminology in Definition 14.0.7.

Theorem 14.0.8 *(Monotone Sequence Theorem) Let (s_n) be a sequence of real numbers.*

(1) If (s_n) is increasing and bounded above, then s_n converges to $\sup\{s_n : n \in \mathbb{N}\}$.

(2) If (s_n) is deccreasing and bounded below, then s_n converges to $\inf\{s_n : n \in \mathbb{N}\}$.

PROOF. We prove (1), leaving the analogous proof of (2) as an exercise. Suppose (s_n) is increasing and bounded above, and let $s = \sup\{s_n : n \in \mathbb{N}\}$ (which exists, as an element of $\mathbb{R}$, by the completeness axiom for $\mathbb{R}$, since $\{s_n : n \in \mathbb{N}\}$ is non-empty and bounded above). Let $\epsilon > 0$. By Lemma 10.0.7, there exists $N \in \mathbb{N}$ such that $s_N > s - \epsilon$. For all $n > N$, we have $s_n \geq s_N > s - \epsilon$, since (s_n) is increasing. Also, s is an upper bound for $\{s_n : n \in \mathbb{N}\}$ (since s is the supremum of that set), so $s_n \leq s$. Hence, for all $n > N$, we have $s - \epsilon < s_n \leq s$. Subtracting s in both inequalities gives $-\epsilon < s_n - s \leq 0$, which implies $|s_n - s| < \epsilon$. ■

If we were working with, say, $\mathbb{Q}$, as our universe of numbers instead of $\mathbb{R}$, then the Monotone Sequence Theorem would fail. For example, we could let $s_1 = 1.4, s_2 = 1.41, s_3 = 1.414, s_4 = 1.4141$, etc., where s_n is the decimal approximation of $\sqrt{2}$ to n decimal places. Then $s_n \in \mathbb{Q}$ for all n, and the sequence (s_n) is increasing and bounded above. However (s_n) does not converge in $\mathbb{Q}$ because what should be $\lim_{n\to\infty} s_n$ is $\sqrt{2}$, which is not in $\mathbb{Q}$. The completeness property of $\mathbb{R}$ guarantees that $\mathbb{R}$ is not "missing" any points it should have, which is exactly what is needed for the proof of the Monotone Sequence Theorem.

We will use the Monotone Sequence Theorem to prove the Bolzano-Weierstrass Theorem, which is one of the key results in analysis on $\mathbb{R}$.

Chapter 15

Subsequences, the Bolzano-Weierstrass Theorem, and Cauchy Sequences

The basic philosophy of analysis is to "take it to the limit." Here is an example. Suppose we have a function $f : [0, 1] \to \mathbb{R}$ and suppose $A = \{f(x) : x \in [0, 1]\}$ is bounded above. Our hope is that f has a point where it attains its maximum (which we would then try to find; for example, $f(x)$ might represent the profit we can make depending on some parameter $x \in [0, 1]$ of some production process). We might proceed as follows: since A is bounded above, it has a supremum s (by the completeness property of $\mathbb{R}$). We hope to find a point $x \in [0, 1]$ such that $f(x) = s$, so that the supremum of A is actually attained. We can try to find x as follows. By Lemma 10.0.7, for each $n \in \mathbb{N}$, there is an element, call it $f(x_n)$, of the set A, such that $s - \frac{1}{n} < f(x_n) \leq s$ (the inequality $f(x_n) \leq s$ holds because $s = \sup A$). Now it is certainly true that $\lim_{n\to\infty} f(x_n) = s$, since $|f(x_n) - s| < \frac{1}{n}$, but we don't know that there is a point $x \in [0, 1]$ such that $f(x) = \lim_{n\to\infty} f(x_n) = s$. It is very tempting to say: "let $x = \lim_{n\to\infty} x_n$." This is a mistake because we have no reason to know that the sequence (x_n) is convergent. This mistake is one of the most common errors students make. Just because you can write down the symbols "$\lim_{n\to\infty} x_n$" doesn't guarantee that the symbols are meaningful. The expression "$\lim_{n\to\infty} x_n$" only makes sense if we know that the limit exists, and it may not.

Is there a way to salvage this plan? It turns out that there is. Although (x_n) may not be convergent, it may be that some portion of the sequence (x_n) forms a convergent sequence of its own. Such a "portion of a sequence" is what is called a *subsequence*, which we define below. If that subsequence converges to some x, we then have a candidate for where f attains its maximum. We need another property of f, called *continuity*, which we will eventually discuss, to guarantee that f evaluated at the subsequence points actually converges to $f(x)$. But ultimately we will prove a theorem that states that all of this works. The key concept is to look at f at points x_n where $f(x_n)$ gets closer and closer to the supremum s, and then use limits to find a point x where some of these points x_n pile up; at that point x we should have the extreme behavior we are looking for.

Subsequences

Definition 15.0.1 *Let (s_n) be a sequence. A subsequence of (s_n) is any sequence (t_k) where $t_k = s_{n_k}$ for each $k \in \mathbb{N}$, where n_k is a strictly increasing sequence of natural numbers. We usually write the subsequence in the form $(s_{n_k}) = (s_{n_1}, s_{n_2}, s_{n_3}, \ldots)$.*

In more detail, $t_1 = s_{n_1}, t_2 = s_{n_2}, t_3 = s_{n_3}$, etc., and $n_1 < n_2 < n_3 < \ldots$. A subsequence of (s_n) could be, for example, $(s_3, s_{14}, s_{26}, s_{1,268}, s_{5,439}, \ldots)$.

It is important to understand that the index for the subsequence (s_{n_k}) is k, since we have entries for $k = 1, 2, 3, \ldots$. Also those integers n_k can be arbitrarily chosen, subject only to the constraint that $n_k < n_{k+1}$ for all $k \in \mathbb{N}$. It is easy to show by induction that for any subsequence (s_{n_k}) of (s_n), we have $n_k \geq k$: for

$k = 1$ this fact holds because $n_1 \geq 1$, and, for the induction step, we have $n_k > k$ by the induction hypothesis, so our condition $n_{k+1} > n_k$ implies (since all $n_k \in \mathbb{N}$) that $n_{k+1} \geq n_k + 1 > k + 1$.

For examples of subsequences, consider the sequence $(s_n) = (0, 1, 2, 0, 1, 2, 0, 1, 2, 0, 1, 2, \ldots)$. If we take s_1, s_4, s_7, etc., we always get 0. So $(0, 0, 0, \ldots,)$ is a subsequence of (s_n). So are $(1, 1, 1, \ldots)$, $(2, 2, 2, \ldots)$, and $(0, 1, 0, 1, 0, 1, \ldots)$. Notice that some (but not all) of these subsequences converge even though the original sequence (s_n) diverges.

For another example, suppose $s_n = 1 + \frac{1}{n}$ for n odd, and $s_n = 3 - \frac{1}{n}$ for n even. Then the subsequence $(s_{2k-1}) = (s_1, s_3, s_5, \ldots)$ converges to 1 and the subsequence $(s_{2k}) = (s_2, s_4, s_6, \ldots)$ converges to 3.

The set of points x such that some subsequence of (s_n) converges to x can be quite large. For example, we know that we can enumerate the rational numbers $\mathbb{Q} = \{r_i : i \in \mathbb{N}\}$ (by Proposition 12.0.13). Now consider the sequence

$$(r_n) = (r_1, r_2, r_3, \ldots).$$

We claim that for any real number x, there is a subsequence (r_{n_k}) of (r_n) such that $\lim_{k\to\infty} r_{n_k} = x$. To see this fact, we know that we can find a rational number r such that $r \in (x, x + 1)$, and that rational number r must equal r_{n_1} for some number $n_1 \in \mathbb{N}$. We claim, inductively, that we can find $n_k \in \mathbb{N}$, for each $k \in \{2, 3, \ldots, \}$, such that $n_k > n_{k-1}$ and $r_{n_k} \in (x, x + \frac{1}{k})$. To prove the inductive step of this claim, suppose n_k as required is given and we need to show the existence of n_{k+1}. The interval $(x, x + \frac{1}{k+1})$ contains infinitely many rational numbers (to verify this fact, choose a rational number $r \in (x, x + \frac{1}{2(k+1)})$ and note that $r + \frac{1}{m(k+1)} \in (x, x + \frac{1}{k+1})$ for all $m \geq 2$). Since $\{r_1, r_2, \ldots, r_{n_k}\}$ is a finite set, it cannot contain all of the rational numbers in $(x, x + \frac{1}{k+1})$, so there must a rational number, which must be of the form $r_{n_{k+1}}$ with $n_{k+1} > n_k$, belonging to $(x, x + \frac{1}{k+1})$. This completes the inductive step, yielding a subsequence (r_{n_k}) of (r_n) satisfying $r_{n_k} \in (x, x + \frac{1}{k})$ for all $k \in \mathbb{N}$. Therefore $|r_{n_k} - x| < \frac{1}{k}$ for all $k \in \mathbb{N}$. It follows then that $\lim_{k\to\infty} r_{n_k} = x$ (because we know that for $\epsilon > 0$, we can find $N \in \mathbb{N}$ such that $\frac{1}{k} < \epsilon$ for all $k > N$, and then we have $|r_{n_k} - x| < \frac{1}{k} < \epsilon$ for all $k > N$). So we have shown that every real number is the limit of a subsequence of (r_n); i.e., the set of subsequential limit points of the particular sequence (r_n) is all of $\mathbb{R}$.

The first useful fact to notice is that if the whole sequence (s_n) converges to s, for some $s \in \mathbb{R}$, then every subsequence of (s_n) converges to s as well.

Proposition 15.0.2 *Suppose (s_n) is a sequence of real numbers, and $s_n \to \ell$, for some $\ell \in \mathbb{R}$. Then for any subsequence (s_{n_k}) of (s_n), we have $\lim_{k\to\infty} s_{n_k} = \ell$.*

Proof. Let $\epsilon > 0$. Since $s_n \to \ell$, there exists $N \in \mathbb{N}$ such that $|s_n - \ell| < \epsilon$ for all $n > N$. Then for all $k > N$, we have $n_k \geq k > N$, so $|s_{n_k} - \ell| < \epsilon$. ■

What interests us most is when a divergent sequence has a convergent subsequence. This situation does not always happen; for example, the sequence $(n) = (1, 2, 3, 4, \ldots)$ does not have any convergent subsequences. The problem with the sequence (n) is that it is unbounded. So let's restrict to bounded sequences and study them a little further.

The lim sup and lim inf of a bounded sequence

Suppose (s_n) is a bounded sequence, say $-M \leq s_n \leq M$ for all $n \in \mathbb{N}$. Let's define a sequence (t_n) as follows: let

$t_1 = \sup\{s_1, s_2, s_3, \ldots\} = \sup\{s_k : k \geq 1\}$;
$t_2 = \sup\{s_2, s_3, s_4, \ldots\} = \sup\{s_k : k \geq 2\}$;
$t_3 = \sup\{s_3, s_4, s_5 \ldots\} = \sup\{s_k : k \geq 3\}$;
and, more generally, for all $k \in \mathbb{N}$,
$t_n = \sup\{s_n, s_{n+1}, s_{n+2} \ldots\} = \sup\{s_k : k \geq n\}$.

We should be careful to check that these suprema exist. However, since $-M \leq s_n \leq M$ for all $n \in \mathbb{N}$, we have in particular that $-M \leq s_k \leq M$ for all $k \geq n$, so the set $\{s_k : k \geq n\}$ is non-empty and bounded above, so it has a supremum. Moreover, we must have $t_n \leq M$ since M is an upper bound for $\{s_k : k \geq 1\}$, and hence also for $\{s_k : k \geq n\}$, and the supremum t_n is smaller than or equal to any other upper bound.

Also we have $t_n \geq -M$ just because $t_n \geq s_n \geq -M$. So the sequence (t_n) is bounded, with $-M \leq t_n \leq M$ for all $k \in \mathbb{N}$.

The critical observation is that, for each $n \in \mathbb{N}$, we have

$$\{s_k : k \geq n+1\} = \{s_{n+1}, s_{n+2}, s_{n+3}, \ldots\} \subseteq \{s_n, s_{n+1}, s_{n+2}, s_{n+3}, \ldots\} = \{s_k : k \geq n\}$$

and hence

$$t_{n+1} = \sup\{s_k : k \geq n+1\} \leq \sup\{s_k : k \geq n\} = t_n,$$

by Example 10.0.6 (the supremum of a subset is less than or equal to the supremum of the containing set). Therefore the sequence (t_n) is decreasing and bounded below. Therefore, by the Monotone Sequence Theorem (Theorem 14.0.8), part 2, the sequence t_n is convergent. We define the *limit supremum*, abbreviated lim sup, of (s_n) to be the limit of (t_n).

Analogously, we let $r_n = \inf\{s_k : k \geq n\}$. Then the sequence (r_n) is increasing and bounded above, so it also has a limit, by the Monotone Sequence Theorem, and we define the *limit inferior*, or "lim inf" of (s_n) to be the limit of (r_n). Here is the formal definition, written more concisely.

Definition 15.0.3 *Suppose (s_n) is a bounded sequence of real numbers. Define*

$$\limsup s_n = \lim_{n\to\infty} (\sup\{s_k : k \geq n\})$$

and

$$\liminf s_n = \lim_{n\to\infty} (\inf\{s_k : k \geq n\}).$$

Some texts use the notation $\underline{\lim}$ for lim inf and $\overline{\lim}$ for lim sup. It certainly can happen that $\limsup s_n = \liminf s_n$; in fact, we will soon see that this equality happens if and only if (s_n) is convergent.

Example 15.0.4 *Let $s_n = 3 + \frac{1}{n}$ for n even and $s_n = 1 - \frac{1}{n}$ for n odd; then*

$$(s_n) = \left(0, 3+\frac{1}{2}, \frac{2}{3}, 3+\frac{1}{4}, \frac{4}{5}, 3+\frac{1}{6}, \frac{6}{7}, 3+\frac{1}{8}, \frac{8}{9}, \ldots\right).$$

Then

$t_1 = \sup\{s_1, s_2, s_3, \ldots\} = \sup\{0, 3+\frac{1}{2}, \frac{2}{3}, 3+\frac{1}{4}, \frac{4}{5}, 3+\frac{1}{6}, \frac{6}{7}, 3+\frac{1}{8}, \frac{8}{9}, \ldots\} = 3+\frac{1}{2}$,
$t_2 = \sup\{s_2, s_3, s_4, \ldots\} = \sup\{3+\frac{1}{2}, \frac{2}{3}, 3+\frac{1}{4}, \frac{4}{5}, 3+\frac{1}{6}, \frac{6}{7}, 3+\frac{1}{8}, \frac{8}{9}, \ldots\} = 3+\frac{1}{2}$,
$t_3 = \sup\{s_3, s_4, s_5 \ldots\} = \sup\{\frac{2}{3}, 3+\frac{1}{4}, \frac{4}{5}, 3+\frac{1}{6}, \frac{6}{7}, 3+\frac{1}{8}, \frac{8}{9}, \ldots\} = 3+\frac{1}{4}$,
$t_4 = \sup\{s_4, s_5, s_6, \ldots\} = \sup\{3+\frac{1}{4}, \frac{4}{5}, 3+\frac{1}{6}, \frac{6}{7}, 3+\frac{1}{8}, \frac{8}{9}, \ldots\} = 3+\frac{1}{4}$,
$t_5 = \sup\{s_5, s_6, s_7, \ldots\} = \sup\{\frac{4}{5}, 3+\frac{1}{6}, \frac{6}{7}, 3+\frac{1}{8}, \frac{8}{9}, \ldots\} = 3+\frac{1}{6}$,

and, more generally, for each $n \in \mathbb{N}$,

$t_{2k-1} = 3 + \frac{1}{2k} = t_{2k}$. *In other words, for n even, $t_n = 3 + \frac{1}{n}$, and for n odd, $t_n = 3 + \frac{1}{n+1}$. That is,*

$$(t_n) = \left(3+\frac{1}{2}, 3+\frac{1}{2}, 3+\frac{1}{4}, 3+\frac{1}{4}, 3+\frac{1}{6}, 3+\frac{1}{6}, 3+\frac{1}{8}, \ldots\right).$$

We can see then that (t_n) is decreasing with limit 3, so

$$\limsup s_n = \lim_{n\to\infty} t_n = 3.$$

Working similarly, we see for $r_n = \inf\{s_k : k \geq n\}$, we have $r_1 = 0, r_2 = \frac{2}{3}, r_3 = \frac{2}{3}, r_4 = \frac{4}{5}, r_5 = \frac{4}{5}$, and, more generally, $r_n = 1 - \frac{1}{n}$ for n odd, and $r_n = 1 - \frac{1}{n+1}$ for n even, or

$$(r_n) = \left(0, \frac{2}{3}, \frac{2}{3}, \frac{4}{5}, \frac{4}{5}, \frac{6}{7}, \frac{6}{7}, \ldots\right).$$

Thus r_n is increasing with limit 1, so

$$\liminf s_n = \lim_{n\to\infty} r_n = 1.$$

Notice that the subsequence $(s_{2n}) = (3 + \frac{1}{2n})$ *of* (s_n) *converges to* $3 = \limsup s_n$*, and the subsequence* $(s_{2n-1}) = (1 - \frac{1}{2n-1})$ *of* (s_n) *converges to* $1 = \liminf s_n$*. We will see that this behavior always occurs: for a bounded sequence, there is a subsequence converging to the* lim sup *and a subsequence converging to the* lim inf*.*

It is possible, if we allow infinite values, to define the lim sup and lim inf of sequences that are not bounded, but for simplicity we prefer not to consider those possibilities at the moment. If we do so, however, then $\limsup s_n$ and $\liminf s_n$ are always defined, whereas $\lim s_n$ is not defined if (s_n) is not convergent. Since we have only defined $\limsup s_n$ and $\liminf s_n$ for bounded sequences so far, we can only say that $\limsup s_n$ and $\liminf s_n$ are always defined for bounded sequences, which is sufficient for our purposes.

One of the reasons for the importance of the lim sup and lim inf is that they put bounds on the possibilities for limits of subsequences of a sequence (s_n). Before stating this result, we require a lemma, which seems obvious, and whose proof (which can be done similarly to the proof of Proposition 14.1) is left as an exercise.

Lemma 15.0.5 *Suppose* (s_n) *and* (t_n) *are convergent sequences, and* $s_n \leq t_n$ *for all* $n \in \mathbb{N}$*. Then*

$$\lim_{n\to\infty} s_n \leq \lim_{n\to\infty} t_n.$$

Theorem 15.0.6 *Suppose* (s_n) *is a bounded sequence of real numbers. Suppose* (s_{n_k}) *is a convergent subsequence of* (s_n)*. Then*

$$\liminf s_n \leq \lim_{k\to\infty} s_{n_k} \leq \limsup s_n.$$

Proof. For each $k \in \mathbb{N}$, we have $r_{n_k} = \inf\{s_{n_k}, s_{n_k+1}, s_{n_k+2}, \ldots\} \leq s_{n_k}$, since the inf of a set is smaller than or equal to any of the set's elements, which in this case includes s_{n_k}. Similarly, $s_{n_k} \leq \sup\{s_{n_k}, s_{n_k+1}, s_{n_k+2}, \ldots\} = t_{n_k}$. That is,

$$r_{n_k} \leq s_{n_k} \leq t_{n_k}, \tag{15.1}$$

for each $k \in \mathbb{N}$. By definition, $\lim_{k\to\infty} r_{n_k} = \liminf s_n$ and $\lim_{k\to\infty} t_{n_k} = \limsup s_n$. Since (s_{n_k}) is assumed to be convergent, taking the limit as $k \to \infty$ in equation (15.1) and applying Lemma 15.0.5 twice gives

$$\liminf s_n = \lim_{k\to\infty} r_{n_k} \leq \lim_{k\to\infty} s_{n_k} \leq \lim_{k\to\infty} t_{n_k} = \limsup s_n.$$

■

The *Bolzano-Weierstrass Theorem* is often stated as: a bounded sequence has a convergent subsequence. The next result includes that fact, but is more precise in that it states that there are subsequences converging to the lim inf and to the lim sup, respectively. This fact makes Theorem 15.0.6 more interesting, because it says that set of subsequential limit points contains its maximum, the lim sup, and its minimum, the lim inf.

The idea of the next proof is fairly natural. To find a subsequence of (s_n) converging to $\limsup s_n$, let $t_n = \sup\{s_n, s_{n+1}, s_{n+2}, \ldots\}$. By definition, the sequence (t_n) converges to $\limsup s_n$. However, for each n, there must be some element of the set $\{s_n, s_{n+1}, s_{n+2}, \ldots\}$ which closely approximates t_n (there must be elements close to the supremum, by Lemma 10.0.7). These close elements will form our subsequence (s_{n_k}). To get convergence, we can choose s_{n_k} closer and closer to t_{n_k} as k gets larger. Since we are choosing s_{n_k} close to t_{n_k}, and t_{n_k} is getting close to $\limsup s_n$ (the subsequence (t_{n_k}) still converges to $\limsup s_n$, by Theorem 15.0.2), then s_{n_k} is getting close to $\limsup s_n$ also. The thing that is tricky is choosing the subsequence elements s_{n_k} to guarantee that the index n_k is increasing, to give a true subsequence. To do that, we have to choose the indices n_k inductively.

Theorem 15.0.7 *(Bolzano-Weierstrass Theorem) Suppose* (s_n) *is a bounded sequence of real numbers. Then* (s_n) *has a subsequence converging to* $\limsup s_n$*. Also* (s_n) *has a subsequence converging to* $\liminf s_n$*.*

Proof. We prove the existence of a subsequence of (s_n) converging to $\limsup s_n$, and leave the analogous argument for the lim inf as an exercise. For each $n \in \mathbb{N}$, let $t_n = sup\{s_n, s_{n+1}, s_{n+2} \ldots\}$ as usual. Then $\lim_{n\to\infty} t_n = \limsup s_n$, by definition.

Since $t_1 = \sup\{s_1, s_2, s_3, \ldots\}$, there exists an index $n_1 \in \mathbb{N}$ such that $s_{n_1} > t_1 - 1$ (by Lemma 10.0.7 with $\epsilon = 1$). Since t_1 is the supremum, we have $s_{n_1} \leq t_1$. So we have $t_1 - 1 < s_{n_1} \leq t_1$. Subtracting t_1 from both inequalities gives, $-1 < s_{n_1} - t_1 \leq 0$, which implies that $|s_{n_1} - t_1| < 1$.

To guarantee that our next index n_2 that we will select satifies $n_2 > n_1$, consider

$$t_{n_1+1} = \sup\{s_{n_1+1}, s_{n_1+2}, s_{n_1+3} \ldots\}.$$

By Lemma 10.0.7 with $\epsilon = \frac{1}{2}$, there exists $n_2 \in \{n_1 + 1, n_1 + 2, n_1 + 3 \ldots\}$ such that $s_{n_2} > t_{n_1+1} - \frac{1}{2}$. We also have $s_{n_2} \leq t_{n_1+1}$, since the supremum is an upper bound. So $t_{n_1+1} - \frac{1}{2} < s_{n_2} \leq t_{n_1+1}$, hence $|s_{n_2} - t_{n_1+1}| < \frac{1}{2}$. Since $n_2 \in \{n_1 + 1, n_1 + 2, n_1 + 3 \ldots\}$, we have $n_2 > n_1$.

Next we consider $t_{n_2+1} = \sup\{s_{n_2+1}, s_{n_2+2}, \ldots\}$ and select $n_3 > n_2$ such that $|s_{n_3} - t_{n_2+1}| < \frac{1}{3}$, and continue in this way. To be more formal, we state our selection inductively, as follows.

We will establish by induction on k that there exist natural numbers n_k for all $k \in \mathbb{N}$ satisfying

$$n_k > n_{k-1} \text{ and } |s_{n_k} - t_{n_{k-1}+1}| < \frac{1}{k}$$

(set $n_0 = 0$ for the case $k = 1$). The case $k = 1$ is proved above. Suppose now that n_k has been determined. Consider

$$t_{n_k+1} = \sup\{s_{n_k+1}, s_{n_k+2}, s_{n_k+3}, \ldots\}.$$

By Lemma 10.0.7 with $\epsilon = \frac{1}{k+1}$, there exists a natural number $n_{k+1} \in \{n_k + 1, n_k + 2, n_k + 3, \ldots\}$ (in other words, $n_{k+1} > n_k$) such that $s_{n_{k+1}} > t_{n_k+1} - \frac{1}{k+1}$. By the definition of the supremum, we automatically have $s_{n_{k+1}} \leq t_{n_k+1}$. So $t_{n_k+1} - \frac{1}{k+1} < s_{n_{k+1}} \leq t_{n_k+1}$. Subtracting t_{n_k+1} from all terms gives $-\frac{1}{k+1} < s_{n_{k+1}} - t_{n_k+1} \leq 0$, which implies $|s_{n_{k+1}} - t_{n_k+1}| < \frac{1}{k+1}$. This argument completes the inductive step. Hence n_k meeting the conditions stated exists for all $k \in \mathbb{N}$.

Note that (s_{n_k}) is a subsequence of (s_n), since $n_k > n_{k-1}$ for all $k \in \mathbb{N}$. Since $|s_{n_k} - t_{n_{k-1}+1}| < \frac{1}{k}$, we have $\lim_{k\to\infty}(s_{n_k} - t_{n_{k-1}+1}) = 0$ (given $\epsilon > 0$, pick $N > \frac{1}{\epsilon}$; then for $k > N$ we have $|s_{n_k} - t_{n_{k-1}+1} - 0| < \frac{1}{k} < \frac{1}{N} < \epsilon$). Note also that $(t_{n_{k-1}+1})_{k=1}^{\infty}$ is a subsequence of (t_n), since $n_k > n_{k-1}$ for each $k \in \mathbb{N}$. Therefore $\lim_{k\to\infty} t_{n_{k-1}+1} = \lim_{n\to\infty} t_n = \limsup s_n$, by Theorem 15.0.2. Hence by Theorem 14.0.4 part (2),

$$\lim_{k\to\infty} s_{n_k} = \lim_{k\to\infty} (s_{n_k} - t_{n_{k-1}+1} + t_{n_{k-1}+1})$$

$$= \lim_{k\to\infty} (s_{n_k} - t_{n_{k-1}+1}) + \lim_{k\to\infty} t_{n_{k-1}+1} = 0 + \limsup s_n = \limsup s_n.$$

Thus (s_{n_k}) is a subsequence of (s_n) converging to $\limsup s_n$. ■

At this point we can clarify the relation between the lim sup and lim inf and the limit of a sequence. First we need a lemma, whose proof is left as an exercise.

Lemma 15.0.8 *(Squeeze Lemma) Suppose $(r_n), (s_n)$, and (t_n) are sequences of real numbers, satisfying $r_n \leq s_n \leq t_n$ for all $n \in \mathbb{N}$. Suppose also that $r_n \to s$ and $t_n \to s$, for some $s \in \mathbb{R}$. Then $s_n \to s$.*

Corollary 15.0.9 *Let (s_n) be a bounded sequence of real numbers. Then (s_n) is convergent if and only if $\limsup s_n = \liminf s_n$, and in that case,*

$$\lim s_n = \limsup s_n = \liminf s_n. \tag{15.2}$$

Proof. First suppose (s_n) is convergent. By Theorem 15.0.7, there is a subsequence (s_{n_k}) converging to $\limsup s_n$. However, by Proposition 15.0.2 any subsequence of (s_n) converges to $\lim_{n\to\infty} s_n$. Hence $\limsup s_n = \lim_{n\to\infty} s_n$. The same argument applies to $\liminf s_n$, establishing $\liminf s_n = \lim_{n\to\infty} s_n$, and hence (15.2).

Conversely, suppose $\limsup s_n = \liminf s_n$. Let $s = \limsup s_n = \liminf s_n$. For each $n \in \mathbb{N}$, let $r_n = \inf\{s_n, s_{n+1}, s_{n+2}, \ldots\}$ and $t_n = \sup\{s_n, s_{n+1}, s_{n+2}, \ldots\}$. Then by definition, $r_n \to \liminf s_n = s$ and $t_n \to \limsup s_n = s$. Also, by the definition of the supremum and infimum of a set, we have $r_n \leq s_n \leq t_n$ for each $n \in \mathbb{N}$. Hence by Lemma 15.0.8, we have that s_n converges to s also. ■

We look back momentarily at an example given in the Section 1, Example 1.1, where we set $x_1 = 2$ and defined x_n recursively by $x_{n+1} = 6 - x_n$. If $\lim_{n\to\infty}$ exists, say $x = \lim_{x\to\infty} x_n$, then taking the limit on both sides of the equation $x_{n+1} = 6 - x_n$ gives $x = 6 - x$ ((x_{n+1}) is a subsequence of (x_n), so converges to x also), hence $x = 3$. But in fact $x_1 = 2, x_2 = 4, x_3 = 2, x_4 = 4$, etc. and so (x_n) is not convergent. This example shows the danger of assuming the convergence of a sequence; in particular writing the expression $\lim x_n$ does not make sense until after the convergence of (x_n) has been proved.

By contrast, if (x_n) is known to be bounded, it always makes sense to write down $\limsup x_n$ and $\liminf x_n$. This point is one of the key advantages of the limit supremum and limit infimum. Moreover, if one can demonstrate that $\limsup x_n = \liminf x_n$, then one concludes that x_n is convergent to the common value $\limsup x_n = \liminf x_n$.

Cauchy sequences

To prove that a sequence (s_n) converges using the definition, we need to know the limit value s in order to estimate $|s_n - s|$. In many cases we don't know s, but we mostly just want to know that (s_n) converges. If the sequence is increasing and bounded above, or decreasing and bounded below, the monotone sequence lemma allows us to conclude the convergence of (s_n), but monotone sequences are very special cases. The lim sup and lim inf might be helpful, because they are somewhat explicit and one can prove convergence of (s_n) by showing the equality of the lim sup and lim inf, but this approach can be difficult to carry out. Another procedure for showing convergence is to consider *Cauchy sequences.*

Definition 15.0.10 *A sequence (s_n) is Cauchy if, for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $|s_n - s_m| < \epsilon$ for all $n, m > N$.*

The terminology $|s_n - s_m| < \epsilon$ for all $n, m > N$ means that if both $n > N$ and $m > N$, then we must have $|s_n - s_m| < \epsilon$. Whereas the definition of (s_n) converging to s is roughly that s_n is getting close to s, a sequence is Cauchy is s_n and s_m are getting close to each other for n and m large enough. If s_n and s_m are always close for large enough n and m, that means that the values s_n are bunching up on the number line. In a sense, if (s_n) is Cauchy (also stated as: "(s_n) is a Cauchy sequence"), then (s_n) is trying to converge. The primary issue is whether the point that (s_n) is trying to converge to, is in the space (look back at the example at the end of Chapter 14 where $s_1 = 1.4, s_2 = 1.41, s_3 = 1.414, s_4 = 1.4141$, etc., that is, s_n is the decimal approximation of $\sqrt{2}$ to n decimal places). Thus we will see that the convergence of Cauchy sequences of real numbers is essentially the same thing as the completeness property of $\mathbb{R}$.

We first observe that any convergent sequence is Cauchy: if the s_n are getting close to a limit s, they must be getting close to each other.

Lemma 15.0.11 *Suppose (s_n) is a convergent sequence of real numbers. Then (s_n) is Cauchy.*

Proof. Suppose (s_n) converges to s, where $s \in \mathbb{R}$. Let $\epsilon > 0$. Since $s_n \to s$, there exists $N \in \mathbb{N}$ such that $|s_n - s| < \frac{\epsilon}{2}$ for all $n > N$. Therefore if $n, m > N$, we have (using the triangle inequality)

$$|s_n - s_m| = |s_n - s + s - s_m| \leq |s_n - s| + |s - s_m| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

■

Notice that adding and subtracting s and using the triangle inequality in the previous proof is the formal way of expressing the intuition that s_n and s_m must become close to each other because they both become close to s.

To prove that a Cauchy sequence converges, you just have to prove that it has a convergent subsequence, according to the following lemma.

Lemma 15.0.12 *Suppose (s_n) is a Cauchy sequence of real numbers. Suppose (s_n) has a subsequence (s_{n_k}) with $\lim_{k\to\infty} s_{n_k} = s$, for some $s \in \mathbb{R}$. Then (s_n) is convergent with $\lim_{n\to\infty} s_n = s$.*

PROOF. Let $\epsilon > 0$. Since (s_n) is Cauchy, there exists $N \in \mathbb{N}$ such that $|s_n - s_m| < \frac{\epsilon}{2}$ if $n, m > N$. Since $\lim_{k\to\infty} s_{n_k} = s$, there exists $N_1 \in \mathbb{N}$ such that $|s_{n_k} - s| < \frac{\epsilon}{2}$ if $k > N_1$. Choose $k > \max(N, N_1)$. Then $n_k \geq k > \max(N, N_1)$. Hence for all $n > N$, we have

$$|s_n - s| = |s_n - s_{n_k} + s_{n_k} - s| \leq |s_n - s_{n_k}| + |s_{n_k} - s| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

(The estimate $|s_n - s_{n_k}| < \frac{\epsilon}{2}$ holds because $n > N$ and $n_k \geq k > N$, whereas the estimate $|s_{n_k} - s| < \frac{\epsilon}{2}$ holds because we choose $k > N_1$.) ■

The previous lemma supports that idea that a Cauchy sequence is trying to converge, but the issue is whether the limit is in the space. Once a subsequence is known to converge, then we know that the potential limit of the entire sequence is in the space.

Her is another key property of Cauchy sequences.

Lemma 15.0.13 *A Cauchy sequence is bounded.*

PROOF. Let (s_n) be a Cauchy sequence. By the definition of a Cauchy sequence with $\epsilon = 1$, there exists $N \in \mathbb{N}$ such that $|s_n - s_m| < 1$ for all $n, m > N$. Letting $m = N + 1$, we have that if $n > N$, then

$$|s_n| = |s_n - s_{N+1} + s_{N+1}| \leq |s_n - s_{N+1}| + |s_{N+1}| < 1 + |s_{N+1}|.$$

This bounds all but finitely many of the terms s_n, so we can just add in the remaining values and observe that

$$|s_n| \leq \max\{|s_1|, |s_2|, \ldots, |s_N|, 1 + |s_{N+1}|\},$$

where $1 + |s_{N+1}|$ is a bound if $n > N$, and the right side is trivially a bound if $n \leq N$. ■

We now have all of the ingredients to deduce something remarkable.

Theorem 15.0.14 *(Cauchy Criterion) Let (s_n) be a sequence of real numbers. Then (s_n) is convergent if and only if (s_n) is Cauchy.*

PROOF. The fact that a convergent sequence is Cauchy is Lemma 15.0.11. Now suppose (s_n) is Cauchy. By Lemma 15.0.13, (s_n) is bounded. By the Bolzano-Weierstrass Theorem (Theorem 15.0.7), (s_n) has a convergent subsequence. Hence by Lemma 15.0.12, (s_n) is convergent. ■

Thus one way to prove that a sequence is convergent is to prove that it is Cauchy, which we may be able to do without knowing the limit of the sequence explicitly. Just to keep track of the logic, the hard work in proving the Cauchy Criterion was done by the Bolzano-Weierstrass Theorem. The proof of the Bolzano-Weierstrass theorem depended on knowing that the lim sup and the lim inf of a bounded sequence exist, which relied on the monotone sequence lemma. In turn, the monotone sequence lemma follows from the completeness property of the real numbers (the fact that every bounded above set has a supremum in the real numbers).

If we go further in analysis and extend beyond the one dimensional setting of $\mathbb{R}$, such as $\mathbb{R}^n$ for $n \geq 2$ or infinite dimensional spaces like those referred to in Examples 1.0.2 and 1.0.3, the underlying space no longer has an order, so the concept of the supremum of a set of elements does not make sense for them. Thus the idea that the space is "complete" or has no "missing points" cannot be expressed in terms of the existence of suprema. Instead, the notion of a Cauchy sequence is used to define completeness: a space is complete if every Cauchy sequence in the space converges.

Also, one of the ways of constructing the real numbers from the rationals is via Cauchy sequences. Roughly, one looks at all Cauchy sequences of rational numbers (these are the sequences that are trying to converge to a real number, intuitively). One regards two Cauchy sequences (s_n) and (t_n) as equivalent if the sequence $(s_1, t_1, s_2, t_2, s_3, t_3, \ldots)$ obtained by intertwining (s_n) and (t_n) is Cauchy; intuitively that means that (s_n) and (t_n) are trying to converge to the same number. The relation of being equivalent in this case is in fact an equivalence relation, and one can identify $\mathbb{R}$ with the collection of equivalence classes under this relation. In other words, we define $\mathbb{R}$ by identifying each real number x with the set of all Cauchy sequences of rational numbers which are trying to converge to x. Carrying out this process and verifying that all of the properties of $\mathbb{R}$ are obtained as consequences is quite lengthy, and we won't carry it out. However, we remark that this approach has other applicatons: it can be applied more generally to infinite dimensional spaces that are not complete to form their *completion*.

Chapter 16

Open and Closed Sets

We call an interval of the form $(a,b) = \{x \in \mathbb{R} : a < x < b\}$ (or (a,∞) or $(-\infty,b)$ or $(-\infty,\infty) = \mathbb{R}$) *open intervals*, and we call an interval of the form $[a,b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ (or $(-\infty,b]$ or $[a,\infty)$ or $(-\infty,\infty) = \mathbb{R}$) *closed intervals*. In this section we generalize to consider *open* sets and *closed* sets.

Open Sets

Definition 16.0.1 *Let A be a subset of $\mathbb{R}$. Then A is open if, for each $x \in A$, there exists $r > 0$ such that $(x-r, x+r) \subseteq A$.*

In this definition, $(x-r, x+r)$ is, as usual, the interval $\{t \in \mathbb{R} : x - r < t < x + r\}$. It is important to understand that the $r > 0$ in the definition depends on x, that is, $r = r(x)$. This point is understood logically because the existence of r is stated after x is introduced, so implicitly it is understood that r may be different for different x. The set $(x-r, x+r)$ is the same as $\{t \in \mathbb{R} : |x - t| < r\}$, which is sometimes denoted $B(x,r)$ or $B_r(x)$, and called the *ball of radius r around x*. A set A is open if every point in A has a ball (of some positive radius) around it which is contained in A.

One reason that open sets are important is that they are the natural domains to consider when we study derivatives of functions. For the derivative of a function f to exist at a point x, we need the difference quotients $\frac{f(x+h)-f(x)}{h}$ to exist for all h in some interval $(-r,r)$ around 0. Therefore we need $f(x+h)$ to be defined for all $h \in (-r,r)$, so we need $(x-r, x+r)$ to be in the domain of f. So the domain of a differentiable function should be an open set.

Example 16.0.2 *The interval $[0,1)$ is not open.*

Proof. Let $x = 0$ and $r > 0$. Then $(x-r, x+r) = (0-r, 0+r) = (-r,r) \not\subseteq [0,1)$ since, for example, the point $-r/2 \in (-r,r)$ but $-r/2 \notin [0,1)$. So no ball of positive radius around $0 \in [0,1)$ is contained in $[0,1)$. Hence $[0,1)$ is not open. ■

Notice that to prove that the set in the last example is not open, we only had to show that the criterion in the definition of an open set fails at the single point $x = 0$. In fact, that criterion is met at all other points of $[0,1)$.

Intuitively, an open set is one that does not contain its boundary points. The set $[0,1)$ is not open because it contains one of its boundary points, namely 0. On the other hand, open intervals do not contain either of their boundary points, and hence are open, as the hext example shows.

Example 16.0.3 *Suppose $a, b \in \mathbb{R}$ and $a < b$. Then the interval (a,b) is open.*

Proof. Let $x \in (a,b)$. Then $x > a$, so $x - a > 0$, and $x < b$, so $b - x > 0$. Define $r = \min(x-a, b-x)$. Then $r > 0$ because r is the minimum of two strictly positive quantities. We claim that $(x-r, x+r) \subseteq (a,b)$.

To prove the claim, suppose $y \in (x-r, x+r)$. Then $x - r < y < x + r$, hence $-r < y - x < r$. Therefore, using the fact that $r \leq b - x$ by definition of r, we have

$$y = y - x + x < r + x \leq b - x + x = b,$$

so $y < b$. Also, since $r \leq x - a$, we have $-r \geq a - x$, so

$$y = y - x + x > -r + x \geq a - x + x = a,$$

so $y > a$. Hence $y \in (a, b)$. This proves the claim. Thus for each $x \in (a, b)$, we have found $r > 0$ such that $(x - r, x + r) \subseteq (a, b)$. Hence (a, b) is open. ■

Although the proof given is the most straightforward for an interval contained in $\mathbb{R}$, it is instructive to see a proof using the notation for balls above. We first notice that if $c = \frac{a+b}{2}$ is the midpoint of (a, b), and $R = \frac{b-a}{2}$ is the distance from c to either a or b, then $(a, b) = B(c, R)$. This fact is geometrically clear, and we leave the formal proof to the reader. To prove that $B(c, R)$ is open, let $x \in B(c, R)$. Then $|x - c| < R$, by definition of $B(c, R)$. Let $r = R - |x - c|$, so that $r > 0$ and $|x - c| = R - r$. If $y \in B(x, r)$, then $|y - x| < r$, so, by the triangle inequality,

$$|y - c| = |y - x + x - c| \leq |y - x| + |x - c| < r + R - r = R.$$

Thus $y \in B(c, R)$. We have shown that $B(x, r) \subseteq B(c, R)$. Since $x \in B(c, R)$ is arbitrary, $B(c, R)$ is open. Notice how the triangle inequality handles both of the inequalities in the first proof simultaneously. The second proof is more transparent (especially if you draw a picture to motivate the choice of r), and extends to $\mathbb{R}^n$ and, more generally, to metric spaces.

Note that the entire real line $\mathbb{R}$ is open (take any $r > 0$ at any $x \in \mathbb{R}$), and $\emptyset$ is open, because the condition "for every $x \in \emptyset$, there exists $r > 0$ such that $(x - r, x + r) \subseteq \emptyset$" is vacuously true, because there are no $x \in \emptyset$.

It is important to understand how open sets behave under unions and intersections. The next result says that an arbitrary union of open sets is open.

Proposition 16.0.4 *Let Λ be any set, and suppose that O_λ is an open subset of $\mathbb{R}$ for all $\lambda \in \Lambda$. Then $\cup_{\lambda \in \Lambda} O_\lambda$ is open.*

Proof. Let $x \in \cup_{\lambda \in \Lambda} O_\lambda$. Then there exists some $\lambda \in \Lambda$, let's call it λ_x, such that $x \in O_{\lambda_x}$. Since O_{λ_x} is open, there exists $r > 0$ such that

$$(x - r, x + r) \subseteq O_{\lambda_x} \subseteq \cup_{\lambda \in \Lambda} O_\lambda.$$

Hence $(x - r, x + r) \subseteq \cup_{\lambda \in \Lambda} O_\lambda$. Therefore $\cup_{\lambda \in \Lambda} O_\lambda$ is open. ■

The next result states that a finite intersection of open sets is open.

Proposition 16.0.5 *Let $n \in \mathbb{N}$ and suppose that $O_1, O_2, \ldots, O_n$ are open subsets of $\mathbb{R}$. Then $\cap_{i=1}^n O_i$ is open.*

Proof. Let $x \in \cap_{i=1}^n O_i$. Then for each $i \in \{1, 2, \ldots, n\}$, we have $x \in O_i$. Since O_i is open, there exists $r_i > 0$ such that $(x - r_i, x + r_i) \subseteq O_i$. Let $r = \min\{r_1, r_2, \ldots, r_n\}$. Note that $r > 0$ because r is the minimum of finitely many positive numbers. Then $r \leq r_i$ for all $i = 1, 2, \ldots, n$, so

$$(x - r, x + r) \subseteq (x - r_i, x + r_i) \subseteq O_i.$$

Thus $(x - r, x + r) \subseteq O_i$ for all $i = 1, 2, \ldots, n$, so $(x - r, x + r) \subseteq \cap_{i=1}^n O_i$. Hence $\cap_{i=1}^n O_i$ is open. ■

It is not true that countable intersections of open sets are necessarily open. For an example, note that $(-\frac{1}{i}, 1)$ is open for any $i \in \mathbb{N}$ by Example 16.0.3, but $\cap_{i=1}^\infty (-\frac{1}{i}, 1) = [0, 1)$ is not open by Example 16.0.2.

The notion of open sets can be formulated in the general context of metric spaces, just by replacing intervals in the definition of an open set by balls determined by the metric. There is an even more general notion of a *topological space*, which includes all metric spaces as examples. A topological space is a set X

together with a collection $\mathcal{A}$ of subsets of X which satisfies the properties: (i) $\emptyset, X \in \mathcal{A}$, (ii) any union of elements of $\mathcal{A}$ belongs to $\mathcal{A}$, and (iii) any finite intersection of elements of $\mathcal{A}$ belongs to $\mathcal{A}$. Thus the properties we have derived for open sets in metric spaces becomes the definition of a collection of open sets in topological spaces. For most applications in analysis, it is sufficient to consider metric spaces.

Closed Sets

Definition 16.0.6 *A subset E of $\mathbb{R}$ is closed if every sequence (x_n) of points of E which converges (in $\mathbb{R}$) satisfies $\lim_{n\to\infty} x_n \in E$.*

In other words, E is closed if E contains all of the limits of convergent sequences of E.

Example 16.0.7 *The interval $(0, 1]$ is not closed, because $\frac{1}{n} \in E$ for every $n \in \mathbb{N}$, and the sequence $(1/n)$ converges to 0, but $0 \notin E$.*

Note that $(0, 1]$ is not open either, because there is no interval $(1 - r, 1 + r)$, for $r > 0$, around the point $1 \in (0, 1]$, that is contained in $(0, 1]$. Sets don't have to be either open or closed.

The next result shows that closed intervals are, in fact, closed sets, just as Example 16.0.3 showed that open intervals are open sets.

Example 16.0.8 *Let $a, b \in \mathbb{R}$ with $a < b$. Then the interval $[a, b]$ is a closed set.*

Proof. Suppose (x_n) is a sequence with $x_n \in [a, b]$ for all $n \in \mathbb{N}$, and suppose (x_n) converges to some x. Since $a \le x_n \le b$ for all $n \in \mathbb{N}$, then $a \le x = \lim_{n\to\infty} x_n \le b$ (by direct proof, or, for example, by Lemma 15.0.5, with one of the sequences (s_n) or (t_n) being constant). Hence $x \in [a, b]$. ■

Whereas open sets don't contain their boundary points, closed sets must contain their boundary points. We note that the entire real line $\mathbb{R}$ is closed (trivially, since any limit point x of any sequence must be a real number), and the empty set is closed, because it satisfies the criterion in the definition vacuously, since there are no sequences of elements of the empty set. Also a single point $\{x\}$ in $\mathbb{R}$ is a closed set, because, if $x_n \in \{x\}$ for all $n \in \mathbb{N}$, then $x_n = x$ for all $n \in \mathbb{N}$, so $\lim_{n\to\infty} x_n = x \in \{x\}$.

Whereas open sets are the natural sets to consider for defining derivatives, we will eventually see that closed intervals are the natural sets on which to define the Riemann integral of a function.

Let's consider how closed sets behave under unions and intersections. Whereas arbitrary unions of open sets are open, closed sets have the property that an arbitrary intersection of closed sets is closed.

Proposition 16.0.9 *Suppose Λ is any set, and E_λ is a closed subset of $\mathbb{R}$, for each $\lambda \in \Lambda$. Then $\cap_{\lambda\in\Lambda} E_\lambda$ is closed.*

Proof. Suppose $x_n \in \cap_{\lambda\in\Lambda} E_\lambda$ for every $n \in \mathbb{N}$ and $x_n \to x$, for some $x \in \mathbb{R}$. For any $\lambda' \in \Lambda$, we have $x_n \in E_{\lambda'}$ for all $n \in \mathbb{N}$, since $x_n \in \cap_{\lambda\in\Lambda} E_\lambda$. Since $E_{\lambda'}$ is closed and $x_n \to x$, we have $x \in E_{\lambda'}$. Since this inclusion holds for every $\lambda' \in \Lambda$, we have $x \in \cap_{\lambda\in\Lambda} E_\lambda$. Hence we have shown that $\cap_{\lambda\in\Lambda} E_\lambda$ is closed. ■

The next result states that a finite union of closed sets is closed.

Proposition 16.0.10 *Suppose $E_1, E_2, \ldots, E_n$ are closed subsets of $\mathbb{R}$, for some $n \in \mathbb{N}$. Then $\cup_{i=1}^n E_i$ is closed.*

Proof. Suppose (x_k) is a sequence of points such that $x_k \in \cup_{i=1}^n E_i$ for all $k \in \mathbb{N}$, and $x_k \to x$. Let

$$A_i = \{k \in \mathbb{N} : x_k \in E_i\},$$

for each $i = 1, 2, \ldots, n$. Then $\cup_{i=1}^n A_i = \mathbb{N}$ (for each $k \in \mathbb{N}$ x_k belongs to E_i, hence k belongs to A_i, for some i). Since $\mathbb{N}$ is infinite, then at least one of the sets A_i is infinite. Let $i_0 \in \{1, 2, \ldots, n\}$ be such that A_{i_0} is infinite. Letting $k_1 = \min A_{i_0}, k_2 = \min A_{i_0} \setminus \{k_1\}$, and , inductively $k_{i+1} = \min A_{i_0} \setminus \{k_1, k_2, \ldots, k_i\}$, we obtain

$$A_{i_0} = \{x_{k_1}, x_{k_2}, x_{k_3}, \ldots\} = \{x_{k_i}\}_{i=1}^\infty,$$

with $k_i < k_{i+1}$ for all $i \in \mathbb{N}$. Therefore (x_{k_i}) is a subsequence of (x_k) such that $x_{k_i} \in E_{i_0}$ for all $i \in \mathbb{N}$. Since (x_k) converges to x, we also have that (x_{k_i}) converges to x (by Proposition 15.0.2). Since E_{i_0} is closed, and $x_{k_i} \in E_{i_0}$ for all i, we conclude that $x \in E_{i_0}$, and hence $x \in \cup_{i=1}^{n} E_i$. ■

It is not true that a countable union of closed sets is closed. For example, for each $n \in \mathbb{N}$, the interval $[\frac{1}{n}, 1]$ is closed (by Example 16.0.8), but $\cup_{n=1}^{\infty}[\frac{1}{n}, 1] = (0, 1]$ is not closed (Example 16.0.7).

Relation Between Open and Closed Sets

Open and closed sets are intuitively complementary in the sense that closed sets don't contain their boundary points but open sets do. In fact, the relation between open and closed sets is complementary in the more precise sense that a set is open if and only if its complement is closed.

Proposition 16.0.11 *Let O be an open subset of $\mathbb{R}$. Then $O^c = \mathbb{R} \setminus O$ is closed.*

PROOF. Suppose (x_n) is a sequence with $x_n \in O^c$ for each $n \in \mathbb{N}$, with $x_n \to x$. For each $r > 0$, there exists $n \in \mathbb{N}$ such that $x_n \in (x - r, x + r)$ (in fact, this property holds for all $n > N$, for some $N \in \mathbb{N}$, by the definition of sequence convergence). Since $x_n \in O^c$, this fact guarantees that $(x - r, x + r)$ is not a subset of O. Hence $x \notin O$, since every point in O has such an interval around x which is contained in O, since O is open. Hence $x \in O^c$. Thus O^c is closed. ■

We also have the following.

Proposition 16.0.12 *Let E be a closed subset of $\mathbb{R}$. Then $E^c = \mathbb{R} \setminus E$ is open.*

PROOF. Let $x \in E^c$. We claim that there exists $n \in \mathbb{N}$ such that $(x - \frac{1}{n}, x + \frac{1}{n}) \subseteq E^c$. We establish this claim by contradiction: if not, then for all $n \in \mathbb{N}$, there exists $x_n \in E$ such that $x_n \in (x - \frac{1}{n}, x + \frac{1}{n})$. Since $x_n \in (x - \frac{1}{n}, x + \frac{1}{n})$, we have $|x_n - x| < \frac{1}{n}$, and hence $x_n \to x$. Since $x_n \in E$ for all $n \in \mathbb{N}$ and E is closed, we obtain $x \in E$, which contradicts our assumption that $x \in E^c$. This contradiction establishes the claim that $(x - \frac{1}{n}, x + \frac{1}{n}) \subseteq E^c$, for some $n \in \mathbb{N}$. Since $x \in E^c$ was arbitrary, E^c is open. ■

Putting Propositions 16.0.11 and 16.0.12 together yields the following.

Proposition 16.0.13 *Let A be a subset of $\mathbb{R}$. Then A is open if and only if A^c is closed. Equivalently, for a subset B of $\mathbb{R}$, B is closed if and only if B^c is open.*

PROOF. If A is open, then A^c is closed, by Proposition 16.0.11. Conversely, if A^c is closed, then $A = (A^c)^c$ is open by Proposition 16.0.12. Similarly, if B is closed, then B^c is open by Proposition 16.0.12, whereas if B^c is open, then $B = (B^c)^c$ is closed by Proposition 16.0.11. ■

Propositions 16.0.11 and 16.0.12 can be used to derive facts about closed sets from corresponding facts about open sets, and vice versa. For example, given the fact that an arbitrary union of open sets is open (Proposition 16.0.4), we can deduce the fact that an arbitrary intersection of closed sets is closed (Proposition 16.0.9), as follows. Suppose E_λ is closed, for every $\lambda \in \Lambda$. Then E_λ^c (meaning $(E_\lambda)^c$) is open, by Proposition 16.0.12. Hence $\cup_{\lambda\in\Lambda} E_\lambda^c$ is open, by Proposition 16.0.4. Hence $(\cup_{\lambda\in\Lambda} E_\lambda^c)^c$ is closed, by Proposition 16.0.11. By De Morgan's law,

$$(\cup_{\lambda\in\Lambda} E_\lambda^c)^c = \cap_{\lambda\in\Lambda}(E_\lambda^c)^c = \cap_{\lambda\in\Lambda} E_\lambda,$$

so $\cap_{\lambda\in\Lambda} E_\lambda$ is closed. Similarly, we could prove Proposition 16.0.4 is we assume Proposition 16.0.9. In the same way, Proposition 16.0.5 is equivalent to Proposition 16.0.10.

In the general setting of topological spaces, where the collection of open sets is given at the start, a set is defined to be closed if its complement is open.

The Closure and Interior of a Set

Given a set $A \subseteq \mathbb{R}$, we can form a minimal closed set containing A, called $\overline{A}$, the *closure* of A.

Definition 16.0.14 *For $A \subseteq \mathbb{R}$, let*

$$\overline{A} = \{x \in \mathbb{R} : \text{ there exists a sequence } (x_n) \text{ with } x_n \in A \text{ for all } n \in \mathbb{N}, \text{ and } x_n \to x\}.$$

That is, $\overline{A}$ consists of all limits of sequences from A. We will see, for example, that if $a < b$ and $A = (a, b)$, then $\overline{A} = [a, b]$. The key properties of the closure are described in the following Proposition.

Proposition 16.0.15 *Suppose $A \subseteq \mathbb{R}$. Then*

(i) $A \subseteq \overline{A}$;

(ii) $\overline{A}$ is closed;

(iii) $A = \overline{A}$ if and only if A is closed;

(iv) $\overline{\overline{A}} = \overline{A}$;

(v) If F is a closed subset of $\mathbb{R}$ and $A \subseteq F$, then $\overline{A} \subseteq F$;

(vi) $\overline{A} = \cap\{F \subseteq \mathbb{R} : F \text{ is closed and } A \subseteq F\}$.

Proof. (i) Let $x \in A$. Let $x_n = x$ for all $n \in \mathbb{N}$. The each x_n belongs to A and $x_n \to x$. Therefore $x \in \overline{A}$.

(ii) Suppose $x_n \in \overline{A}$ for each $n \in \mathbb{N}$, and $x_n \to x$. We need to show that $x \in \overline{A}$, which is not obvious because we only have $x_n \in \overline{A}$, not $x_n \in A$. Since $x_n \in \overline{A}$, there exists a sequence of elements of A which converge to x_n. By going out far enough in that sequence, we can find an element of A, call it y_n, satisfying $|y_n - x_n| < \frac{1}{n}$. Using the triangle inequality,

$$|y_n - x| = |y_n - x_n + x_n - x| \leq |y_n - x_n| + |x_n - x| < \frac{1}{n} + |x_n - x| \to 0,$$

as $n \to \infty$. Hence $y_n \to x$. Since (y_n) is a sequence of elements in A and $y_n \to x$, we have $x \in \overline{A}$.

(iii) If $A = \overline{A}$ then A is closed since $\overline{A}$ is closed by (ii). Conversely, suppose A is closed. We claim $A = \overline{A}$. We have $A \subseteq \overline{A}$ by (i). To prove $\overline{A} \subseteq A$, suppose $x \in \overline{A}$. By definition of closure, then, there exists a sequence (x_n) with $x_n \in A$ for all $n \in \mathbb{N}$, and $x_n \to x$. But since A is closed, we have $x \in A$. Hence $\overline{A} \subseteq A$, and so altogether we have $A = \overline{A}$.

(iv) Since $\overline{A}$ is closed, by (ii), applying (iii) to $\overline{A}$ gives $\overline{A} = \overline{\overline{A}}$.

(v) Suppose F is a closed subset of $\mathbb{R}$ and $A \subseteq F$. To show $\overline{A} \subseteq F$, let $x \in \overline{A}$. Then there exists a sequence (x_n) with $x_n \in A$ for all $n \in \mathbb{N}$, such that $x_n \to x$. Then $x_n \in F$ for all $n \in \mathbb{N}$ (since $A \subseteq F$), and F is closed, so $x \in F$. Hence $\overline{A} \subseteq F$.

(vi) We first prove that $\overline{A} \subseteq \cap\{F \subseteq \mathbb{R} : F \text{ is closed and } A \subseteq F\}$. Suppose $F \subseteq \mathbb{R}$ is closed and $A \subseteq F$. By (v), $\overline{A} \subseteq F$. Since $A \subseteq F$ holds for all closed F such that $A \subseteq F$, we obtain $\overline{A} \subseteq \cap\{F \subseteq \mathbb{R} : F \text{ is closed and } A \subseteq F\}$.

Now we show that $\cap\{F \subseteq \mathbb{R} : F \text{ is closed and } A \subseteq F\} \subseteq \overline{A}$. To see this fact, note that $\overline{A}$ is closed by (ii) and $A \subseteq \overline{A}$ by (i), so $\overline{A}$ is one of the sets F in the intersection. The intersection of sets is always a subset of any set in the intersection, so $\cap\{F \subseteq \mathbb{R} : F \text{ is closed and } A \subseteq F\} \subseteq \overline{A}$. ■

The last proposition justifies the statement that $\overline{A}$ is the minimal closed set containing A: by (i) and (ii), $\overline{A}$ is a closed set containing A, and by (v), any closed set F containing A contains $\overline{A}$, hence is larger than (or equal to) $\overline{A}$.

Condition (vi) in Proposition 16.0.15 is used to define the closure of a set in a topological space. That characterization can often be used to simplify proofs involving closures. For example, consider the following simple result.

Proposition 16.0.16 *Suppose A and B are subsets of $\mathbb{R}$ and $A \subseteq B$. Then $\overline{A} \subseteq \overline{B}$.*

PROOF. (1^{st} proof) Let $x \in \overline{A}$. Then there exists a sequence (x_n) with $x_n \in A$ for all $n \in \mathbb{N}$, such that $x_n \to x$. Since $A \subseteq B$ by assumption, we have $x_n \in B$ for all $n \in \mathbb{N}$. Since $x_n \to x$, we have $x \in \overline{B}$. Therefore $\overline{A} \subseteq \overline{B}$.

(2^{nd} proof) We have $A \subseteq B$ by assumption, and $B \subseteq \overline{B}$ by Proposition 16.0.15 (i). So $A \subseteq \overline{B}$, and $\overline{B}$ is closed by Proposition 16.0.15 (ii). By Proposition 16.0.15, it follows that $\overline{A} \subseteq \overline{B}$. ■

The first proof is direct and intuitive, using the definition of the closure, but the second proof is clean and elegant.

Parallel to the notion of the closure of a set, which has to do with closed sets, there is the notion of the interior of a set, which is related to open sets. However, whereas the closure of a set is a larger set, the interior of a set A is a subset of A (and may be empty).

Definition 16.0.17 *For a subset A of $\mathbb{R}$, define*

$$A^\circ = \{x \in A : \text{ there exists } r > 0 \text{ such that } (x - r, x + r) \subseteq A\}.$$

We call A°, the *interior* of A. We leave the proof of the following Proposition as an exercise.

Proposition 16.0.18 *Let A be a subset of $\mathbb{R}$. Prove that*

(i) A° is open;

(ii) A is open if and only if $A = A^\circ$;

(iii) $(A^\circ)^\circ = A^\circ$;

(iv) if $B \subseteq A$ and B is open, then $B \subset A^\circ$;

(v) $A^\circ = \cup\{B : B \text{ is open and } B \subseteq A\}$.

Chapter 17

Compact Sets

Compactness plays a critical role in analysis. It is a subtle concept. We will consider two versions of the idea, which turn out to be equivalent in the context of $\mathbb{R}$: *open cover compactness* and *sequential compactness*. We will follow the usual terminology, using the term *compact* to mean open cover compact, only using the phrase "open cover compact" when we want to emphasize which notion we are using.

We will apply compactness when we study continuous functions. It will turn out that if the domain of a continuous function is a compact set, then the function must have properties that would not necessarily hold if its domain is not compact. We will see that a continuous function, whose domain is a compact set, must be bounded, must attain its maximum and minimum values, and must be uniformly continuous (we haven't defined continuity or uniformly continuity yet).

We start with sequential compactness.

Sequential Compactness

The Bolzano-Weierstass Theorem (Theorem 15.0.7) tells us that a bounded sequence of real numbers has a convergent subsequence. Considerations along these lines lead to the next definition.

Definition 17.0.1 *A set $K \subseteq \mathbb{R}$ is sequentially compact if, for every sequence (x_n) with $x_n \in K$ for all $n \in \mathbb{N}$, there exists a subsequence (x_{n_k}) which converges to a point $x \in K$.*

It is important to understand that the definition of sequential compactness of a set K requires both that any sequence (x_n) of points of K has a convergent subsequence, and that the limit of that convergent subsequence belongs to K.

Example 17.0.2 *Let $a, b \in \mathbb{R}$ with $a < b$. Then $[a, b]$ is sequentially compact.*

Proof. Suppose (x_n) is a sequence of points with $x_n \in [a, b]$ for each $n \in \mathbb{N}$. Since $x_n \in [a, b]$ for all $n \in \mathbb{N}$, the sequence (x_n) is bounded. By the Bolzano-Weierstrass Theorem (Theorem 15.0.7), (x_n) has a convergent subsequence (x_{n_k}). Let $\lim_{k\to\infty} x_{n_k} = x$. Since $x_{n_k} \in [a, b]$ for all $k \in \mathbb{N}$ and $[a, b]$ is closed (Example 16.0.8), we have $x \in [a, b]$. So $[a, b]$ is sequentially compact. ■

Example 17.0.3 *Let $a, b \in \mathbb{R}$ with $a < b$. Then $(a, b]$ is not sequentially compact.*

Proof. For $n \in \mathbb{N}$, let $x_n = a + \frac{b-a}{n}$. Then $x_n \in (a.b]$ for all $n \in \mathbb{N}$ (because $a + \frac{b-a}{n} > a$ and $a + \frac{b-a}{n} \leq a + b - a = b$). Note that (x_n) converges to a (because $\lim_{n\to\infty} \frac{b-a}{n} = (b-a)\lim_{n\to\infty} \frac{1}{n} = 0$). let (x_{n_k}) be any subsequence of (x_n). Then $\lim_{k\to\infty} x_{n_k} = a$ (by Proposition 15.0.2). But $a \notin (a, b]$, so (x_n) is a sequence of elements of $(a, b]$ which has no subsequence converging to an element of $(a, b]$ (since all such subsequences converge to a). ■

The property of being sequentially compact can be reduced to two apparently simpler properties.

Theorem 17.0.4 *Let E be a subset of $\mathbb{R}$. Then E is sequentially compact if and only if E is closed and bounded.*

PROOF. First suppose E is sequentially compact. We first show that E is closed. Suppose (x_n) is a sequence of points with $x_n \in E$ for all $n \in \mathbb{N}$, such that (x_n) converges to some $x \in \mathbb{R}$. We need to show that $x \in E$. Since E is sequentially compact, (x_n) has a subsequence converging to a point $y \in E$. But any subsequence of (x_n) must converge to x (by Proposition 15.0.2, because (x_n) converges to x). Thus $y = x$, and since $y \in E$, we have $x \in E$.

We now show that the sequential compactness of E implies that E is bounded. We prove the contrapositive: if E is unbounded, then E is not sequentially compact. Suppose E is unbounded. Then in particular, $E \not\subseteq [-n, n]$, for each $n \in \mathbb{N}$, so there exists $x_n \in E$ satisfying $|x_n| > n$. So (x_n) is a sequence in E. Let (x_{n_k}) be any subsequence of (x_n). Then $|x_{n_k}| \geq n_k \geq k$, so the sequence (x_{n_k}) is unbounded and hence divergent (Proposition 14.0.3). Thus (x_n) is a sequence of points of E which has no subsequence converging to a point of E (or to any other point, in this case). So E is not sequentially compact.

We now show that if E is closed and bounded, then E is sequentially compact. Let (x_n) be any sequence of points satifying $x_n \in E$ for all $n \in \mathbb{N}$. Since E is bounded, the sequence (x_n) is bounded. By the Bolzano-Weierstrass Theorem (Theorem 15.0.7), (x_n) has a subsequence (x_{n_k}) converging to some point $x \in \mathbb{R}$. But since E is closed and $x = \lim_{k\to\infty} x_{n_k}$, we conclude that $x \in E$. Thus (x_n) has a subsequence converging to a point of E, so E is sequentially compact. ■

Open Cover Compactness

Sequential compactness is defined in terms of the convergence of sequences (or subsequences), which is closely related to closed sets. Because of the close relation between open and closed sets (i.e., A is open if and only if A^c is closed, by Proposition 16.0.13), it should not be surprising that there is a second notion of compactness that is formulated in terms of open sets. We start with the definition of an open cover.

Definition 17.0.5 *Let A be a subset of $\mathbb{R}$. An open cover of A is a collection of open sets O_λ of $\mathbb{R}$, for $\lambda \in \Lambda$, where Λ is some index set, such that*

$$A \subseteq \cup_{\lambda\in\Lambda} O_\lambda.$$

If $A \subseteq \cup_{\lambda\in\Lambda} O_\lambda$, we say that $\{O_\lambda\}_{\lambda\in\Lambda}$ covers A.

In other words, an open cover of a set A is just a collection of open sets whose union contains A. These open sets are subsets of $\mathbb{R}$; they are not required to be subsets of A.

Example 17.0.6 *Let $A = [0, 1]$. Then $\left\{\left(x - \frac{1}{10}, x + \frac{1}{10}\right)\right\}_{x\in[0,1]}$ is an open cover of A, because, for each $x_0 \in [0, 1]$, we have*

$$x_0 \in \left(x_0 - \frac{1}{10}, x_0 + \frac{1}{10}\right) \subseteq \bigcup_{x\in[0,1]} \left(x - \frac{1}{10}, x + \frac{1}{10}\right).$$

Example 17.0.7 *Let $A = (0, 1)$. Then $\left\{\left(\frac{x}{2}, 1\right)\right\}_{x\in(0,1)}$ is an open cover of A, because, for each $x_0 \in (0, 1)$, we have*

$$x_0 \in \left(\frac{x_0}{2}, 1\right) \subseteq \bigcup_{x\in(0,1)} \left(\frac{x}{2}, 1\right).$$

These last two examples give open covers that cover the set A in a trivial way, because for each point x of the set A there is an element of the open cover which both corresponds to x and contains x. Such covers may seem extravagant, in the sense that there are so many sets in the cover - an uncountable number in both examples. We raise the question of whether we can extract a smaller number of sets from the cover, that still do the job of covering A.

Definition 17.0.8 *Let A be a subset of $\mathbb{R}$, let Λ be a set and let $\{O_\lambda\}_{\lambda\in\Lambda}$ be an open cover of A. A subcover of A is a set of the form $\{O_{\lambda'}\}_{\lambda\in\Lambda'}$, where $\Lambda' \subseteq \Lambda$, such that $\{O_{\lambda'}\}_{\lambda\in\Lambda'}$ is an open cover of A. A finite subcover of A is a subcover of A with only finitely many elements (that is, Λ' is a finite set).*

We are particularly interested in whether a cover has a finite subcover. In some important cases, having a finite subcover allows us to take a minimum of some positive quantity associated with the cover (for example, the radius, if the elements of the cover are all intervals) and still have a positive quantity (whereas the inifimum of an infinite set of positive numbers may be 0), which then may allow us to obtain a uniform estimate. We will see examples of this phenomenon later.

Although Examples 17.0.6 and 17.0.7 appear to be constructed similarly (put an open set around each point of A), they are different from the standpoint of finding a finite subcover.

Example 17.0.9 *Let $A = [0,1]$ and let $\left\{\left(x - \frac{1}{10}, x + \frac{1}{10}\right)\right\}_{x\in[0,1]}$ be the open cover of A from Example 17.0.6. This cover has a finite subcover. Specifically, we can let*

$$I_i = \left(x_i - \frac{1}{10}, x_i + \frac{1}{10}\right), \text{ for } x_i = \frac{i}{10}, \ i = 0, 1, 2, \ldots, 10.$$

Then $[0,1] \subseteq \cup_{i=0}^{10} I_i$, because every real number in $[0,1]$ has distance less than $\frac{1}{10}$ from one of the numbers $0, \frac{1}{10}, \frac{2}{10}, \ldots, \frac{9}{10}, 1$.

Example 17.0.10 *Let $A = (0,1)$ and let $\left\{\left(\frac{x}{2}, 1\right)\right\}_{x\in(0,1)}$ be the open cover of A from Example 17.0.7. This cover has no finite subcover. To see why, consider any finite set $x_1, x_2, \ldots, x_n \in (0,1)$. Let $y = \min\{x_1, x_2, \ldots, x_n\}$. Then, for example, $\frac{y}{4} \in (0,1)$ but $\frac{y}{4} \notin \left(\frac{x_i}{2}, 1\right)$ for all $i = 1, 2, \ldots, n$, because $\frac{y}{4} < \frac{y}{2} \leq \frac{x_1}{2}$ for all i because $y \leq x_i$. Hence $\frac{y}{4} \notin \cup_{i=1}^{n} \left(\frac{x_i}{2}, 1\right)$, so $\left\{\left(\frac{x_i}{2}, 1\right)\right\}_{i=1}^{n}$ is not a subcover of $(0,1)$. So $\left\{\left(\frac{x}{2}, 1\right)\right\}_{x\in(0,1)}$ has no finite subcover.*

It turns out that certain sets A have the property that any open cover of A has a finite subcover.

Definition 17.0.11 *A subset K of $\mathbb{R}$ is compact, or open cover compact, if every open cover of K has a finite subcover.*

Example 17.0.12 *Any finite set of points of $\mathbb{R}$ is compact.*

Proof. Let $A = \{x_1, x_2, \ldots, x_n\}$ be a set of points in $\mathbb{R}$, for some $n \in \mathbb{N}$ and some set Λ. Let $\{O_\lambda\}_{\lambda\in\Lambda}$ be an open cover of A. For each $i = 1, 2, \ldots, n$, we have $x_i \in A \subseteq \cup_{\lambda\in\Lambda} O_\lambda$, so there exists $\lambda_i \in \Lambda$ such that $x_i \in O_{\lambda_i}$. Then $A \subseteq \cup_{i=1}^{n} O_{\lambda_i}$, so A has a finite subcover. ■

A simple fact about compact sets is that they must be bounded.

Proposition 17.0.13 *Let K be a compact subset of $\mathbb{R}$. Then K is bounded.*

Proof. For each $n \in \mathbb{N}$, the interval $(-n, n)$ is open, and $K \subseteq \cup_{n=1}^{\infty}(-n, n)$, because each point $x \in K$ satisfies $x \in (-n, n)$ for every $n > |x|$. (In fact, $\cup_{n=1}^{\infty}(-n, n) = \mathbb{R}$.) Thus $\{(-n, n)\}_{n=1}^{\infty}$ forms an open cover of K. Since K is compact, there is a finite set $\{n_1, n_2, \ldots, n_k\}$ of natural numbers, for some $k \in \mathbb{N}$, such that $K \subseteq \cup_{i=1}^{k}(-n_i, n_i)$. Let $N = \max\{n_1, n_2, \ldots, n_k\}$. Then $(-n_1, n_i) \subseteq (-N, N)$ for each $i = 1, 2, \ldots, k$, so $K \subseteq (-N, N)$. That is, if $x \in K$, then $x \in (-N, N)$, so $|x| < N$. Thus K is bounded. ■

The proof of Proposition 17.0.13 shows how compactness is used in many proofs to obtain a uniform bound from pointwise bounds. For each point x in K, of course we have $|x| < N_x$ for some N_x, which allows us to construct an infinite open cover of K. By compactness, we pass to a finite subcover. Then using the fact that the maximum of finitely many numbers is finite (whereas the infinitely many numbers may be unbounded), we obtain a single N, independent of x, such that $|x| < N$ for all $x \in K$.

The assumption that the sets in the cover are open did not play a role in Example 17.0.12, but the role of that assumption is clarified in a comparison of the next two examples.

Example 17.0.14 *Let $A = \{\frac{1}{i} : i \in \mathbb{N}\}$. Then A is not compact.*

Proof. To show that A is not compact, we exhibit an open cover of A that has no finite subcover. The idea is to cover the points $\frac{1}{i}$ of A with sufficiently small open intervals that no two intervals in the cover overlap. Then removing even one set from the cover would result in a point of A not being in

the remaining union. So there cannot be any finite subcover. To be precise, let I_i be the open interval $(\frac{1}{i} - \frac{1}{2i(i+1)}, \frac{1}{i} + \frac{1}{2i(i+1)})$. (This radius $\frac{1}{2i(i+1)}$ is chosen to be half of the distance from $\frac{1}{i}$ to $\frac{1}{i+1}$, since $\frac{1}{i+1}$ is the nearest point of the form $\frac{1}{j}$ to $\frac{1}{i}$.) Then for $i, j \in \mathbb{N}$ with $i \neq j$, we claim that

$$I_i \cap I_j = \left(\frac{1}{i} - \frac{1}{2i(i+1)}, \frac{1}{i} + \frac{1}{2i(i+1)}\right) \cap \left(\frac{1}{j} - \frac{1}{2j(j+1)}, \frac{1}{j} + \frac{1}{2j(j+1)}\right) = \emptyset.$$

To prove this fact, we may assume $i < j$. Since $i, j \in \mathbb{N}$ and $j \geq i+1$, we have $1 \leq i^2 \leq i(j-1) = ij - i$. Hence $0 \leq ij - i - 1$, which implies that $j < (i+1)(j-i)$, or $\frac{1}{i+1} \leq \frac{j-i}{j}$, and finally $\frac{1}{i(i+1)} \leq \frac{j-i}{ij}$. By way of contradiction, suppose $x \in I_i \cap I_j$. Since $x \in I_i$, we have $\left|x - \frac{1}{i}\right| \leq \frac{1}{2i(i+1)}$, and similarly $\left|x - \frac{1}{j}\right| \leq \frac{1}{2j(j+1)} \leq \frac{1}{2i(i+1)}$ (the last inequality holds because we assumed $i < j$). Then by the triangle inequality,

$$\frac{j-i}{ij} = \left|\frac{1}{i} - \frac{1}{j}\right| \leq \left|\frac{1}{i} - x\right| + \left|x - \frac{1}{j}\right| < \frac{1}{2i(i+1)} + \frac{1}{2j(j+1)} \leq 2 \cdot \frac{1}{2i(i+1)} = \frac{1}{i(i+1)} \leq \frac{j-i}{ij},$$

using the result from above at the end. Thus we have $\frac{j-i}{ij} < \frac{j-i}{ij}$, a contradiction. Hence no $x \in I_i \cap I_j$ exists, so $I_i \cap I_j = \emptyset$.

Now suppose we have any subcover of $\{I_i\}_{i=1}^{\infty}$. For each $i \in \mathbb{N}$, there must be an element of the subcover containing $\frac{1}{i}$. But by the disjointness of the intervals $\{I_j\}_{j=1}^{\infty}$, there is only one interval in the original cover containing $\frac{1}{i}$, nameley I_i. So I_i must belong to the subcover, for each $i \in \mathbb{N}$. That is, there is no subcover other than the full original cover, which has infinitely many distinct elements. So the cover $\{I_i\}_{i=1}^{\infty}$ of A has no finite subcover. Hence A is not compact. ■

One might think that compactness is a measure of the size of the set, so that a subset of a compact set would have to be compact. That supposition is not true, however, as the next example shows: we can add one point to the set A of Example 17.0.14 and obtain a compact set.

Example 17.0.15 *Let A be the set of Example 17.0.14, and let $K = A \cup \{0\}$. Then K is compact.*

Proof. Let $\{O_\lambda\}_{\lambda \in \Lambda}$ be an open cover of K, where Λ is some set. Since $0 \in A$, there exists $\lambda_0 \in \Lambda$ such that $0 \in O_{\lambda_0}$. Since O_{λ_0} is open and contains 0, there exists $h > 0$ such that $(-h, h) \subseteq O_{\lambda_0}$. Let $N \in \mathbb{N}$ be such that $N > \frac{1}{h}$. For $n > N$, we have $n > \frac{1}{h}$, so $0 < \frac{1}{n} < h$. Hence $\frac{1}{n} \in (-h, h) \subseteq O_{\lambda_0}$, for all $n > N$. For each $n = 1, 2, \ldots, N$, there exists $\lambda_n \in \Lambda$ such that $\frac{1}{n} \in O_{\lambda_n}$, since $\{O_\lambda\}_{\lambda \in \Lambda}$ covers K. Then the finite collection $\{O_{\lambda_0}, O_{\lambda_1}, \ldots O_{\lambda_n}\}$ forms a subcover of A, since O_{λ_0} contains $\frac{1}{n}$ for all $n > N$, and O_{λ_n} contains $\frac{1}{n}$, for each $n = 1, 2, \ldots, N$. ■

Notice how the requirement that the sets in the cover be open played a role in Example 17.0.15. There had to be an element of the open cover containing the point 0, since $0 \in K$, but since an open set containing 0 has to contain an interval around 0, that element of the cover contained all but finitely many of the points in K. Thus compactness measures the structure of a set, not exactly its size. More precisely, we have the following key property of compact sets.

Proposition 17.0.16 *Let K be a compact subset of $\mathbb{R}$. Then K is closed.*

Proof. We will prove that $K^c = \mathbb{R} \setminus K$ is open, which, by Proposition 16.0.13, implies that K is closed. Let $x \in K^c$. For $n \in \mathbb{N}$, let

$$O_n = \left(-\infty, x - \frac{1}{n}\right) \cup \left(x + \frac{1}{n}, \infty\right) = \left\{y \in \mathbb{R} : |y - x| > \frac{1}{n}\right\}.$$

Notice that

$$O_n^c = \left\{y \in \mathbb{R} : |y - x| \leq \frac{1}{n}\right\} = \left[x - \frac{1}{n}, x + \frac{1}{n}\right].$$

Also, each O_n is open (e.g., since O_n^c is closed), $O_n \subseteq O_{n+1}$ for each $n \in \mathbb{N}$ (i.e., the sets O_n are nested), and $\cup_{n=1}^{\infty} O_n = \mathbb{R} \setminus \{x\}$. Hence $K \subseteq \cup_{n=1}^{\infty} O_n$, since $x \notin K$. That is, $\{O_n\}_{n \in \mathbb{N}}$ is an open cover of K. Since

K is compact, there exists a finite subcover. That is, there exist $k \in \mathbb{N}$ and $n_1, n_2, \ldots, n_k \in \mathbb{N}$ such that $K \subseteq \cup_{i=1}^{k} O_{n_i}$. Let $N = \max(n_1, n_2, \ldots, n_k)$. Then $O_{n_i} \subseteq O_N$ for all $i = 1, 2, \ldots, k$, since the sets O_n are nested. Hence $\cup_{i=1}^{k} O_{n_i} = O_N$, and so $K \subseteq \cup_{i=1}^{k} O_{n_i} = O_N$. Taking complementary sets, we have $O_N^c \subseteq K^c$, hence

$$\left(x - \frac{1}{N}, x + \frac{1}{N}\right) \subseteq \left[x - \frac{1}{N}, x + \frac{1}{N}\right] = O_N^c \subseteq K^c.$$

Hence the open interval around x of radius $1/N$ is contained in K^c. Since $x \in K^c$ is arbitrary, we obtain that K^c is open. Hence K is closed. ■

Notice how the last proof used compactness to reduce to finitely many $1/n_i$, thus obtaining a positive radius $1/N$ as the minimum of finitley many positive numbers.

We have already seen that subsets of compact sets are not necessarily compact (by Examples 17.0.14 and 17.0.15). Proposition 17.0.16 identifies one obstruction: the subset would have to be closed to have a chance to be compact. It turns out that this obstruction is the only one.

Proposition 17.0.17 *Let K be a compact subset of $\mathbb{R}$ and suppose $E \subseteq K$ and E is closed. Then E is compact.*

Proof. Let $\{O_\lambda\}_{\lambda \in \Lambda}$ be an open cover of E, for some index set Λ. Since E is closed, E^c is open. Then $E^c \cup \cup_{\lambda \in \Lambda} O_\lambda = \mathbb{R}$ since, for $x \in \mathbb{R}$, either $x \in E \subseteq \cup_{\lambda \in \Lambda} O_\lambda$ or $x \in E^c$. In particular, then, $K \subseteq E^c \cup \cup_{\lambda \in \Lambda} O_\lambda$, so $\{O_\lambda\}_{\lambda \in \Lambda} \cup \{E^c\}$ is an open cover of K. Since K is compact, there is a finite subcover of K. There are two possibilities: either E^c is an element of this subcover or not.

If E^c is not an element of this subcover, then there exists $n \in \mathbb{N}$ and $\lambda_1, \lambda_2, \ldots, \lambda_n \in \Lambda$ such that $K \subseteq \cup_{i=1}^{n} O_{\lambda_i}$. Since $E \subseteq K$, we have $E \subseteq \cup_{i=1}^{n} O_{\lambda_i}$. Hence $\{O_{\lambda_1}, O_{\lambda_2}, \ldots, O_{\lambda_n}\}$ is a finite subset of $\{O_\lambda\}_{\lambda \in \Lambda}$ which covers E, as required.

If E^c is an element of the finite subcover of K, then then there exists $n \in \mathbb{N}$ and $\lambda_1, \lambda_2, \ldots, \lambda_n \in \Lambda$ such that $K \subseteq E^c \cup \cup_{i=1}^{k} O_{\lambda_i}$. Since $E \subseteq K$, we have $E \subseteq E^c \cup \cup_{i=1}^{k} O_{\lambda_i}$. We claim that in fact $E \subseteq \cup_{i=1}^{k} O_{\lambda_i}$. To verify this fact, if $x \in E$, then because $E \subseteq E^c \cup \cup_{i=1}^{k} O_{\lambda_i}$ we know that either $x \in E^c$ or $x \in \cup_{i=1}^{k} O_{\lambda_i}$. But $x \notin E^c$ because $x \in E$, so it must be that $x \in \cup_{i=1}^{k} O_{\lambda_i}$. Therefore again $\{O_{\lambda_1}, O_{\lambda_2}, \ldots, O_{\lambda_n}\}$ is a finite subset of $\{O_\lambda\}_{\lambda \in \Lambda}$ which covers E.

Hence E is compact. ■

The key to our general result below (the Heine-Borel Theorem, Theorem 17.0.19) characterizing compact sets is the next Lemma.

Lemma 17.0.18 *Let $a, b \in \mathbb{R}$ with $a < b$. Then the closed interval $[a, b]$ is compact.*

Proof. Let $\{O_\lambda\}_{\lambda \in \Lambda}$ be an open cover of $[a, b]$. By way of contradiction, we suppose that $[a, b]$ has no finite subcover from $\{O_\lambda\}_{\lambda \in \Lambda}$. We introduce the following notation: for any interval $J = [\alpha, \beta]$, let $\ell(J) = \beta - \alpha$ denote the length of the interval J. Let $I_0 = [a, b]$. We claim that there exist closed intervals $\{I_j\}_{j=1}^{\infty}$ such that, for all $j \in \mathbb{N}$

(i) $I_j \subseteq I_{j-1}$; that is, the intervals $\{I_j\}_{j=1}^{\infty}$ are decreasing and hence all are contained in $[a, b] = I_0$ because $I_1 \subseteq I_0$;

(ii) $\ell(I_j) = \frac{1}{2}\ell(I_{j-1}) = \frac{b-a}{2^j}$;
and
(iii) I_j has no finite subcover from $\{O_\lambda\}_{\lambda \in \Lambda}$.

We prove the claim by induction. To prove the case $j = 1$, let $I_0^\ell = [a, a + \frac{b-a}{2}]$ be the closed left half of $I_0 = [a, b]$ and let $I_0^r = [a + \frac{b-a}{2}, b]$ be the closed right half of I_0. Since I_0 is asumed to have no finite subcover from $\{O_\lambda\}_{\lambda \in \Lambda}$, then either I_0^ℓ or I_0^r (or both) has no finite subcover from $\{O_\lambda\}_{\lambda \in \Lambda}$ (since, if each has a finite subcover, then the union of these two subcovers would be a finite subcover of I_0, which is impossible). Denote by I_1 either I_0^ℓ or I_0^r, whichever has no finite subcover from $\{O_\lambda\}_{\lambda \in \Lambda}$ (if neither has a finite subcover, pick one). Then I_1 is a closed interval, (i) $I_1 \subseteq I_0$, (ii) $\ell(I_1) = \frac{1}{2}\ell(I_0) = \frac{b-a}{2}$, and I_1 has no finite subcover from $\{O_\lambda\}_{\lambda \in \Lambda}$.

We now prove the inductive step, by iterating this bisection process. Suppose $I_1, I_2, \ldots, I_j$ satisfying (i), (ii), and (iii) have been found. To find I_{j+1}, let I_j^ℓ be the closed left half of I_j and let I_j^r be the closed right half of I_j. Since I_j has no finite subcover from $\{O_\lambda\}_{\lambda\in\Lambda}$, then at least one of I_j^ℓ and I_j^r, call it I_{j+1}, has no finite subcover from $\{O_\lambda\}_{\lambda\in\Lambda}$. Then I_{j+1} is a closed interval, (i) $I_{j+1} \subseteq I_j$, (ii) $\ell(I_{j+1}) = \frac{1}{2}\ell(I_j) = \frac{1}{2}\cdot\frac{b-a}{2^j} = \frac{b-a}{2^{j+1}}$, and I_{j+1} has no finite subcover from $\{O_\lambda\}_{\lambda\in\Lambda}$.

This argument completes the induction step, and hence we obtain closed intervals $\{I_j\}_{j=1}^\infty$ satisfying (i), (ii), and (iii), for all $j \in \mathbb{N}$, by induction.

By (i), the closed intervals $\{I_j\}_{j=0}^\infty$ are nested, so by the Nested Interval Property (Theorem 11.0.8), $\cap_{i=0}^\infty I_i \neq \emptyset$. Let $x \in \cap_{i=0}^\infty I_i \neq \emptyset$. Since $x \in [a,b]$, there exists $\lambda' \in \Lambda$ such that $x \in O_{\lambda'}$ since $\{O_\lambda\}_{\lambda\in\Lambda}$ covers $[a,b]$. Since $O_{\lambda'}$ is open and contains x, there exists $r > 0$ such that $(x-r, x+r) \subseteq O_{\lambda'}$. Choose j sufficiently large that $\ell(I_j) = \frac{b-a}{2^j} < r$. Since $x \in I_j$ and $\ell(I_j) < r$, we see that $I_j \subseteq (x-r, x+r) \subseteq O_{\lambda'}$ (proof: if $y \in I_j$, then $|x-y| \leq \ell(I_j) < r$, so $y \in (x-r, x+r)$). Hence I_j has a finite subcover from $\{O_\lambda\}_{\lambda\in\Lambda}$, namely the single set $O_{\lambda'}$. This contradicts (iii), and hence shows that the assumption that $[a,b]$ has no finite subcover from $\{O_\lambda\}_{\lambda\in\Lambda}$ is false. So a finite subcover of $[a,b]$ from $\{O_\lambda\}_{\lambda\in\Lambda}$ exists. Since $\{O_\lambda\}_{\lambda\in\Lambda}$ is an arbitrary open cover of $[a,b]$, we conclude that $[a,b]$ is compact. ■

Lemma 17.0.18 is the most difficult step in the proof of the next result, which is one of the deepest results in this course. Note that the key ingredient of Lemma 17.0.18 was the Nested Interval Property, whose proof depends essentially on the completeness property of $\mathbb{R}$.

Theorem 17.0.19 *(Heine-Borel Theorem) A subset E of $\mathbb{R}$ is compact if and only if E is closed and bounded.*

Proof. First, suppose E is compact. By Proposition 17.0.13, E is bounded, and by Proposition 17.0.16, E is closed.

Now suppose E is closed and bounded. Since E is bounded, there exists $N \in \mathbb{N}$ such that $E \subseteq [-N, N]$. By Lemma 17.0.18, $[-N, N]$ is compact, and E is closed, by assumption. By Proposition 17.0.17, E is compact. ■

This characterization of compact sets holds in $\mathbb{R}$ and in $\mathbb{R}^n$. It does not hold in a general metric space: although compact sets are closed and bounded, it is not necessarily true that closed, bounded sets are compact in general metric spaces. In particular, in metric spaces which are also vector spaces, it generally does not hold if the dimension of the vector space is infinite. The fact that closed, bounded sets in finite dimensions are compact greatly simplifies analysis in finite dimensions compared to analysis in infinite-dimensional spaces, like those of Example 1.0.2 and 1.0.3.

Corollary 17.0.20 *Let E be a subset of $\mathbb{R}$. Then E is sequentially compact if and only if E is open cover compact.*

Proof. By Theorems 17.0.4 and 17.0.19, both the sequential compactness and the open cover compactness of E are equivalent to the property that E is closed and bounded. ■

It turns out that the two notions of compactness, sequential compactness and open cover compactness, are equivalent more generally in metric spaces, even though neither is equivalent to being closed and bounded. The proof of this equivalence is much more difficult, because one does not have a simpler characterization of either type of compactness. In the more general setting of topological spaces, there are examples where sequential compactness and open cover compactness are not equivalent.

The nested interval property was used to prove that a finite closed interval is compact. The finite closed intervals in the nested interval property are examples of compact sets. The nested interval property can be generalized to compact sets in the following way.

Proposition 17.0.21 *Suppose that K_j is a compact subset of $\mathbb{R}$ for each $j \in \mathbb{N}$. If $\cap_{j=1}^n K_j \neq \emptyset$ for all $n \in N$, then $\cap_{j=1}^\infty K_j \neq \emptyset$.*

Notice that the nested interval property (Theorem 11.0.8) is a special case of Proposition 17.0.21. We leave the proof of Proposition 17.0.21 as an exercise. The contrapositive form of the statement in Proposition 17.0.21 is sometimes useful: If each K_j is a compact subset of $\mathbb{R}$ and $\cap_{j=1}^\infty K_j = \emptyset$, then there exists $n \in \mathbb{N}$ such that $\cap_{j=1}^n K_j = \emptyset$.

Compact sets will be particularly useful when studying continuous functions, which is our next topic.

Chapter 18

Limits of Functions on $\mathbb{R}$

Definition of the Limit of a Function Defined on $\mathbb{R}$

In Chapters 13, 14, and 15, we studied sequences and their limits. Recall that a sequence is a function defined on the natural numbers. Now we want to consider functions defined on more general subsets of the real numbers, often on uncountable sets like intervals. We want to see if $f(x)$ approaches some value L as x approaches some value $c \in \mathbb{R}$. If so, we will write $\lim_{x\to c} f(x) = L$. Let's try to formulate what that should mean.

The rough idea of "$\lim_{x\to c} f(x) = L$" is that $f(x)$ gets close to L as x gets close to c. From our experience with sequences in Chapter 13, we know how to make the statement "$f(x)$ gets close to L" precise. We introduce the notation ϵ to denote a positive number which we think of as small, and say that "eventually" we have $|f(x) - L| < \epsilon$. By eventually we no longer mean that $n \to \infty$, as we did for sequences; instead we imagine points x getting closer and closer to the limit value c. We should have the conclusion $|f(x) - L| < \epsilon$ for all x which are close enough to c. To be more precise, once $\epsilon > 0$ is given, there should be some positive number, which we call δ, so that if $|x - c| < \delta$, then $|f(x) - L| < \epsilon$. This definition is almost correct, except for one detail. The limit as $x \to c$ of $f(x)$ should be something that depends on the values of $f(x)$ for x near *but not equal to* c. In fact, we often want to know if $\lim_{x\to c} f(x)$ exists when f is not defined at c (so that possibly we can extend f to be defined at c), or if f is defined at c we want to know if the value of f at c agrees with $\lim_{x\to c} f(x)$ (if that limit exists). So we exclude $x = c$ from the points we consider when defining $\lim_{x\to c} f(x)$. This exclusion is easily expressed by saying that the conclusion $|f(x) - L| < \epsilon$ should hold for x satisfying $0 < |x - c| < \delta$; since $|x - c| \neq 0$ we know that $x = c$ is not considered. Hence we have the notorious '$\epsilon - \delta$" definition of limits, as follows.

Definition 18.0.1 *Suppose $c \in \mathbb{R}, L \in \mathbb{R}$, and there exists $r > 0$ such that f is a real-valued function defined on $(c - r, c) \cup (c, c + r)$. Then $\lim_{x\to c} f(x) = L$ (or $f(x) \xrightarrow{x\to c} L$) if, for all $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x) - L| < \epsilon$ for all $x \in (c - r, c) \cup (c, c + r)$ such that $0 < |x - c| < \delta$. If there is no $L \in \mathbb{R}$ such that $\lim_{x\to c} f(x) = L$, we say that $\lim_{x\to c} f(x)$ does not exist.*

It is important to understand that the $\delta > 0$ in the definition depends on ϵ; this point is implied by the fact that in the definition, the existence of δ is stated after ϵ has been introduced. Notice that in this definition we are assuming f is defined for all x satisfying $0 < |x - c| < r$. Later we will be more careful about the definition and allow f to be defined on more general sets. First let's do some examples of proving certain limit values using the definition.

Example 18.0.2 *Prove that* $\lim_{x\to 1}(4x + 2) = 6$.

As we saw with limit examples for sequences, there is often some side work done before writing the proof. Let's do that now. For $\epsilon > 0$ given, we want to obtain $|f(x) - L| < \epsilon$ for x close enough to 1. So we compute

$$|f(x) - L| = |4x + 2 - 6| = |4x - 4| = 4|x - 1|.$$

To make $4|x - 1| < \epsilon$, we just need to make $|x - 1| < \frac{\epsilon}{4}$, so we take $\delta = \frac{\epsilon}{4}$. Here is the formal proof.

PROOF. Let $\epsilon > 0$. Let $\delta = \frac{\epsilon}{4}$. Suppose $0 < |x-1| < \delta$. Then

$$|4x + 2 - 6| = |4x - 4| = 4|x - 1| < 4\delta = \epsilon.$$

■

Example 18.0.3 *Prove that* $\lim_{x\to 3} x^2 = 9$.

Again let's first do the side work. We need to show that $|x^2 - 9| < \epsilon$ when $0 < |x - 3| < \delta$. We factor $x^2 - 9$ to write

$$|x^2 - 9| = |(x + 3)(x - 3)| = |x + 3||x - 3| < |x + 3|\delta,$$

if $|x - 3| < \delta$. We want to have $|x + 3|\delta < \epsilon$, but we can't just let $\delta = \frac{\epsilon}{x+3}$ because δ can't depend on x. (Note the definition says that given ϵ, there exists δ that works for all x satisfying $|x - c| < \delta$, which means that the same δ must work for all these x, and so δ can't depend on x.) However, this concern seems a little silly, since we really only care about x values close to 3, and for these the term $|x + 3|$ can't be too big. In particular, if, say, $|x - 3| < 1$, then $2 < x < 4$, so $5 < x + 3 < 7$, hence $|x + 3| < 7$. Thus if $|x - 3| < 1$ and $|x - 3| < \frac{\epsilon}{7}$, we have

$$|x^2 - 9| = |x + 3||x - 3| < 7 \cdot \frac{\epsilon}{7} = \epsilon.$$

The number "1" in the statement "$|x-3| < 1$" above, was chosen just because it is the easiest to work with. One could use any positive number. How do we make both conditions ($|x - 3| < 1$ and $|x - 3| < \frac{\epsilon}{7}$) hold for $|x - 3| < \delta$? We just define $\delta = \min\left(1, \frac{\epsilon}{7}\right)$. We still have $\delta > 0$ because the minimum of two strictly positive numbers is still strictly positive. The formal proof is as follows.

PROOF. Let $\epsilon > 0$. Let $\delta = \min\left(1, \frac{\epsilon}{7}\right)$. Suppose $0 < |x - 3| < \delta$. Then $|x - 3| < 1$, so $2 < x < 4$, hence $5 < x + 3 < 7$, and therefore $|x + 3| < 7$. Therefore

$$|x^2 - 9| = |x + 3||x - 3| < 7\delta \leq \epsilon,$$

since $\delta \leq \frac{\epsilon}{7}$. ■

One of the most important things to keep in mind about limits is that they don't necessarily exist. To see this in examples, we first formulate the negation of the statement $\lim_{x\to c} f(x) = L$. If it is not true that $\lim_{x\to c} f(x) = L$, then there exists $\epsilon > 0$ such that for all $\delta > 0$, there exists $x \in (c - r, c) \cup (c, c + r)$ such that $0 < |x - c| < \delta$ and $|f(x) - L| \geq \epsilon$.

Example 18.0.4 *(Dirichlet Function) Define* $f : \mathbb{R} \to \mathbb{R}$ *by* $f(x) = 1$ *if* $x \in \mathbb{Q}$ *and* $f(x) = 0$ *if* $x \in \mathbb{R} \setminus \mathbb{Q}$. *Then for every* $c \in \mathbb{R}$, $\lim_{x\to c} f(x)$ *does not exist.*

PROOF. Suppose, by way of contradiction, that there exists some $L \in \mathbb{R}$ such that $\lim_{x\to c} f(x) = L$. We apply the definition of $\lim_{x\to c} f(x) = L$ with $\epsilon = \frac{1}{2}$. Then there exists some $\delta > 0$ such that $|f(x) - L| < \frac{1}{2}$ for all $x \in \mathbb{R}$ satisfying $0 < |x - c| < \delta$. Then for any points x_1 and x_2 satisfying $0 < |x_1 - c| < \delta$ and $0 < |x_2 - c| < \delta$, we have

$$|f(x_1) - f(x_2)| = |f(x_1) - L + L - f(x_2)| \leq |f(x_1) - L| + |L - f(x_2)| < \frac{1}{2} + \frac{1}{2} = 1.$$

But, no matter how small δ is (as long as $\delta > 0$), there exist both a rational number x_1 satisfying $0 < |x_1 - c| < \delta$ (by Theorem 11.0.4) and an irrational number x_2 satisfying $0 < |x_2 - c| < \delta$ (by Theorem 11.0.7). Then $|f(x_1) - f(x_2)| = |0 - 1| = 1$, contradicting the inequality $|f(x_1) - f(x_2)| < 1$ above. Hence there is no L such that $\lim_{x\to c} f(x) = L$. ■

The function in Example 18.0.4 is called the *Dirichlet Function*. It is a very strange function, which can't really be graphed accurately. However, there is also the following related example. For this example we assume the standard properties of the sine function.

Example 18.0.5 *Define* $f : (-\infty, 0) \cup (0, \infty) \to \mathbb{R}$ *by* $f(x) = \sin\left(\frac{1}{x}\right)$. *Then* $\lim_{x\to 0} f(x)$ *does not exist.*

PROOF. Notice that $f\left(\frac{1}{2n\pi}\right) = \sin(2n\pi) = 0$, whereas $f\left(\frac{1}{2n\pi+\frac{\pi}{2}}\right) = \sin(2n\pi + \frac{\pi}{2}) = 1$. For any $r > 0$, if we choose $N > \frac{1}{2\pi r}$, then $\frac{1}{2\pi N} < r$. Then for all $n > N$ we have $0 < \frac{1}{2\pi n} < r$ and $0 < \frac{1}{2\pi n+\frac{\pi}{2}} < r$. Thus f takes the value 0 infinitely many times in $(0, r)$ (at all points $\frac{1}{2\pi n}$ for $n > N$) and f takes the value 1 infinitely many times in $(0, r)$ (at all points $\frac{1}{2\pi n+\frac{\pi}{2}}$ for $n > N$). Hence for the same reason as in Example 18.0.4, $\lim_{x\to 0} f(x)$ does not exist. ■

The function $\sin\left(\frac{1}{x}\right)$ oscillates between $+1$ and -1 infinitely many times in any interval around 0, oscillating more and more quickly as x approaches 0. This example is a good one to keep in mind; it will be used many times in this course.

Another situation where a limit does not exist is if a function blows up to $+\infty$ or $-\infty$. For example, $\lim_{x\to 0}\left(\frac{1}{x}\right)$ does not exist. Many texts give a separate definition for $\lim_{x\to c} f(x) = +\infty$ or $\lim_{x\to c} f(x) = -\infty$, but we don't do so now to avoid confusion with finite limits.

To get the idea of limits on $\mathbb{R}$, we assumed in Definition 18.0.1 that f is defined on the intervals $(c-r, c)$ and $(c, c+r)$, for some $r > 0$. However, limits can be defined under weaker conditions. Let's ask the question: suppose $A \subseteq \mathbb{R}$ and $f : A \to \mathbb{R}$ is defined. When can we define $\lim_{x\to c} f(x)$? The first observation regarding this question is that we don't need $c \in A$. For example, it turns out that $\lim_{x\to 0} \frac{\sin x}{x} = 1$, althougth of course $\frac{\sin x}{x}$ is not defined at 0. In fact, one of the main purposes of finding limits is to extend the definition of f in a reasonable way to points where f is not originally defined.

On the other hand, we need f to be defined near the point c. For example, if f is defined only on $(0, 1)$, it doesn't make much sense to ask whether $\lim_{x\to 2} f(x)$ exists. In this example, it wouldn't help if f is defined on $(0, 1) \cup \{2\}$ because $\lim_{x\to 2}$ should not depend on the value of f at 2, or even whether f is defined at 2.

Nevertheless, we don't need f to be defined on $(c - r, c) \cup (c, c + r)$ to define $\lim_{x\to c} f(x)$. For example, if f is defined only on the set $\left\{\frac{1}{n} : n \in \mathbb{N}\right\}$, it may happen that $f\left(\frac{1}{n}\right)$ tends to a limit as $\frac{1}{n} \to 0$, so it could be reasonable to define $\lim_{x\to 0} f(x)$ in this case. In general, what is needed to define whether $\lim_{x\to c} f(x)$ exists or not is for f to be defined at least on a sequence of points (x_n), with $x_n \neq c$ for all $n \in \mathbb{N}$, such that $x_n \to c$. This idea leads to the next definition.

Definition 18.0.6 *Let $A \subseteq \mathbb{R}$. We say that x is a limit point of A if there exists a sequence (x_n) with $x_n \in A \setminus \{x\}$ for all $n \in \mathbb{N}$, such that $\lim_{n\to\infty} x_n = x$. Let*

$$A' = \{x \in \mathbb{R} : x \text{ is a limit point of } A\}.$$

With this understanding, we can make the following more general definition of a limit.

Definition 18.0.7 *Suppose $A \subseteq \mathbb{R}, f : A \to \mathbb{R}, c \in A'$, and $L \in \mathbb{R}$. We say $\lim_{x\to c} f(x) = L$, or $f(x) \xrightarrow{x\to c} L$, if, for all $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x) - L| < \epsilon$ for all $x \in A$ such that $0 < |x - c| < \delta$.*

Properties of Limits

Limits of functions on $\mathbb{R}$ can be defined in terms of sequences. If $c \in A'$, then there is a sequence (x_n) with $x_n \in A \setminus \{c\}$ for all $n \in \mathbb{N}$, such that $x_n \to c$. Then $(f(x_n))$ forms another sequence, and we consider whether this sequence converges.

Proposition 18.0.8 *Suppose $A \subseteq \mathbb{R}, f : A \to \mathbb{R}, c \in A'$, and $L \in \mathbb{R}$. Then $\lim_{x\to c} f(x) = L$ if and only if $\lim_{n\to\infty} f(x_n) = L$ for all sequences (x_n) satisfying $x_n \in A \setminus \{c\}$ for all $n \in \mathbb{N}$ and $x_n \to c$.*

PROOF. First suppose $\lim_{x\to c} f(x) = L$. Let (x_n) be a sequence with $x_n \in A \setminus \{c\}$ for all $n \in \mathbb{N}$, such that $x_n \to c$. We need to show that $\lim_{n\to\infty} f(x_n) = L$. Let $\epsilon > 0$. Since $\lim_{x\to c} f(x) = L$, there exists $\delta > 0$ such that $|f(x) - L| < \epsilon$ for all $x \in A$ such that $0 < |x - c| < \delta$. Since $x_n \to c$, there exists $N \in \mathbb{N}$ such that $|x_n - c| < \delta$ for all $n > N$, and since $x_n \neq c$ for all $n \in \mathbb{N}$, we also have $0 < |x_n - c| < \delta$. Thus for all $n > N$, we have $|f(x_n) - L| < \epsilon$. Hence $\lim_{n\to\infty} f(x_n) = L$.

We now prove the converse statement, namely that if $\lim_{n\to\infty} f(x_n) = L$ for all sequences (x_n) with $x_n \in A \setminus \{c\}$ for all $n \in \mathbb{N}$, such that $x_n \to c$, then $\lim_{x\to c} f(x) = L$. We prove the contrapositive: if it is not true that $\lim_{x\to c} f(x) = L$, then there exists a sequence (x_n) with $x_n \in A \setminus \{c\}$ for all $n \in \mathbb{N}$ such that

$x_n \to c$, but it is not true that $\lim_{x\to c} f(x) = L$. So suppose that it is not true that $\lim_{x\to c} f(x) = L$. This means that there exists some $\epsilon > 0$ such that for all $\delta > 0$, there exists $x \in A$ with $0 < |x - c| < \delta$ but $|f(x) - L| > \epsilon$. For this ϵ, we apply this property with $\delta = \frac{1}{n}$, for each $n \in \mathbb{N}$, to obtain $x_n \in A$ satisfying $0 < |x_n - c| < \frac{1}{n}$ and $|f(x_n) - L| \geq \epsilon$. Consider the sequence (x_n), which satisfies $x_n \in A \setminus \{c\}$ for all n and $\lim_{n\to\infty} x_n = c$ since $|x_n - c| < \frac{1}{n}$. Since $|f(x_n) - L| \geq \epsilon$ for all $n \in \mathbb{N}$, we have that $(f(x_n))$ does not converge to L. Hence the contrapositive is proved, which establishes the converse direction. ■

We next consider some basic facts about limits of functions defined on subsets of $\mathbb{R}$. These facts are analogous to the properties of limits of sequences stated and proved in Chapter 14. They can be proved directly from the definition of limits on $\mathbb{R}$ by arguments analogous to those in Chapter 14, but it is easier and quicker to use the results in Chapter 14 together with Proposition 18.0.8, as follows. The first property is uniqueness of limits.

Proposition 18.0.9 *(Uniqueness of limits) Suppose $A \subseteq \mathbb{R}, f : A \to \mathbb{R}, c \in A'$, and $L_1, L_2 \in \mathbb{R}$. If $\lim_{x\to c} f(x) = L_1$ and $\lim_{x\to c} f(x) = L_2$, then $L_1 = L_2$.*

Proof. Since $c \in A'$, there exists a sequence (x_n) satisfying $x_n \in A \setminus \{c\}$ for all $n \in \mathbb{N}$ such that $x_n \to c$. Since $\lim_{x\to c} f(x) = L_1$, Proposition 18.0.8 guarantees that $\lim_{n\to\infty} f(x_n) = L_1$. However, the same argument applies to L_2 to give $\lim_{n\to\infty} f(x_n) = L_2$. By uniqueness for sequence limits (Proposition 14.0.1), applied to the sequence $(f(x_n))$, we have $L_1 = L_2$. ■

We also have the analogue of Theorem 14.0.4.

Proposition 18.0.10 *Suppose $A \subseteq \mathbb{R}, f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ are functions, $c \in A', \lim_{x\to c} f(x) = L$, and $\lim_{x\to c} g(x) = M$, for some $L, M \in \mathbb{R}$. Then*

(1) for $\alpha \in \mathbb{R}$, $\lim_{x\to c} (\alpha f(x)) = \alpha L$;

(2) $\lim_{x\to c} (f(x) + g(x)) = L + M$;

(3) $\lim_{x\to c} (f(x)g(x)) = LM$;

(4) if $M \neq 0$ and there exists $r > 0$ such that $g(x) \neq 0$ for all $x \in ((c - r, c) \cup (c, c + r)) \cap A$, then $\lim_{x\to c} \dfrac{f(x)}{g(x)} = \dfrac{L}{M}$.

Proof. These results can all be proved by using Proposition 18.0.8 to reduce to Theorem 14.0.4. We illustrate by proving (2) this way. The other parts are proved similarly; we leave those proofs as exercises.

To prove (2), let (x_n) be any sequence satisfying $x_n \in A \setminus \{c\}$ for all $n \in \mathbb{N}$ and $x_n \to c$. Since $\lim_{x\to c} f(x) = L$ and $\lim_{x\to x} g(x) = M$, we have $\lim_{n\to\infty} f(x_n) = L$ and $\lim_{n\to\infty} g(x_n) = M$, by Proposition 18.0.8. Then by Theorem 14.0.4, $\lim_{n\to\infty}(f(x_n) + g(x_n)) = L + M$. Since this conclusion holds for any sequence (x_n) satisfying $x_n \in A\setminus\{c\}$ for all $n \in \mathbb{N}$ and $x_n \to c$, Proposition 18.0.8 implies that $\lim_{x\to c}(f(x)+ g(x)) = L + M$. ■

Chapter 19

Continuous Functions

Definition of Continuity at a Point

One of the fundamental and most powerful ideas in analysis is *continuity*. In this chapter we define what it means for a function defined on a set $A \subseteq \mathbb{R}$ to be continuous at a point of A. To do so, we first have to distinguish the special, but somewhat trivial, case of *isolated points* of A.

Definition 19.0.1 *Let $A \subseteq \mathbb{R}$. A point $x \in A$ is isolated (or, is an isolated point of A) if there exists some $r > 0$ such that $(x - r, x + r) \cap A = \{x\}$.*

That is, x is an isolated point of A if there is an open set containing x which contains no points of A other than x.

Notice that if $x \in A$, and x is not an isolated point of A, then x is a limit point of A (Definition 18.0.6) (proof: If x is not an isolated point of A, then for every $n \in \mathbb{N}$, there exists an $x_n \in A \setminus \{x\}$ such that $|x_n - x| < \frac{1}{n}$, hence $x_n \to x$).

Definition 19.0.2 *Let $A \subseteq \mathbb{R}, c \in A$, and suppose $f : A \to \mathbb{R}$ is a function. We say f is continuous at c if either*

(1) c is an isolated point of A, or

(ii) c is a limit point of A, $\lim_{x\to c} f(x)$ exists, and $\lim_{x\to c} f(x) = f(c)$.

Case (ii) in this definition is the important one to understand: not only does the limit of $f(x)$ as $x \to c$ exist, but f is defined at c and has the "right" value, namely the value of the limit. Case (i) is just needed because $\lim_{x\to c} f(x)$ is not defined if c is an isolated point of A; in that case we declare f to be continuous at c by default.

Example 19.0.3 *The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is continuous at $x = 3$: we showed in Example 18.0.3 that $\lim_{x\to 3} x^2 = 9$, and $f(3) = 3^2 = 9$.*

The next two examples demonstrate the two ways a function can fail to be continuous at a point: the limit may exist but not coincide with the value of the function, or the limit may not exist at all.

Example 19.0.4 *Define $f : \mathbb{R} \to \mathbb{R}$ by $f(3) = -47$ and $f(x) = x^2$ if $x \neq 3$. Then f is not continuous at 3: although $\lim_{x\to 3} f(x)$ exists, it has the value 9, which does not agree with $f(3)$.*

One is tempted to say that the value of $f(3)$ in the last example is the "wrong" value; if we redefine $f(3)$ to be 9, we obtain continuity at 3. In the next example, there is no "right" value.

Example 19.0.5 *Let $\alpha \in \mathbb{R}$. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \sin\left(\frac{1}{x}\right)$ for $x > 0$,and let $f(0) = \alpha$. Then f is not continuous at $x = 0$ (no matter what α is), because $\lim_{x\to 0} f(x)$ does not exist, by Example 18.0.5.*

Example 19.0.6 *Let $\alpha \in \mathbb{R}$. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = 1$ if $x \in \mathbb{Q} \setminus \{0\}, f(x) = 0$ if $x \in \mathbb{R} \setminus \mathbb{Q}$, and let $f(0) = \alpha$. Then f is not continuous at $x = 0$ (no matter what α is), because $\lim_{x\to 0} f(x)$ does not exist, by Example 18.0.4.*

Continuity is considered a nice property for a function to have. Intuitively, a continuous function varies gradually. If f is continuous at c, the function f does not oscillate wildly (as in Example 19.0.5) or jump around erratically near c (as in Example 19.0.6), so that $\lim_{x\to c} f(x)$ exists, and f is defined at the point c in a way that is consistent with the values of f nearby. Under normal conditions, one expects functions that describe physical phenomena (such as the speed of an object at time t, the temperature at a point on the Earth's surface, or the concentration of a chemical in the bloodstream at time t) to be continuous nearly all of the time. Points at which discontinuities occur are exceptional and hence generally meaningful (such as the time when a falling object hits the earth, so that its velocity instantaneously changes from some possibly large value to 0).

There are several important equivalent definitions of continuity. In the next Proposition, condition (2) is the famous $\epsilon - \delta$ definition of continuity, and condition (3) characterizes continuity in terms of sequences.

Proposition 19.0.7 *Suppose $A \subseteq \mathbb{R}, c \in A$, and $f : A \to \mathbb{R}$ is a function. The following are equivalent:*

(1) f is continuous at c;

(2) for all $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(c)| < \epsilon$ for all $x \in A$ such that $|x - c| < \delta$;

(3) $\lim_{n\to\infty} f(x_n) = f(c)$ for all sequences (x_n) such that $x_n \in A$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} x_n = c$.

Proof. First suppose c is an isolated point of A. Then there exists $r > 0$ such that $(c-r, c+r) \cap A = \{c\}$. Then f is continuous at c, by definition, so (1) holds. Also (2) holds, as follows. Given $\epsilon > 0$, let $\delta = r$. If $x \in A$ and $|x - c| < \delta$, then $x = c$, and hence $|f(x) - f(c)| = |f(c) - f(c)| = 0 < \epsilon$. We claim that (3) holds as well. To see why, suppose (x_n) is a sequence of elements of A converging to c. Since $x_n \to c$, there exists $N \in \mathbb{N}$ such that $|x_n - c| < r$ for all $n > N$. Then for $n > N$ we have $x_n = c$, and so $|f(x_n) - f(c)| = |f(c) - f(c)| = 0$ for all $n > N$. Hence $f(x_n) \to f(c)$.

Now suppose c is a limit point of A. Suppose (1) holds. To prove (2), let $\epsilon > 0$. Then by the definition of continuity (Definition 19.0.1 and the definition of limits, Definition 18.0.7), there exists $\delta > 0$ such that $|f(x) - f(x)| < \epsilon$ for all $x \in A$ such that $0 < |x - c| < \delta$. Then (2) holds, because if $|x - c| < \delta$ then either $x = c$, in which case $|f(x) - f(c)| = 0$, or $0 < |x - c| < \delta$, and we know that $|f(x) - f(c)| < \epsilon$. So (1) implies (2). Conversely, if (2) holds, and $\epsilon > 0$ is given, then by (2) there exists $\delta > 0$ such that $|f(x) - f(c)| < \epsilon$ for all $x \in A$ satisfying $|x - c| < \delta$. In particular, if $x \in A$ and $0 < |x - c| < \delta$, we have $|f(x) - f(c)| < \epsilon$. Thus $\lim_{x\to c} f(x) = f(c)$, so f is continuous at c. Thus (1) and (2) are equivalent.

Still assuming that c is a limit point of A, we now show that (2) implies (3). Suppose (x_n) is a sequence such that $x_n \in A$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} x_n = c$. To show that $\lim_{n\to\infty} f(x_n) = f(c)$, let $\epsilon > 0$. By (2), there exists $\delta > 0$ such that $|f(x) - f(c)| < \epsilon$ for all $x \in A$ such that $|x - c| < \delta$. Since $x_n \to c$, there exists $N \in \mathbb{N}$ such that $|x_n - c| < \delta$ for all $n > N$. Therefore for all $n > N$, $|f(x) - f(c)| < \epsilon$. Hence $\lim_{n\to\infty} f(x_n) = f(c)$, so (3) holds.

To complete the proof, it suffices to show that (3) implies (1). Suppose (3) holds. Then in particular, we have $\lim_{n\to\infty} f(x_n) = f(c)$ for all sequences (x_n) such that $x_n \in A, x_n \neq c$ for all $n \in \mathbb{N}$, and $\lim_{n\to\infty} x_n = c$. Then by the characterization of limits via sequences (Proposition 18.0.8), $\lim_{x\to c} f(x) = f(c)$. Hence (1) holds. ■

Next we consider some basic properties of continuous functions that correspond to the properties of limits in Theorem 14.0.4 and Proposition 18.0.10. First we need a lemma, whose proof we leave as an exercise.

Lemma 19.0.8 *Suppose $A \subseteq \mathbb{R}, f : A \to \mathbb{R}$ is a function, $c \in A, f$ is continuous at c, and $f(c) \neq 0$. Then there exists $r > 0$ such that $f(x) \neq 0$ for all $x \in (c - r, c + r) \cap A$.*

Proposition 19.0.9 *Suppose $A \subseteq \mathbb{R}, f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ are functions, $c \in A$ and f and g are continuous at c. Then*

(1) αf is continuous at c, for all $\alpha \in \mathbb{R}$;

(2) $f+g$ is continuous at c;

(3) $f \cdot g$ is continuous at c;

(4) if $g(c) \neq 0$, then $\frac{f}{g}$ is continuous at c.

To be precise about (4), the domain of $\frac{f}{g}$ is restricted to the points $x \in A$ where $g(x) \neq 0$. By Lemma 19.0.8, g is not zero on an interval containing c, and (4) means that $\frac{f}{g}$, defined only at the points where g is not zero, is continuous at c.

PROOF. If c is an isolated point of A, then $\alpha f, f+g, f \cdot g$, and $\frac{f}{g}$ are automatically continuous at c. If c is a limit point of A, then by definition of continuity, $\lim_{x\to c} f(x)$ and $\lim_{x\to c} g(x)$ exist, $\lim_{x\to c} f(x) = f(c)$, and $\lim_{x\to c} g(x) = g(c)$. Then by Proposition 18.0.10, $\lim_{x\to c}(\alpha f), \lim_{x\to c}(f+g)$, and $\lim_{x\to c}(f \cdot g)$ exist, with

(1) $\lim_{x\to c}(\alpha f)(x) = \alpha \lim_{x\to c} f(x) = \alpha f(c) = (\alpha f)(c)$;

(2) $\lim_{x\to c}(f+g)(x) = \lim_{x\to c} f(x) + \lim_{x\to c} g(x) = f(c) + g(c) = (f+g)(c)$;

(3) $\lim_{x\to c}(f \cdot g)(x) = (\lim_{x\to c} f(x)) \cdot (\lim_{x\to c} g(x)) = f(c) \cdot g(c) = (f \cdot g)(c)$.

By Lemma 19.0.8, there exists $r > 0$ such that $g(x) \neq 0$ for $x \in (c-r, c+r) \cap A$, so we can apply part (4) of Proposition 18.0.10 to conclude that

(4) $\lim_{x\to c} \left(\frac{f}{g}\right)(x) = \frac{\lim_{x\to c} f(x)}{\lim_{x\to c} g(x)} = \frac{f(c)}{g(c)} = \left(\frac{f}{g}\right)(c)$.

Hence $\alpha f, f+g, f \cdot g$, and $\frac{f}{g}$ are continuous at c. ■

Corollary 19.0.10 *If $p : \mathbb{R} \to \mathbb{R}$ is a polynomial (that is, $p(x) = \sum_{k=0}^{n} a_k x^k$ for some $n \in \mathbb{N}$ and coefficients $a_0, a_1, \ldots, a_n \in \mathbb{R}$), then p is continuous at every point of $\mathbb{R}$. If p and q are polynomials and $q(c) \neq 0$, where $c \in \mathbb{R}$, then $\frac{p}{q}$ (defined at all points where q is non-zero) is continuous at c.*

PROOF. Since the functions $f(x) = 1$ and $f(x) = x$ is continuous, and the product of continuous functions is continuous (Proposition 19.0.9 (3)), a simple induction argument gives that x^n is continuous for each $n \in \mathbb{N} \cup 0$. Then cx^n is continuous for any constant c (Proposition 19.0.9 (1)), and the sum of continuous functions is continuous (Proposition 19.0.9 (2)), so p is continuous. The continuity of $\frac{p}{q}$ at points where q is non-zero follows from Proposition 19.0.9 (4). ■

A function of the form $\frac{p}{q}$, where p and q are polynomials, is called a *rational function*, because it is the ratio of polynomials.

Continuity on a Set

So far we have only defined continuity at a point. We define continuity on a set in the natural way, as follows.

Definition 19.0.11 *Let $A \subseteq \mathbb{R}$, and let $f : A \to \mathbb{R}$ be a function. We say that f is continuous on A is f is continuous at c, for all $c \in A$.*

There is an important characterization of continuity of a function on a set in terms of inverse images. If the domain of the function is an open set, that characterization can be stated in the following way.

Theorem 19.0.12 *Suppose $O \subseteq \mathbb{R}$ is an open set, and $f : O \to \mathbb{R}$ is a function. Then f is continuous on O if and only if $f^{-1}(U)$ is an open set for all open sets $U \subseteq \mathbb{R}$.*

PROOF. First suppose f is continuous on O. Let $U \subseteq \mathbb{R}$ be open. Let $c \in f^{-1}(U)$. Then $f(c) \in U$. Since U is open, there exists $\epsilon > 0$ such that $(f(c)-\epsilon, f(c)+\epsilon) \subseteq U$. Since f is continuous, there exists $\delta > 0$ such that if $x \in (c-\delta, c+\delta) \cap O$, then $|f(x) - f(c)| < \epsilon$, which means that $f(x) \in (f(c)-\epsilon, f(c)+\epsilon) \subseteq U$.

Therefore $(c-\delta, c+\delta) \cap O \subseteq f^{-1}(U)$. However, since O is open and $c \in S$, there exists $s > 0$ such that $(c-s, c+s) \subseteq O$. Hence for $r = \min(\delta, s) > 0$, we have that $(c-r, c+r) \subseteq f^{-1}(U)$. Thus, $f^{-1}(U)$ is open.

Conversely, let $c \in O$ and $\epsilon > 0$. Let $U = (f(c)-\epsilon, f(c)+\epsilon)$. Then U is open, so by assumption, $f^{-1}(U)$ is open. Since $c \in f^{-1}(U)$, then there exists $\delta > 0$ such that $(c-\delta, c+\delta) \subseteq f^{-1}(U)$. That means that for all $x \in (c-\delta, c+\delta)$, we have $f(x) \in U = (f(c)-\epsilon, f(c)+\epsilon)$. In other words, if $|x-c| < \delta$, then $|f(x)-f(c)| < \epsilon$. Therefore f is continuous on O. ■

There is a version of Theorem 19.0.12 for f defined on a set A that is not necessarily open, but that version requires the notion of *relatively open* sets. For a set $A \subseteq \mathbb{R}$, we say that a subset V of A is relatively open in A if, for each $x \in V$, there exists an $r > 0$ such that $(x-r, x+r) \cap A \subseteq V$. For example, the set $(\frac{1}{2}, 1]$ is relatively open in $A = [0, 1]$, because, for example at $x = 1 \in A$, we have $(1-\frac{1}{2}, 1+\frac{1}{2}) \cap [0, 1] = (\frac{1}{2}, 1] \subseteq V$. In other words, in considering relatively open sets in A, we only consider the points in A, as if A were the entire space we are considering. Alternately, a set $V \subseteq A$ is relatively open if and only if there exists O open in $\mathbb{R}$ such that $V = A \cap O$. In the example just considered, where $A = [0, 1]$, we have that $V = (\frac{1}{2}, 1]$ is relatively open in A because $V = (\frac{1}{2}, 1] = (\frac{1}{2}, \frac{3}{2}) \cap [0, 1] = O \cap A$ where $O = (\frac{1}{2}, \frac{3}{2})$ is open in $\mathbb{R}$.

The analogue of Theorem 19.0.12 for general subsets A of $\mathbb{R}$ states that $f : A \to \mathbb{R}$ is continuous on A if and only if $f^{-1}(U)$ is relatively open in A for every open set $U \subseteq \mathbb{R}$. The proof is the same as the proof of Theorem 19.0.12, except slightly easier in one direction because there is no need to find the quantity $s > 0$.

Continuous Functions on Compact Sets

In calculus we study optimization, which means finding points where some function attains a maximum or minimum. We may be interested in maximizing profit, minimizing cost, or designing a box with maximum volume for a given surface area. The first question we should ask is when the optimization problem has an answer. For example, if we consider the function $f(x) = x$ on the domain $(0, 1)$, there is no point $x \in (0, 1)$ where f attains its largest or its smallest value. The problem is not with the function, which is as nice as possible. The problem is with the domain of definition of the function. Another problem that can occur is that f may be unbounded; for example the function $f(x) = \frac{1}{x}$ is unbounded on the domain $(0, 1)$, even though f is continuous on $(0, 1)$ (e.g., by Corollary 19.0.10). However, if f is continuous and the domain of f is compact, then these problems do not occur. The next result states that a continuous function on a compact set is bounded and attains its maximum and minimum. Readers with good memories will realize that we finally have all of the necessary machinery to carry out the approach outlined in the first paragraph of Chapter 15.

Theorem 19.0.13 *Suppose $K \subseteq \mathbb{R}$ is a non-empty compact set, and $f : K \to \mathbb{R}$ is continuous. Then f is bounded,*

(1) there exists $x_0 \in K$ such that $f(x_0) = \sup f(K)$, and

(2) there exists $x_1 \in K$ such that $f(x_1) = \inf f(K)$.

Proof. We use the fact from Corollary 17.0.20 that K is sequentially compact. We first prove that f is bounded, by contradiction. Suppose f is unbounded. Then for each $n \in \mathbb{N}$, there exists $x_n \in K$ such that $|f(x_n)| > n$. Since K is sequentially compact, there is a convergent subsequence (x_{n_k}) of (x_n), with $x = \lim_{k\to\infty} x_{n_k} \in K$. Since f is continuous at x, we have $\lim_{k\to\infty} f(x_{n_k}) = f(x)$. But $|f(x_{n_k})| \geq n_k \geq k$ for each $k \in \mathbb{N}$, hence $(f(x_{n_k}))$ is an unbounded sequence and hence is divergent (Proposition 14.0.3). This contradiction shows that f is bounded.

We prove (1), leaving the analogous proof of (2) as an exercise. Since f is bounded, the set $f(K)$ is bounded, so it has a supremum. Let $\alpha = \sup f(K)$. Then (by Lemma 10.0.7, or just because $\alpha - \frac{1}{n}$ is not an upper bound for $f(K)$), there exists $x_n \in K$ such that $f(x_n) > \alpha - \frac{1}{n}$. Then $\alpha - \frac{1}{n} < f(x_n) \leq \alpha$ (since α is an upper bound for $f(K)$), or $|f(x_n) - \alpha| < \frac{1}{n}$. Hence $\lim_{n\to\infty} f(x_n) = \alpha$. Since K is sequentially compact, there exists a a convergent subsequence (x_{n_k}) of (x_n), with $x_0 = \lim_{k\to\infty} x_{n_k} \in K$. Then $(f(x_{n_k}))$ is a subsequence of $f(x_n)$, so $\lim_{k\to\infty} f(x_{n_k}) = \alpha$ also (Proposition 15.0.2). Since f is continuous at x_0, we have $f(x_0) = \lim_{k\to\infty} f(x_{n_k}) = \alpha$. ■

Notice how this last proof brings together several major themes of this course so far: sups, sequences, subsequences, compactness, and continuity.

Chapter 20

Uniform Continuity

Continuity is a good property for a function to have, but for some purposes continuity is not good enough. For example, the function $f(x) = \frac{1}{x}$ is continuous on the domain $(0, 1)$, but f is not bounded. There is a stronger notion, called *uniform continuity*, that often is sufficient.

Definition 20.0.1 *Suppose $A \subseteq \mathbb{R}$. A function $f : A \to \mathbb{R}$ is uniformly continuous (or, uniformly continuous on A) if, for all $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ for all $x, y \in A$ such that $|x - y| < \delta$.*

This definition may appear to be the same as the $\epsilon - \delta$ characterization of continuity (Proposition 19.0.7 (2)), but it is not. In that characterization, the point c and $\epsilon > 0$ are given; then the statement is that there exists $\delta > 0$ such that $|f(x) - f(c)| < \epsilon$ for all $x \in A$ satisfying $|x - c| < \delta$. The fact that the existence of δ is stated after c and ϵ have been introducted means that δ may depend on both c and ϵ. In the definition of uniform continuity, the existence of δ is stated after only ϵ is introduced, so δ is only allowed to depend on ϵ. In particular, this δ gives the estimate $|f(x) - f(y)| < \epsilon$ for all $x, y \in A$ satisfying $|x - y| < \delta$. That is, in uniform continuity, the same δ works uniformly for all $x, y \in A$. Although δ depends on ϵ in the definition of uniform continuity, δ must be independent of x and y.

It should be evident that the uniform continuity of f on A implies the continuity of f on A. Uniform continuity is not easily characterized in terms of limits or sequences (compare to Proposition 19.0.7), because those characterizations are naturally local, meaning holding at each point, rather than uniform. We consider some examples, beginning with a trivial one.

Example 20.0.2 *Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x$. Then f is uniformly continuous.*

Proof. Let $\epsilon > 0$. Let $\delta = \epsilon$. Then if $x, y \in \mathbb{R}$ with $|x - y| < \delta = \epsilon$, it follows that

$$|f(x) - f(y)| = |x - y| < \epsilon.$$

Hence f is uniformly continuous on $\mathbb{R}$. ■

Next we consider three examples of continuous functions which are not uniformly continuous.

Example 20.0.3 *Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Then f is not uniformly continuous on $\mathbb{R}$.*

Proof. Let $c \in \mathbb{R}$ be fixed, with $c > 0$ positive, and consider $x \in \mathbb{R}$ with $x > 0$. We consider $\epsilon = 1$ and see what conditions on δ must hold if $|f(x) - f(c)| < \epsilon = 1$. Suppose

$$|f(x) - f(c)| = |x^2 - c^2| = |x - c||x + c| < 1.$$

Then we must have $|x - c| < \frac{1}{|x+c|} = \frac{1}{x+c} < \frac{1}{c}$. Thus for $\epsilon = 1$ we would need $\delta \leq \frac{1}{c}$ As $c \to \infty$, this would force δ to go to 0. Hence there is no $\delta > 0$ that works for $\epsilon = 1$ and all c. So f is not uniformly continuous. ■

Notice that Examples 20.0.2 and 20.0.3 show that the product of two uniformly continuous functions (or even the square of a uniformly continuous function) is not necessarily uniformly continuous, in contrast to the fact (Proposition 19.0.9 (3)) that the product of two continuous functions is continuous.

The failure of uniform continuity in Example 20.0.3 resulted from the behavior of $f(x)$ as $x \to \infty$, but uniform continuity can fail for a continuous function on a bounded set also.

Example 20.0.4 *Define $f : (0,1) \to \mathbb{R}$ by $f(x) = \frac{1}{x}$. Then f is not uniformly continuous.*

PROOF. For a point $c \in (0,1)$ and $\epsilon = 1$, we consider what δ must satisfy so that

$$|f(x) - f(c)| = \left|\frac{1}{x} - \frac{1}{c}\right| < 1.$$

Then $\frac{1}{x} \in \left(\frac{1}{c} - 1, \frac{1}{c} + 1\right) = \left(\frac{1-c}{c}, \frac{1+c}{c}\right)$, hence $x \in \left(\frac{c}{1+c}, \frac{c}{1-c}\right)$. The length of the interval $\left(\frac{c}{1+c}, \frac{c}{1-c}\right)$ is $\frac{c}{1-c} - \frac{c}{1+c} = \frac{2c^2}{1-c^2}$. As c approaches 0, $\frac{2c^2}{1-c^2} \to 0$, so the value of δ corresponding to $\epsilon = 1$ goes to 0 as $c \to 0$. Thus there is no uniform δ for $\epsilon = 1$, so f is not uniformly continuous. ■

Example 20.0.5 *Define $f : (0,1) \to \mathbb{R}$ by $f(x) = \sin\left(\frac{1}{x}\right)$. Then f is not uniformly continuous.*

PROOF. As noted in Example 18.0.5, $f\left(\frac{1}{2n\pi + \frac{\pi}{2}}\right) = 1$ and $f\left(\frac{1}{2n\pi}\right) = 0$. Let $\epsilon = \frac{1}{2}$. Then

$$\left|f\left(\frac{1}{2n\pi + \frac{\pi}{2}}\right) - f\left(\frac{1}{2n\pi}\right)\right| = |1 - 0| = 1,$$

but $\left|\frac{1}{2n\pi+\frac{\pi}{2}} - \frac{1}{2n\pi}\right| \leq |0 - \frac{1}{2n\pi}| = \frac{1}{2n\pi} \to 0$. Thus for $\epsilon = \frac{1}{2}$, then no matter how small $\delta > 0$ is, we can find points $x, y \in (0,1)$ with $|x - y| < \delta$ and $|f(x) - f(y)| = 1 > \frac{1}{2} = \epsilon$. So f is not uniformly continuous. ■

In the last two examples, the limit of $f(x)$ as x approaches the boundary point 0 does not exist. It turns out that this situation does not occur if f is unifomly continuous. In fact, a uniformly continuous function can be extended continuously to the closure of its domain. To prove this important fact, first we need a lemma that says that uniformly continuous functions take Cauchy sequences to Cauchy sequences.

Lemma 20.0.6 *Suppose $A \subseteq \mathbb{R}$ and $f : A \to \mathbb{R}$ is uniformly continuous. Suppose (x_n) is a Cauchy sequence with $x_n \in A$ for all $n \in \mathbb{N}$. Then $(f(x_n))$ is a Cauchy sequence.*

PROOF. Let $\epsilon > 0$. Since f is uniformly continuous, there exists $\delta > 0$ be such that $|f(x) - f(y)| < \epsilon$ for all $x, y \in A$ such that $|x - y| < \delta$. Since (x_n) is Cauchy, there exists $N \in \mathbb{N}$ such that $|x_n - x_m| < \delta$ for all $n, m > N$. Hence $|f(x_n) - f(x_m)| < \epsilon$ for all $n > N$. Therefore $(f(x_n))$ is a Cauchy sequence. ■

Theorem 20.0.7 *Let $A \subseteq \mathbb{R}$ and suppose $f : A \to \mathbb{R}$ is uniformly continuous. Then there exists a function $g : \overline{A} \to \mathbb{R}$ such that g is uniformly continuous on $\overline{A}$ and $g(x) = f(x)$ for $x \in A$.*

PROOF. Let $x \in \overline{A}$. Then there exists a sequence (x_n) with $x_n \in A$ for all $n \in \mathbb{N}$ such that $\lim_{n\to\infty} x_n = x$. (If $x \in A$, we can take $x_n = x$ for all n, but this is the trivial case.) Since (x_n) is convergent, (x_n) is Cauchy (Lemma 15.0.11). By Lemma 20.0.6, $(f(x_n))$ is a Cauchy sequence. Hence $(f(x_n))$ converges (Theorem 15.0.14). We would like to define $f(x) = \lim_{n\to\infty} f(x_n)$, but first we have to know that this proposed value is well-defined, in the sense that it does not depend on the choice of sequence (x_n) converging to x.

To reach this conclusion, suppose (x_n) and (y_n) are sequences of points of A, each of which converges to x. Consider the sequence

$$z = (x_1, y_1, x_2, y_2, x_3, y_3, \ldots);$$

in detail, $z_n = x_{(n+1)/2}$ if n is odd, and $z_n = y_{n/2}$ if n is even. Then $z_n \to x$, since $x_n \to x$ and $y_n \to x$. Thus, by the argument in the last paragraph, $\lim_{n\to\infty} f(z_n) = \alpha$ exists. But $(f(x_n))$ and $(f(y_n))$ are subsequences of $(f(z_n))$, hence by Proposition 15.0.2,

$$\lim_{n\to\infty} f(x_n) = \alpha = \lim_{n\to\infty} f(y_n).$$

That is, $\lim_{n\to\infty} f(x_n)$ is the same for all sequences (x_n) of elements of A which converge to x.

We define $g(x) = f(x)$ for all $x \in A$, and for $x \in \overline{A} \setminus A$ we define $g(x) = \lim_{n\to\infty} f(x_n)$ for any sequence (x_n) of points of A converging to x (which is well-defined by the last paragraph). By definition, g agrees with f on A.

We show that g is uniformly continuous on $\overline{A}$. Let $\epsilon > 0$ and let $x, y \in \overline{A}$. Since f is uniformly continuous on A, there exists $\delta > 0$ such that if $z, w \in A$ and $|z-w| < \delta$, then $|f(z)-f(w)| < \frac{\epsilon}{2}$. Suppose $|x-y| < \delta$. Let (x_n) be a sequence of elements of A converging to x and let (y_n) be a sequence of elements of A converging to y. Then there exists $N \in \mathbb{N}$ such that $|x - x_n| < \frac{\delta-|x-y|}{2}$ and $|y - y_n| < \frac{\delta-|x-y|}{2}$ for all $n > N$. Hence for all $n > N$ we have

$$|x_n - y_n| \le |x_n - x| + |x - y| + |y - y_n| < \frac{\delta - |x-y|}{2} + |x-y| + \frac{\delta - |x-y|}{2} = \delta - |x-y| + |x-y| = \delta.$$

Therefore $|f(x_n) - f(y_n)| < \frac{\epsilon}{2}$. Thus for all $n > N$, we have

$$|g(x) - g(y)| \le |g(x) - f(x_n)| + |f(x_n) - f(y_n)| + |f(y_n) - g(y)| < |g(x) - f(x_n)| + \frac{\epsilon}{2} + |f(y_n) - g(y)|.$$

Since $f(x_n) \to g(x)$ and $f(y_n) \to g(y)$ (by definition of g), we can choose $n > N$ sufficiently large so that $|g(x) - f(x_n)| < \frac{\epsilon}{4}$ and $|f(y_n) - g(y)| < \frac{\epsilon}{4}$. Then we obtain $|g(x) - g(y)| < \frac{\epsilon}{4} + \frac{\epsilon}{2} = \frac{\epsilon}{4} = \epsilon$, for all $x, y \in \overline{A}$ satisfying $|x - y| < \delta$. Hence g is absolutely continuous on $\overline{A}$. ■

Theorem 20.0.7 explains why a uniformly continuous function cannot blow up (i.e., go to $\pm\infty$) at a boundary point of its domain. In fact, a uniformly cointinuous function must be bounded on bounded sets. This fact can be proved directly, but it is an easy consequence of Theorem 20.0.7 and facts we learned earlier about compact sets.

Corollary 20.0.8 *Suppose $A \subseteq \mathbb{R}$ is a bounded set and $f : A \to \mathbb{R}$ is uniformly continuous. Then f is bounded.*

Proof. By Theorem 20.0.7, there exists a continuous function $g : \overline{A} \to \mathbb{R}$ which agrees with f on A. Since A is bounded, so is $\overline{A}$. Hence $\overline{A}$ is compact, since it is closed and bounded (Theorem 17.0.19). By Theorem 19.0.13, g is bounded on $\overline{A}$. Since f agrees with g on A, f is bounded on A. ■

We have seen in Theorem 19.0.13 how compactness, in the form of sequential compactness, can be used to turn a local property (like boundedness in a small interval aroound a point) into a global one (like boundedness on an entire compact set). The issue of local versus global control is exactly the distinction between continuity and uniform continuity. Hence perhaps the next theorem is not as surprising as it might otherwise be.

Theorem 20.0.9 *Suppose $K \subseteq \mathbb{R}$ is a compact set, and $f : K \to \mathbb{R}$ is continuous. Then f is uniformly continuous.*

Proof. Let $\epsilon > 0$. For each $z \in K$, there exists $\delta_z > 0$ such that $|f(x) - f(z)| < \frac{\epsilon}{2}$ for all $x \in K$ such that $|x - z| < \delta_z$. Let I_z denote the interval $I_z = \left(z - \frac{\delta_z}{2}, z + \frac{\delta_z}{2}\right)$. Then each I_z is open, and $K \subseteq \cup_{z\in K} I_z$ (since if $z_0 \in K$, then $z_0 \in I_{z_0} \subseteq \cup_{z\in K} I_z$). Since K is compact, there exist $n \in \mathbb{N}$ and $z_1, z_2, \ldots, z_n \in K$ such that $K \subseteq \cup_{j=1}^n I_{z_j}$. Let $\delta = \min\left\{\frac{\delta_{z_1}}{2}, \frac{\delta_{z_2}}{2}, \ldots, \frac{\delta_{z_n}}{2}\right\}$. Note that $\delta > 0$ since the minimum of finitely many positive numbers is positive.

Now let $x, y \in K$ satisfy $|x - y| < \delta$. (Observe that $\delta > 0$ has been chosen before x, y, so δ is independent of x, y). Since $K \subseteq \cup_{j=1}^n I_{z_j}$, there exists $j \in \{1, 2, \ldots, , n\}$ such that $x \in I_{z_j} = \left(z - \frac{\delta_{z_j}}{2}, z + \frac{\delta_{z_j}}{2}\right)$, so $|x - z_j| < \frac{\delta_{z_j}}{2}$. Note $\delta \le \frac{\delta_{z_j}}{2}$ since $\delta = \min\left\{\frac{\delta_{z_1}}{2}, \frac{\delta_{z_2}}{2}, \ldots, \frac{\delta_{z_n}}{2}\right\}$. Hence

$$|y - z_j| \le |y - x| + |x - z_j| < \delta + \frac{\delta_{z_j}}{2} \le \frac{\delta_{z_j}}{2} + \frac{\delta_{z_j}}{2} = \delta_{z_j}.$$

Since $|y - z_j| < \delta_{z_j}$, we have $|f(y) - f(z_j)| < \frac{\epsilon}{2}$ and similarly $|f(x) - f(z_j)| < \frac{\epsilon}{2}$, since $|x - z_j| < \frac{\delta_{z_j}}{2} < \delta_{z_j}$. Therefore

$$|f(x) - f(y)| \le |f(x) - f(z_j)| + |f(z_j) - f(y)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

■

Again we see the advantage of compactness: if the domain of a continuous function is a compact set, then the function automatically has the stronger property of uniform continuity. This fact will turn out to be the key to proving that a continuous function on a closed finite interval is Riemann integrable, which is one of the fundamental results about Riemann integration.

Chapter 21

Differential Calculus on $\mathbb{R}$

We are now ready to introduce the notion of the derivative of a function.

Definition 21.0.1 *Suppose $a, b \in \mathbb{R}$ with $a < b$. Suppose $f : (a, b) \to \mathbb{R}$ is a function and $x \in (a, b)$. We say that f is differentiable at x if*

$$\lim_{y \to x} \frac{f(y) - f(x)}{y - x}$$

exists. If this limit exists, we call it $f'(x)$, the derivative of f at x.

If we write $y = x + h$, then $y \to x$ is equivalent to $h \to 0$, so an alternate formulation of the definition of $f'(x)$ is

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

if that limit exists. The quantity $\frac{f(y)-f(x)}{y-x}$, or equivalently $\frac{f(x+h)-f(x)}{h}$, is called a *difference quotient* of f. Graphically it represents the slope of the line through the points $(x, f(x))$ and $(y, f(y))$, which can be interpreted as the average rate of change of f with respect to x over the interval $[x, y]$. Letting $y \to x$, we obtain the physical interpretation of $f'(x)$ as the instantaneous rate of change of the function f with respect to x, at the point x, and the graphical interpretation of $f'(x)$ as the slope of the tangent line to the graph of f at the point x.

Example 21.0.2 *Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Prove that f is differentiable at every point $x \in \mathbb{R}$, and $f'(x) = 2x$.*

Proof. For a given x and $y \neq x$,

$$\frac{f(y) - f(x)}{y - x} = \frac{y^2 - x^2}{y - x} = y + x.$$

Since $\lim_{y \to x}(y + x)$ exists and equals $2x$, we obtain that f is differentiable at x and $f'(x) = 2x$. ■

Differentiability implies a degree of "smoothness" of the function. In particular a function that is differentiable at a point must be continuous at that point.

Proposition 21.0.3 *Suppose Suppose $a, b \in \mathbb{R}$ with $a < b$. Suppose $f : (a, b) \to \mathbb{R}$ is a function and $x \in (a, b)$. If f is differentiable at x, then f is continuous at x.*

Proof. We have $\lim_{y \to x} \frac{f(y)-f(x)}{y-x} = f'(x)$ and $\lim_{y \to x}(y - x) = 0$, so by Proposition 18.0.10 (3),

$$f(y) = f(x) + \frac{f(y) - f(x)}{y - x} \cdot (y - x) \xrightarrow{y \to x} f(x) + f'(x) \cdot 0 = f(x).$$

Hence f is continuous at x. ■

Although differentiability implies continuity, continuity does not imply differentiability.

Example 21.0.4 *Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = |x|$. Then f is continuous at 0 but not differentiable at $x = 0$.*

PROOF. The continuity of f at 0 follows because $\lim_{x\to 0} |x| = 0$. However, for $y > 0$, $\frac{f(y)-f(0)}{y-0} = \frac{|y|}{y} = +1$, and for $y < 0$, $\frac{f(y)-f(0)}{y-0} = \frac{|y|}{y} = -1$. Hence $\lim_{y\to 0} \frac{f(y)-f(0)}{y-0}$ does not exist. ■

In Example 21.0.4, the graph of f has a sharp corner at 0. In addition to being continuous, differentiable functions should not have sharp corners. So differentiability implies more smoothness than continuity.

In calculus we learn the following rules for differentiation.

Proposition 21.0.5 *Suppose $a, b \in \mathbb{R}$ with $a < b$, f and g are functions defined on the interval (a, b), and $x \in (a, b)$. Suppose f and g are differentiable at x. Then for $c \in \mathbb{R}$, we have that $cf, f + g$, and fg are differentiable at x, with*

(i) $(cf)'(x) = cf'(x)$ (scalar homogeneity);
(ii) $(f + g)'(x) = f'(x) + g'(x)$ (additivity);
and
(iii) $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$ (product, or Leibniz, rule).

If $g(x) \neq 0$, then $\frac{f}{g}$ is differentiable at x and
(iv) $\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$ (quotient rule).

PROOF. In each case we compute the difference quotient and use the assumed existence of f' and/or g' to take the limit. The existence of the limit shows the asserted differentiability, and the value obtained for the limit yields the differentiation formula. For (i), we use the scalar homogeneity property of limits (Proposition 18.0.10 (1)) to obtain

$$(cf)'(x) = \lim_{y\to x} \frac{(cf)(y) - (cf)(x)}{y - x} = \lim_{y\to x} c\frac{(f)(y) - (f)(x)}{y - x} = c \lim_{y\to x} \frac{f(y) - f(x)}{y - x} = cf'(x).$$

For (ii), we use the additivity property of limits (Proposition 18.0.10 (2)):

$$(f + g)'(x) = \lim_{y\to x} \frac{(f + g)(y) - (f + g)(x)}{y - x} = \lim_{y\to x} \frac{(f(y) + g(y) - f(x) - g(x)}{y - x}$$

$$= \lim_{y\to x} \left(\frac{f(y) - f(x)}{y - x} + \frac{g(y) - g(x)}{y - x} \right) = \lim_{y\to x} \frac{f(y) - f(x)}{y - x} + \lim_{y\to x} \frac{g(y) - g(x)}{y - x} = f'(x) + g'(x).$$

For (iii), we have

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

$$= \lim_{y\to x} \left(\frac{f(x + h)g(x + h) - f(x)g(x + h)}{y - x} + \frac{f(x)g(x + h) - f(x)g(x)}{y - x} \right)$$

$$= \lim_{y\to x} g(x + h)\frac{f(x + h) - f(x)}{y - x} + f(x) \lim_{y\to x} \frac{g(x + h) - g(x)}{y - x}$$

$$= g(x)f'(x) + f(x)g'(x),$$

using Proposition 18.0.10 (2) and (3), and the fact that $\lim_{y\to x} g(x + h) = g(x)$ because the differentiability of g at x implies that g is continuous at x (Proposition 21.0.3).

We leave the proof of (iv) as an exercise.

■

Perhaps the most powerful basic differentiation rule is the *chain rule*, as follows.

Proposition 21.0.6 *(Chain rule) Suppose $a, b, r, s \in \mathbb{R}$ with $a < b$ and $r < s$, $f : (a, b) \to (r, s)$ is differentiable at some point $x \in (a, b)$, and $g : (r, s) \to \mathbb{R}$ is differentiable at $f(x)$. Then $g \circ f : (a, b) \to \mathbb{R}$ is differentiable at x, with*

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

It is tempting to give the following "proof" of the chain rule:

$$(g \circ f)'(x) = \lim_{y \to x} \frac{g \circ f(y) - g \circ f(x)}{y - x} = \lim_{y \to x} \frac{g(f(y) - g(f(x))}{f(y) - f(x)} \cdot \frac{f(y) - f(x)}{y - x}$$

$$= \lim_{y \to x} \frac{g(f(y) - g(f(x))}{f(y) - f(x)} \cdot \lim_{y \to x} \frac{f(y) - f(x)}{y - x} = g'(f(x))f'(x),$$

where we use the continuity of f at x to say that $f(y) \to f(x)$ as $y \to x$. This argument is almost right, but it doesn't take into account that the quantity $f(y) - f(x)$ could take the value 0 for y in every interval around x, so that we have divided by 0, which is undefined. Instead, we have to describe the derivative in a more linear manner.

Proof of Proposition 21.0.6. For any function $h : (\alpha, \beta) \to \mathbb{R}$ which is differentiable at $x \in (\alpha, \beta)$, where $\alpha, \beta \in \mathbb{R}$ with $\alpha < \beta$, define

$$E_h(x, y) = \begin{cases} \frac{h(y)-h(x)}{y-x} - h'(x), & \text{if } y \in (\alpha, \beta) \setminus \{x\} \\ 0 & \text{if } y = x\ . \end{cases}$$

(Here "E" stands for the error made in approximating $\dfrac{h(y) - h(x)}{y - x}$ by $h'(x)$.) Since h is differentiable at x, we have $\lim_{y \to x} E_h(x, y) = \lim_{y \to x} \frac{h(y)-h(x)}{y-x} - h'(x) = h'(x) - h'(x) = 0$. Hence $E_h(x, y)$ is a continuous function of y at $y = x$.

Letting $h = g$, replacing x with $f(x)$, and replacing y with z, we have

$$E_g(f(x), z) = \begin{cases} \frac{g(z)-g(f(x))}{z-f(x)} - g'(f(x)) & \text{if } z \in (r, s) \setminus \{f(x)\} \\ 0 & \text{if } z = f(x)\ . \end{cases} \tag{21.1}$$

Since g is assumed to be differentiable at $f(x)$, we have $E_g(f(x), z) \overset{z \to f(x)}{\longrightarrow} 0$. Multiplying equation (21.1) on both sides by $z - f(x)$ and solving for $g(z) - g(f(x))$ gives

$$g(z) - g(f(x)) = (z - f(x)) \left[E_g(f(x), z) + g'(f(x))\right], \tag{21.2}$$

if $z \neq f(x)$. Note that equation (21.2) is valid at $z = f(x)$ also, since both sides of the equation evaluate to 0 when $z = f(x)$. Since equation 21.2 holds for all z, we can replace z with $f(y)$, for $y \in (a, b)$ satisfying $y \neq x$, and divide by $y - x$ on both sides to obtain

$$\frac{g \circ f(y) - g \circ f(x)}{y - x} = \frac{f(y) - f(x)}{y - x} \left[E_g(f(x), f(y)) + g'(f(x))\right], \text{ for } y \neq x.$$

As $y \to x$, the continuity of f at x (which holds because f is differentiable at x, by assumption) implies that $f(y) \to f(x)$, hence $E_g(f(x), f(y)) \to 0$. Thus taking the limit as $y \to x$ gives that

$$(g \circ f)'(x) = \lim_{y \to x} \frac{g \circ f(y) - g \circ f(x)}{y - x} = \lim_{y \to x} \frac{f(y) - f(x)}{y - x} \cdot \lim_{y \to x} \left[E_g(f(x), f(y)) + g'(f(x))\right]$$

$$= f'(x) \left[0 + g'(f(x))\right] = g'(f(x))f'(x).$$

□

For the next example, we assume the basic calculus facts about the functions $\sin x$ and $\cos x$, including the fact that the derivative of $\sin x$ is $\cos x$.

Example 21.0.7 *Define $f : \mathbb{R} \to \mathbb{R}$ by*

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

Determine whether f is differentiable at $x = 0$. If f is differentiable at $x = 0$ determine whether $f'(x)$ is continuous at $x = 0$.

Solution: It is tempting to use the rules we learn in calculus class to compute

$$f'(x) = 2x \sin\left(\frac{1}{x}\right) + x^2 \left[\cos\left(\frac{1}{x}\right)\right] \cdot \left(-\frac{1}{x^2}\right) = 2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right).$$

As $x \to 0$, the term $2x \sin\left(\frac{1}{x}\right)$ converges to 0, because $\sin\left(\frac{1}{x}\right)$ is bounded and the term x goes to 0. However, the term $\cos\left(\frac{1}{x}\right)$ does not have a limit as $x \to 0$, similarly to $\sin\left(\frac{1}{x}\right)$, as discussed in Example 18.0.5. So one might be convinced that f is not differentiable at $x = 0$. However, that conclusion is not correct. What we just did was not to determine whether $f'(0)$ exists. What we did was find $f'(x)$ for $x \neq 0$ and then we showed that $\lim_{x\to 0} f'(x)$ does not exist. To determine whether $f'(0)$ exists, we need to apply the definition of the derivative to determine whether $f'(0) = \lim_{y\to 0} \frac{f(y)-f(0)}{y-0}$ exists. Since $f(0) = 0$ by definition, we have

$$\frac{f(y) - f(0)}{y - 0} = \frac{y^2 \sin\left(\frac{1}{y}\right) - 0}{y - 0} = y \sin\left(\frac{1}{y}\right),$$

which converges to 0 as $y \to 0$, since $\left|y \sin\left(\frac{1}{y}\right)\right| \leq |y| \to 0$ as $y \to 0$. Hence f is differentiable at 0 and $f'(0) = 0$.

Note that f is differentiable at all points $x \in \mathbb{R}$, since we just checked $x = 0$ and the formula applies for all $x \neq 0$. The earlier computation shows that $\lim_{x\to 0} f'(x)$ does not exist, which shows in particular that $f'(x)$ is not continuous at $x = 0$.

One of the primary areas of application of the derivative is in optimization problems. For example, in business, one wants to choose the value of a parameter to obtain the maximum profit or the minimum cost.

Definition 21.0.8 *Suppose I is an interval (open, closed, or half-open), $x_0 \in I$, and $f : I \to \mathbb{R}$ is a function. We say f has a local maximum at x_0 if there exists $\delta > 0$ such that $f(x) \leq f(x_0)$ for all $x \in I \cap (x_0 - \delta, x_0 + \delta)$. We say f has a global maximum at x_0 if $f(x) \leq f(x_0)$ for all $x \in I$. We say f has a local minimum at x_0 if there exists $\delta > 0$ such that $f(x) \geq f(x_0)$ for all $x \in I \cap (x_0 - \delta, x_0 + \delta)$. We say f has a global minimum at x_0 if $f(x) \geq f(x_0)$ for all $x \in I$.*

If f has a global maximum (respectively, minimum) at a point, then clearly it has a local maximum (respectively, minimum) at that point. The converse is not true. For example, consider $f : [-2, 2] \to \mathbb{R}$ defined by $f(x) = x^4 - x^2$. Then $f(0) = 0$ and $f(x) \leq 0$ for $-1 \leq x \leq 1$, so f has a local maximum at $x = 0$. However, $f(-2) = f(2) = 12$, and the global maximum of f on $[-2, 2]$ occurs at the points $x = \pm 2$.

The relevance of derivatives to optimization is demonstrated by the following result.

Proposition 21.0.9 *Suppose I is an interval, $f : I \to \mathbb{R}$ is a function, and f has a local maximum or minimum at $x_0 \in I$. Then either*

(i) x_0 is an endpoint of I;

(ii) f is not differentiable at x_0; or

(iii) f is differentiable at x_0 and $f'(x_0) = 0$.

Proof. Suppose f has a local maximum at $x_0 \in I$, and suppose that neither (i) nor (ii) holds. Then x_0 is not an endpoint of I and f is differentiable at x_0. Since f has a local maximum at x_0, there exists

$\delta > 0$ such that $f(y) \leq f(x_0)$ for all $y \in (x_0 - \delta, x_0 + \delta)$. Hence if $y \in (x_0 - \delta, x_0)$, we have $y - x_0 < 0$ and $f(y) - f(x_0) \leq 0$, hence $\frac{f(y)-f(x_0)}{y-x_0} \geq 0$. Therefore

$$f'(x_0) = \lim_{y \to x_0} \frac{f(y) - f(x_0)}{y - x_0} \geq 0. \tag{21.3}$$

However, if $y \in (x_0, x_0 + \delta)$, we have $y - x_0 > 0$ but $f(y) - f(x_0) \leq 0$, hence $\frac{f(y)-f(x_0)}{y-x_0} \leq 0$. Therefore

$$f'(x_0) = \lim_{y \to x_0} \frac{f(y) - f(x_0)}{y - x_0} \leq 0. \tag{21.4}$$

Equations 21.3 and 21.4 imply that $f'(x_0) = 0$, which completes the proof in the case where f has a local maximum at x_0.

The proof when f has a local minimum at x_0 is very similar, only with $f(y) \geq f(x_0)$ for all $y \in (x_0 - \delta, x_0 + \delta)$, for some $\delta > 0$, so that the signs of the right and left difference quotients are the opposite of what they are in the local maximum case. ■

The previous result justifies the "take the derivative and set it equal to 0" rule that is used in max-min problems in calculus.

Definition 21.0.10 *Suppose $a, b \in \mathbb{R}$ with $a < b$, and suppose $f : (a, b) \to \mathbb{R}$ is a function. We say that f is differentiable on (a, b) if f is differentiable at x, for all $x \in (a, b)$.*

The next result is used to prove the Mean-Value Theorem, which is a key theoretical result about derivatives.

Theorem 21.0.11 *(Rolle's Theorem) Suppose $a, b \in \mathbb{R}, a < b$, and $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b). Suppose $f(a) = 0$ and $f(b) = 0$. Then there exists $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. Since $[a, b]$ is closed and bounded, $[a, b]$ is compact, by the Heine-Borel Theorem (Theorem 17.0.19). Since f is continuous on $[a, b]$, there exist points $x_0, x_1 \in [a, b]$ such that $f(x_0) = \sup\{f(x) : x \in [a, b]\}$ and $f(x_1) = \inf\{f(x) : x \in [a, b]\}$ (by Theorem 19.0.13).

If x_0 and x_1 are both endpoints of the interval $[a, b]$, then since $f(a) = 0$ and $f(b) = 0$ we get $\sup\{f(x) : x \in [a, b]\} = 0$ and $\inf\{f(x) : x \in [a, b]\} = 0$. Since the supremum and infimum are upper and lower bounds, respectively, for $f(x)$ for $x \in [a, b]$, we obtain $0 \leq f(x) \leq 0$ for all $x \in [a, b]$. Then $f(x) = 0$ for all $x \in [a, b]$, in which case f is differentiable at all points of (a, b), and $f'(x) = 0$ for all $x \in (a, b)$. So for any $c \in (a, b)$, we obtain $f'(c) = 0$.

If it is not true that both the supremum and infimum of f on $[a, b]$ are attained at the endpoints, then at least one of them is attained at an interior point, i.e., a point $c \in (a, b)$. Since f has a global maximum or minimum at c, then f has a local maximum or minimum at c, and so $f'(c) = 0$ by Proposition 21.0.9. ■

Rolle's Theorem leads to the following, which is a generalization of Rolle's Theorem.

Theorem 21.0.12 *(Mean-Value Theorem) Suppose $a, b \in \mathbb{R}$, with $a < b$. Suppose $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b). Then there exists $c \in (a, b)$ such that*

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

Proof. Define $\ell : [a, b] \to \mathbb{R}$ by $\ell(x) = f(a) + \frac{f(b)-f(a)}{b-a}(x - a)$. Then $\ell(a) = f(a) + 0 = f(a)$ and $\ell(b) = f(a) + f(b) - f(a) = f(b)$. In other words, the graph of ℓ is the line passing through the points $(a, f(a))$ and $(b, f(b))$. Note that ℓ is continuous on $[a, b]$ and differentiable on (a, b), with $\ell'(x) = \frac{f(b)-f(a)}{b-a}$ for $x \in (a, b)$ (in particular, ℓ' is constant). Define

$$g(x) = f(x) - \ell(x).$$

Then is continuous on $[a, b]$ and differentiable on (a, b) (since f and g are), with $g(a) = f(a) - \ell(a) = 0$ and $g(b) = f(b) - \ell(b) = 0$. Therefore by Rolle's Theorem (Theorem 21.0.11), there exists a point $c \in (a, b)$ such that $g'(c) = 0$. Since $g' = f' - \ell'$, that means that

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

Multiplying by $b-a$ gives $f(b)-f(a)=f'(c)(b-a)$. ■

The picture describing the proof of the Mean-Value Theorem is just a vertical translation and rotation of the picture in the proof of Rolle's Theorem. The Mean-Value Theorem easily implies the standard facts about intervals on which a differentiable function is either increasing or decreasing, which is used in a calculus course to graph functions by hand. If $f'(x)>0$ for all $x\in(a,b)$, we just say $f'>0$ on (a,b); if $f'(x)<0$ for all $x\in(a,b)$, we say $f'<0$ on (a,b), and similarly with "$<$" replaced everywhere by "$\leq$," or with "$>$" replaced everywhere by "$\geq$."

Proposition 21.0.13 *Suppose $a,b\in\mathbb{R}$ with $a<b$. Suppose $f:[a,b]\to\mathbb{R}$ is continuous, and f is differentiable on (a,b). Then:*

(a) if $f'>0$ on (a,b), then f is strictly increasing on $[a,b]$ (that means: if $s,t\in[a,b]$ satisfy $s<t$, then $f(s)<f(t)$),

(b) if $f'\geq 0$ on (a,b), then f is increasing (or nondecreasing) on $[a,b]$ (that means: if $s,t\in[a,b]$ satisfy $s<t$, then $f(s)\leq f(t)$),

(c) if $f'<0$ on (a,b), then f is strictly decreasing on $[a,b]$ (that means: if $s,t\in[a,b]$ satisfy $s<t$, then $f(s)>f(t)$),

(d) if $f'\leq 0$ on (a,b), then f is decreasing (or nonincreasing) on $[a,b]$ (that means: if $s,t\in[a,b]$ satisfy $s<t$, then $f(s)\geq f(t)$),

The proof is left as an exercise. One might think that if $f'(c)>0$, then there must be an interval $(c-\delta,c+\delta)$ for some $\delta>0$ on which f is increasing. If f' is continuous at c, that conclusion is justified, because then f' is positive in some interval around c (we are still assuming that f is differentiable on (a,b)). However, if we only have $f'(c)>0$ but we don't know f' is continuous at c, then there may not be any interval around c on which f is increasing. An example is

$$f(x)=\begin{cases} x^2\sin\left(\frac{1}{x}\right)+\frac{1}{2}x & \text{if } x\neq 0\\ 0 & \text{if } x=0\ . \end{cases}$$

We leave the verification that this example has the stated properties as an exercise.

An important consequence of Proposition 21.0.13 is the following.

Corollary 21.0.14 *Suppose $a,b\in\mathbb{R}$ with $a<b$. Suppose $f:[a,b]\to\mathbb{R}$ is continuous, and f is differentiable on (a,b). If $f'(x)=0$ for all $x\in(a,b)$, then f is a constant function on $[a,b]$; that is, there exists $c\in\mathbb{R}$ such that $f(x)=c$ for all $x\in[a,b]$.*

We leave the proof as an exercise.

If f is differentiable on an open interval (a,b), and if the derivative function f' is itself differentiable at a point $c\in(a,b)$, we let $f''(c)=(f')'(c)$. We call $f''(c)$ the *second derivative* of f at c. Similar reasoning to above (which is also left as an exercise) gives the second derivative test for a local maximum or minimum.

Proposition 21.0.15 *(Second Derivative Test) Suppose $a,b\in\mathbb{R}$ with $a<b$. Suppose $f:[a,b]\to\mathbb{R}$ is continuous, f is differentiable on (a,b), $f''(x)$ exists for all $x\in(a,b)$ and f'' is continuous on (a,b). Suppose $c\in(a,b)$.*

(a) If f has a local maximum at c, then $f''(c)\leq 0$.

(b) If f has a local minimum at c, then $f''(c)\geq 0$.

(c) If $f'(c)=0$ and $f''(c)<0$, then f has a local maximum at c.

(d) If $f'(c)=0$ and $f''(c)>0$, then f has a local minimum at c.

In the case where $f'(c)=0$ and $f''(c)=0$, the second derivative test is inconclusive. In fact, all possibilities may occur in that case. That is, there are examples where $f'(c)=0$ and $f''(c)=0$ but (i) f has a local maximum at c, and there are examples where (ii) f has a local minimum at c, and there are examples where (iii) f has neither a local maximum nor minimum at c.

Chapter 22

Riemann Integration in One Variable

One of the basic problems in mathematics is to determine the area of some two-dimensional region. Formulas for the areas of simple regions like rectangles, triangles, circles, and trapezoids, go back to ancient times. However, a method to determine the area of a relatively general area, call it A, say of a region below the graph of $y = f(x)$ and above the x-axis (for simplicity assume $f(x) \geq 0$) for $a \leq x \leq b$, was lacking until the development of calculus. One strategy is to find finitely many disjoint outer rectangles whose union contains the region, and use the sum of their areas to get an upper bound M for A, and to find finitely many disjoint inner rectangles which are contained in the region, and use the sum of their areas to get a lower bound m for A. Thus we have $m \leq A \leq M$. If we can show that as the number of rectangles increases, the lower and upper bounds converge to the same number, then that number must be the exact area.

This approach has many applications other than area, and is applied when f is not necessarily assumed to be non-negative (in the context of area, that would mean considering the *signed area*, in which area above the x-axis is counted as positive, and area below the x-axis is counted as negative). The general theory along these lines is called Riemann integration.

Definition 22.0.1 *Suppose $a, b \in \mathbb{R}$ with $a \leq b$. A partition of $[a, b]$ is an ordered set*

$$P = \{x_0, x_1, x_2, \ldots, x_n\},$$

for some $n \in \mathbb{N}$, such that

$$a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b.$$

For $j = 1, 2, \ldots, n$, we let $I_j = [x_{j-1}, x_j]$ be the j^{th} subinterval in the partition. We let $\Delta x_j = x_j - x_{j-1}$ be the length of I_j. We let $\|P\| = \max\{\Delta x_1, \Delta x_2, \ldots, \Delta x_n\}$ be the length of the longest interval in the partition; $\|P\|$ is called the norm of P.

Saying that P is an ordered set just means that we always list the elements of P in increasing order. For example, $\{0, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1\}$ is a partition of $[0, 1]$. We would not call $\{\frac{1}{3}, \frac{1}{2}, 1, 0, \frac{1}{4}\}$ a partition of $[0, 1]$ even though the two agree as sets. This slight abuse of notation allows us to use set notation; for example if P and Q are two partitions of $[a, b]$, we can consider the partition $P \cup Q$ whose elements are all points that are either in P or Q (or both), but listed in increasing order.

Notice that the lengths Δx_j of the intervals I_j sum to give the length $b-a$ of a, b. This fact is geometrically obvious, but the proof depends on the fact that the sum is telescoping, which means that all terms but the first and last cancel out:

$$\sum_{j=1}^{n} \Delta x_j = \sum_{j=1}^{n} (x_j - x_{j-1}) = x_n - x_{n-1} + x_{n-1} - x_{n-2} + \ldots + x_3 - x_2 + x_2 - x_1 + x_1 - x_0 = x_n - x_0 = b - a, \quad (22.1)$$

where when we wrote out the sum we wrote the terms in decreasing order of index rather that the usual order.

Associated with a partition P of $[a, b]$ and a bounded function $f : [a, b] \to \mathbb{R}$ are the upper and lower Riemann sums for f and P, as follows.

Definition 22.0.2 *Suppose $a, b \in \mathbb{R}$ with $a \leq b$, $f : [a, b] \to \mathbb{R}$ is a bounded function, and $P = \{x_0, x_1, x_2, \ldots, x_n\}$ is a partition of $[a, b]$. For $j = 1, 2, \ldots, n$, let*

$$m_j(f) = \inf\{f(x) : x \in [x_{j-1}, x_j]\}, \text{ and } M_j(f) = \sup\{f(x) : x \in [x_{j-1}, x_j]\}.$$

The lower Riemann sum $L(f, P)$ for f and P is

$$L(f, P) = \sum_{j=1}^{n} m_j(f)\Delta x_j.$$

The upper Riemann sum $U(f, P)$ for f and P is

$$U(f, P) = \sum_{j=1}^{n} M_j(f)\Delta x_j.$$

The suprema and infima in Definition 22.0.2 exist because of the assumption that f is bounded. Note that we are not assuming that f is continuous. Returning momentarily to the example of area calculation (and $f \geq 0$), if we look at the interval I_j and consider the rectangle with base equal to the interval I_j on the x-axis and height equal to $M_j(f)$, then since $M_j(f)$ is the supremum of the values of f on I_j, this rectangle contains the region under the graph of f over I_j. The height of this rectangle is $M_j(f)$ and its width is Δx_j, so its area is $M_j(f)\Delta x_j$. Summing these areas in the sum defining $U(f, P)$ shows that $U(f, P)$ is greater than the area under the graph of f for $a \leq x \leq b$. Similarly, since $m_j(f)$ is the infimum of f on I_j, the rectangle over I_j with height $m_j(f)$ lies inside the region under the graph of f over I_j. The area of this inscribed rectangle is $m_j(f)\Delta x_j$, so the lower sum $L(f, P)$ is a lower bound for the area under the graph of f for $a \leq x \leq b$.

Since $m_j = \inf\{f(x) : x \in [x_{j-1}, x_j]\} \leq \sup\{f(x) : x \in [x_{j-1}, x_j] = M_j(f)$ for all j, we have

$$L(f, P) = \sum_{j=1}^{n} m_j(f)\Delta x_j \leq \sum_{j=1}^{n} M_j(f)\Delta x_j = U(f, P), \tag{22.2}$$

for any partition P and bounded function f.

Definition 22.0.3 *Suppose $a, b \in \mathbb{R}$ with $a \leq b$. Suppose $P = \{x_0, x_1, \ldots, x_n\}$ and $Q = \{y_1, y_1, \ldots, y_m\}$ are partitions of $[a, b]$. We say Q is a refinement of P if $P \subseteq Q$ as sets; that is, $x_i \in Q$, for all $i = 0, 1, \ldots, n$.*

Lemma 22.0.4 *Suppose $a, b \in \mathbb{R}$ with $a \leq b$, $f : [a, b] \to \mathbb{R}$ is a bounded function, P and Q are partitions of $[a, b]$, and Q is a refinement of P. Then*

$$L(f, P) \leq L(f, Q) \leq U(f, Q) \leq U(f, P). \tag{22.3}$$

Proof. Since this result involves two partitions, we must be careful about notation. Let $P = \{x_0, x_1, \ldots, x_n\}$ and $Q = \{y_1, y_1, \ldots, y_m\}$ and write $m_j(f, P) = \inf\{f(x) : x \in [x_{j-1}, x_j]\}$ and $M_j(f, P) = \sup\{f(x) : x \in [x_{j-1}, x_j]\}$ instead of just $m_j(f)$ and $M_j(f)$, and similarly write $m_j(f, Q) = \inf\{f(x) : x \in [y_{j-1}, y_j]\}$ and $M_j(f, Q) = \sup\{f(x) : x \in [y_{j-1}, y_j]\}$. Since Q is a refinement of P, for each $j \in \{1, 2, \ldots, n\}$, there exist $k_j \in \{0, 1, \ldots, m\}$ such that $x_j = y_{k_j}$. We must have $k_{j-1} < k_j$, but there may be additional points $y_{k_{j-1}+1}, y_{k_{j-1}+2}, \ldots, y_{k_{j-1}+p}$ of Q between $x_{j-1} = y_{k_{j-1}}$ and $x_j = y_{k_j}$. These numbers satisfy

$$x_{j-1} = y_{k_{j-1}} < y_{k_{j-1}+1} < y_{k_{j-1}+2} < \ldots < y_{k_{j-1}+p} < y_{k_j} = x_j.$$

Notice then that

$$\begin{aligned}\sum_{\ell=k_{j-1}+1}^{k_j} \Delta y_\ell &= y_{k_j} - y_{k_j-1} + y_{k_j-1} - y_{k_j-2} + \ldots + y_{k_{j-1}+2} - y_{k_{j-1}+1} + y_{k_{j-1}+1} - y_{k_{j-1}} \\ &= y_{k_j} - y_{k_{j-1}} = x_j - x_{j-1} = \Delta x_j,\end{aligned} \tag{22.4}$$

for each $j \in \{1, 2, \ldots, n\}$, because the sum is telescoping (where we have written out the sum from the largest index to the smallest). Notice also that for $k_{j-1}+1 \leq \ell \leq k_j$, we have $x_{j-1} = y_{k_{j-1}} \leq y_{\ell-1} < y_\ell \leq y_{k_j} = x_j$, and hence $[y_{\ell-1}, y_\ell] \subseteq [x_{j-1}, x_j]$. Therefore

$$m_j(f,P) = \inf\{f(x) : x \in [x_{j-1}, x_j]\} \leq \inf\{f(x) : x \in [y_{\ell-1}, y_\ell]\} = m_\ell(f,Q) \tag{22.5}$$

and

$$M_\ell(f,Q) = \sup\{f(x) : x \in [y_{\ell-1}, y_\ell]\} \leq \sup\{f(x) : x \in [x_{j-1}, x_j]\} = M_j(f,P), \tag{22.6}$$

for $k_{j-1}+1 \leq \ell \leq k_j$ (because larger sets may have smaller infs and/or larger sups). To estimate the sums in $L(f,Q)$ and $U(f,Q)$, we break the sum over the points y_ℓ into segments consisting of those $y_\ell's$ between x_{j-1} and x_j, for each $j \in \{1, 2, \ldots, n\}$. We then use inequalities (22.5) and (22.6) to make estimates, and then use equation (22.4). We obtain

$$L(f,Q) = \sum_{\ell=1}^{m} m_\ell(f,Q)\Delta y_\ell = \sum_{j=1}^{n} \sum_{\ell=k_{j-1}+1}^{k_j} m_\ell(f,Q)\Delta y_\ell \geq \sum_{j=1}^{n} \sum_{\ell=k_{j-1}+1}^{k_j} m_j(f,P)\Delta y_\ell$$

$$= \sum_{j=1}^{n} m_j(f,P) \sum_{\ell=k_{j-1}+1}^{k_j} \Delta y_\ell = \sum_{j=1}^{n} m_j(f,P)\Delta x_j = L(f,P),$$

and

$$U(f,Q) = \sum_{\ell=1}^{m} M_\ell(f,Q)\Delta y_\ell = \sum_{j=1}^{n} \sum_{\ell=k_{j-1}+1}^{k_j} M_\ell(f,Q)\Delta y_\ell \leq \sum_{j=1}^{n} \sum_{\ell=k_{j-1}+1}^{k_j} M_j(f,P)\Delta y_\ell$$

$$= \sum_{j=1}^{n} M_j(f,P) \sum_{\ell=k_{j-1}+1}^{k_j} \Delta y_\ell = \sum_{j=1}^{n} M_j(f,P)\Delta x_j = U(f,P).$$

These estimates prove the first and last inequalities in inequality (22.3); the remaining inequality is inequality (22.2) applied to Q. ■

Corollary 22.0.5 *Let $a, b \in \mathbb{R}$ with $a < b$. Let P and Q be partitions of $[a,b]$. Suppose $f : [a,b] \to \mathbb{R}$ is a bounded function. Then*

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

PROOF. Consider the partition $P \cup Q$ of $[a,b]$ described above, obtained by taking the union of the sets P and Q and then listing the elements in order. Then $P \cup Q$ is a refinement of both P and Q (since $P \subseteq P \cup Q$ and $Q \subseteq P \cup Q$). Applying Lemma 22.0.4 with Q replaced by the refinement $P \cup Q$, and using inequality (22.2) applied to $P \cup Q$, we obtain

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

■

We call the partition $P \cup Q$ the *common refinement* of P and Q.

Corollary 22.0.5 states that any lower sum is less than or equal to any upper sum. This result may seem surprising, but it is natural if we consider the problem of finding the area under the graph of a function f on $[a,b]$ discussed above. Any lower sum is the area of a region inside the region under the graph of f on $[a,b]$, and so is less than A, whereas any upper sum is the area of a region containing the region under the graph of f on $[a,b]$ and hence is greater than A. That is, we expect $L(f,P) \leq A \leq U(f,Q)$.

Definition 22.0.6 *Let $a, b \in \mathbb{R}$ with $a < b$. Suppose $f : [a,b] \to \mathbb{R}$ is a bounded function. Define*

$$L(f) = \sup\{L(f,P) : P \text{ is a partition of } [a,b]\},$$

and

$$U(f) = \inf\{U(f,P) : P \text{ is a partition of } [a,b]\}.$$

We call $L(f)$ the lower Riemann integral of f, and $U(f)$ is called the upper Riemann integral of f.

A few remarks about the last definition are in order. For any partitions P and Q of $[a,b]$, we have $L(f,P) \le U(f,Q)$, by Corollary 22.0.5. Then for any partition Q, the quantity $U(f,Q)$ is an upper bound for the set $\{L(f,P) : P \text{ is a partition of } [a,b]\}$. Since that set is bounded above, it has a supremum, so $L(f)$ is defined. Similarly, for any partition P, $L(f,P)$ is a lower bound for the set $\{U(f,Q) : Q \text{ is a partition of } [a,b]\}$, and hence that set has an infimum, called $U(f)$. Also, starting from the inequality $L(f,P) \le U(f,Q)$, we can take the supremum over all P on the left sice to obtain

$$L(f) \le U(f,Q).$$

(That is, $U(f,Q)$ is an upper bound for the set $\{L(f,P) : P \text{ is a partition of } [a,b]\}$, and hence it is larger than or equal to the least upper bound $L(f)$.) Then one can take the infimum over all Q on the right side to obtain

$$L(f) \le U(f). \tag{22.7}$$

(In more detail, $L(f)$ is a lower bound for $\{U(f,Q) : Q \text{ is a partition of } [a,b]\}$, so $L(f)$ is less than or equal to the greatest lower bound $U(f)$ of that set.)

We are finally in a position to define the Riemann integral.

Definition 22.0.7 *Let $a, b \in \mathbb{R}$ with $a < b$. Suppose $f : [a,b] \to \mathbb{R}$ is a bounded function. We say f is Riemann integrable on $[a,b]$ if $L(f) = U(f)$. In that case, the common value is the Riemann integral $\int_a^b f(x)\,dx$ of f. The set of all Riemann integrable functions on $[a,b]$ is denoted $\mathcal{R}([a,b])$.*

Just for emphasis, we restate that if $f \in \mathcal{R}([a,b])$, then

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

We remark that the variable "x" in $\int_a^b f(x)\,dx$ is a dummy variable, representing the quantity being integrated over. Just as the index j in a sum $\sum_{j=1}^n a_j$ is can be replaced by another index, so that $\sum_{j=1}^n a_j = \sum_{k=1}^n a_k$, we can replace x in the integral by another letter, so that $\int_a^b f(x)\,dx = \int_a^b f(t)\,dt$.

Note that the upper and lower sums $U(f)$ and $L(f)$ are only defined for bounded functions $f : [a,b] \to \mathbb{R}$ and hence only bounded functions have a chance to be Riemann integrable. One might think that any bounded function $f : [a,b] \to \mathbb{R}$ will satisfy $L(f) = U(f)$ and hence be Riemann integrable, but that is not the case, as the following example shows.

Example 22.0.8 *Define $f : [0,1] \to \mathbb{R}$ by*

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q}\ . \end{cases}$$

(This function is the restriction to $[0,1]$ of the Dirichlet function from Example 18.0.4.) Then $f \notin \mathcal{R}([0,1])$.

Proof. Let $P = \{x_0, x_1, x_2, \ldots, x_n\}$ be a partition of $[0,1]$. Since $x_{j-1} < x_j$ for each $j \in \{1,2,\ldots,n\}$, there exist a rational number $r \in [x_{j-1}, x_j]$ (by the density of $\mathbb{Q}$ in $\mathbb{R}$, Theorem 11.0.4), so $M_j(f,P) = \sup\{f(x) : x \in [x_{j-1}, x_j]\} = 1$ (we have $M_j(f,P) \ge 1$ since $f(r) = 1$, and $f(x) \le 1$ for all x, so $M_j(f, P = 1)$). Similarly, since the irrational numbers are dense in $\mathbb{R}$ (Theorem 11.0.7), we have $m_j(f,P) = 0$. Therefore

$$U(f,P) = \sum_{j=1}^n M_j(f)\Delta x_j = \sum_{j=1}^n 1\Delta x_j = 1,$$

using equation (22.1), since the interval $[0,1]$ has length 1. However,

$$L(f,P) = \sum_{j=1}^n m_j(f)\Delta x_j = \sum_{j=1}^n 0 \cdot \Delta x_j = 0.$$

Hence $L(f,P) = 0$ for all partitions P of $[0,1]$, hence the supremum of all such $L(f,P)$ is still 0; i.e., $L(f) = 0$. Similarly $U(f) = 1$. Since $U(f) \ne L(f)$, we see that f is not Riemann integrable on $[0,1]$. ∎

The Dirichlet function is a somewhat wild function. We hope that "reasonable" functions are Riemann integrable. In fact, we will show that functions that are continuous on $[a,b]$ are Riemann integrable (also, some functions that are not continuous on $[a,b]$ are also Riemann integrable). To prove that a continuous function is Riemann integrable, we first need the following criterion for Riemann integrability.

Lemma 22.0.9 *Let $a,b \in \mathbb{R}$ with $a<b$. Suppose $f:[a,b]\to\mathbb{R}$ is a bounded function. Then $f\in\mathcal{R}([a,b])$ if and only if: for all $\epsilon>0$, there exists a partition P of $[a,b]$ such that $U(f,P)-L(f,P)<\epsilon$.*

Proof. First suppose $f\in\mathcal{R}([a,b])$, so that $U(f)=L(f)$. Let $\epsilon>0$. Since $L(f)=\sup\{L(f,Q): Q \text{ is a partition of } [a,b]\}$, then there exists a partition Q of $[a,b]$ such that $L(f,Q)>L(f)-\frac{\epsilon}{2}$. Also, since $U(f)=\inf\{U(f,R): R \text{ is a partition of } [a,b]\}$, there exists a partition R of $[a,b]$ such that $U(f,R)<U(f)+\frac{\epsilon}{2}$. Let $P=Q\cup R$ be the common refinement of Q and R. Then, using Lemma 22.0.4 and the fact that $L(f)=U(f)$,

$$U(f,P)\le U(f,R)<U(f)+\frac{\epsilon}{2}=L(f)+\frac{\epsilon}{2}<L(f,Q)+\frac{\epsilon}{2}+\frac{\epsilon}{2}=L(f,Q)+\epsilon\le L(f,P)+\epsilon.$$

Hence $U(f,p)-L(f,P)<\epsilon$. This conclusion establishes one direction of the result.

For the converse direction, temporarily fix some $\epsilon>0$. By assumption, there exists a partition P of $[a,b]$ such that $U(f,P)-L(f,P)<\epsilon$. Then by the definitions of $L(f)$ and $U(f)$,

$$U(f)\le U(f,P)<L(f,P)+\epsilon\le L(f)+\epsilon.$$

Since the inequality $U(f)<L(f)+\epsilon$ holds for all $\epsilon>0$, it follows that

$$U(f)\le L(f)$$

(if, to the contrary, $U(f)>L(f)$, then for $\epsilon=\frac{U(f)-L(f)}{2}>0$, we would have $U(f)-L(f)=2\epsilon>\epsilon$, contradicting $U(f)<L(f)+\epsilon$). But $L(f)\le U(f)$ is always true (equation (22.7)), so $L(f)=U(f)$. That is, f is Riemann integrable on $[a,b]$. ■

The next result, that continuous functions on $[a,b]$ are Riemann integrable, shows that the theory of Riemann integration applies successfully to a large class of functions.

Theorem 22.0.10 *Let $a,b\in\mathbb{R}$ with $a<b$. Suppose $f:[a,b]\to\mathbb{R}$ is continuous on $[a,b]$. Then $f\in\mathcal{R}([a,b])$.*

Proof. We first note that since the domain $[a,b]$ of f is closed and bounded, it is compact (the Heine-Borel Theorem, Theorem 17.0.19), and since f is continuous on $[a,b]$, it follows that f is bounded (Theorem 19.0.13). We will verify the condition of f in the Lemma 22.0.9. Let $\epsilon>0$. Since f is continuous on the compact set $[a,b]$, f is uniformly continuous on $[a,b]$ (Theorem 20.0.9). Hence there exists $\delta>0$ such that $|f(x)-f(y)|<\frac{\epsilon}{b-a}$ for all $x,y\in[a,b]$ such that $|x-y|<\delta$. Let P be a partition of $[a,b]$ such that

$$\|P\|=\max\{\Delta x_1,\Delta x_2,\ldots,\Delta x_n\}<\delta.$$

(For example, if we choose $n>\frac{b-a}{\delta}$, then the uniform partition $\{x_0,x_1,\ldots,x_n\}$ defined by $x_j=a+\frac{b-a}{n}\cdot j$ satisfies $\|P\|=\frac{b-a}{n}<\delta$.) Then

$$U(f,P)-L(f,P)=\sum_{j=1}^{n}M_j(f)\Delta x_j-\sum_{j=1}^{n}m_j(f)\Delta x_j=\sum_{j=1}^{n}(M_j(f)-m_j(f))\Delta x_j.$$

For each $j\in\{1,2,\ldots,n\}$, the interval $[x_{j-1},x_j]$ is a compact set, and hence the continuous function f attains its maximum and minimum values on the interval. That is, there exist $a_j,b_j\in[x_{j-1},x_j]$ such that $f(a_j)=m_j(f)$ and $f(b_j)=M_j(f)$. Therefore

$$M_j(f)-m_j(f)=f(b_j)-f(a_j)<\frac{\epsilon}{b-a},$$

since $a_j, b_j \in [x_{j-1}, x_j]$ implies that $|b_j - a_j| \leq \Delta x_j \leq \|P\| < \delta$. Hence

$$U(f,P) - L(f,P) = \sum_{j=1}^{n} (M_j(f) - m_j(f))\Delta x_j < \sum_{j=1}^{n} \frac{\epsilon}{b-a}\Delta x_j$$

$$= \frac{\epsilon}{b-a}\sum_{j=1}^{n}\Delta x_j = \frac{\epsilon}{b-a}\cdot(b-a) = \epsilon,$$

using equation (22.1). Thus, by Lemma 22.0.9, f is Riemann integrable on $[a,b]$. ■

Theorem 22.0.10 is reassuring, but there is one big problem at this point with Riemann integration. Given a function f, to show that $\int_a^b f(x)\,dx$ exists using the definition, one must compute $L(f,P)$ and $U(f,P)$ for finer and finer partitions, and show that the supremum of the $L(f,P)$ agrees with the infimum of $U(f,P)$. Then to evaluate $\int_a^b f(x)\,dx$, one must find the supremum of $L(f,P)$. These upper and lower Riemann sums can only be computed explicitly for very simple functions f and regular partitions P. Here is an example where the upper and lower sums can be computed in closed form.

Example 22.0.11 *Define $f : [0,1] \to \mathbb{R}$ by $f(x) = x$. For each $n \in \mathbb{N}$, let $P_n = \{x_0, x_1, \ldots, x_n\}$ be the partition of $[0,1]$ where $x_j = \frac{j}{n}$, for each $j = 0, 1, \ldots, n$. Caculate $L(f,P_n)$ and $U(f,P_n)$ as explicit expressions depending only on n. Then evaluate $L(f)$ and $U(f)$. Deduce that f is Riemann integrable on $[0,1]$ and evaluate $\int_0^1 f(x)\,dx$.*

Solution: We use the formla:

$$\sum_{j=1}^{n} j = \frac{n(n+1)}{2}. \tag{22.8}$$

from Example 5.0.2.

For each $j = 1, 2, \ldots, n$, we have $\Delta x_j = x_j - x_{j-1} = \frac{j}{n} - \frac{j-1}{n} = \frac{1}{n}$. Since $f(x) = x$ is increasing, we have $m_j(f) = \inf\{f(x) : x \in [x_{j-1}, x_j]\} = f(x_{j-1}) = f\left(\frac{j-1}{n}\right) = \frac{j-1}{n}$. Similarly, $M_j(f) = \sup\{f(x) : x \in [x_{j-1}, x_j]\} = f(x_j) = f\left(\frac{j}{n}\right) = \frac{j}{n}$. Then

$$L(f,P_n) = \sum_{j=1}^{n} m_j(f)\Delta x_j = \sum_{j=1}^{n} \frac{(j-1)}{n}\cdot\frac{1}{n} = \frac{1}{n^2}\sum_{j=1}^{n}(j-1),$$

and

$$\sum_{j=1}^{n}(j-1) = 0 + 1 + \cdots + (n-1) = \sum_{j=1}^{n-1} j = \frac{(n-1)n}{2} = \frac{n^2}{2} - \frac{n}{2},$$

by the case p_{n-1} of equation (22.8). Hence $L(f,P_n) = \frac{1}{n^2}\left(\frac{n^2}{2} - \frac{n}{2}\right) = \frac{1}{2} - \frac{1}{2n}$. Also using equation (22.8),

$$U(f,P_n) = \sum_{j=1}^{n} M_j(f)\Delta x_j = \sum_{j=1}^{n} \frac{j}{n}\cdot\frac{1}{n} = \frac{1}{n^2}\sum_{j=1}^{n} j = \frac{1}{n^2}\frac{n(n+1)}{2} = \frac{1}{n^2}\left(\frac{n^2}{2} + \frac{n}{2}\right) = \frac{1}{2} + \frac{1}{2n}.$$

Since $L(f)$ is the supremum of $L(f,P)$ over all partitions P, we have $L(f,P_n) \leq L(f)$ for each n. Hence $\lim_{n\to\infty} L(f,P_n) \leq L(f)$. Hence by part (b), $\frac{1}{2} = \lim_{n\to\infty}\left(\frac{1}{2} - \frac{1}{2n}\right) \leq L(f)$. Similarly, since $U(f)$ is the infimum of $U(f,P)$ over all partitions P, we have $U(f) \leq \lim_{n\to\infty} U(f,P_n) = \lim_{n\to\infty}\left(\frac{1}{2} + \frac{1}{2n}\right) = \frac{1}{2}$. But $L(f) \leq U(f)$, so we have $\frac{1}{2} \leq L(f) \leq U(f) \leq \frac{1}{2}$. We deduce that $L(f) = \frac{1}{2} = U(f)$. Since $L(f) = U(f)$, we have that f is Riemann integrable on $[0,1]$ and $\int_0^1 x\,dx = U(f) = \frac{1}{2}$.

The region under the graph of $f(x) = x$ and above the x-axis for $0 \le x \le 1$ has the shape of a isosceles right triangle with sides of length 1. So all of the above computation just affirms that the area of such a triangle is $\frac{1}{2}$, so this computation may not seem like an impressive achievement. However, the same reasoning can be applied to the function $f(x) = x^2$, for example. If we know the formula $\displaystyle\sum_{j=1}^{n} j^2 = \frac{2n^3 + 3n^2 + n}{6}$ for $n \in \mathbb{N}$, which can be proved by induction, then the same process as above (which we leave as an exercise), leads to the conclusion that the area of the region under the graph of $f(x) = x^2$ and above the x-axis for $0 \le x \le 1$ is $\frac{1}{3}$, which does not follow from the standard elementary area formulas.

As noted above, explicit evaluation of the integral using only the definition is rarely possible, and is long and difficult when possible. Fortunately, the next theorem states that there is an easy way to compute a large number of elementary integrals, using a surprising connection between integration and differentiation. We will see that in some sense these operations are inverses of each other. The significance of the following theorem in mathematics and in the development of modern technology and civilization is hard to overemphasize.

Theorem 22.0.12 *(Fundamental Theorem of Calculus, Part I) Let $a, b \in \mathbb{R}$ with $a < b$. Suppose $f \in \mathcal{R}([a,b])$, and there exists a function $F : [a,b] \to \mathbb{R}$ such that F is continuous on $[a,b]$, F is differentiable on (a,b), and $F'(x) = f(x)$ for all $x \in (a,b)$. Then*

$$\int_a^b f(x)\,dx = F(b) - F(a). \tag{22.9}$$

Proof. Let $P = \{x_0, x_1, \ldots, x_n\}$ be any partition of $[a,b]$. The for each $j \in \{1, 2, \ldots, n\}$, F is continuous on $[x_{j-1}, x_j]$ and differentiable on (x_{j-1}, x_j). Hence by the Mean Value Theorem (Theorem 21.0.12), there exists $c_j \in (x_{j-1}, x_j)$ such that

$$F(x_j) - F(x_{j-1}) = F'(c_j)(x_j - x_{j-1}) = f(c_j)\Delta x_j. \tag{22.10}$$

Since $m_j(f) = \inf\{f(x) : x \in [x_{j-1}, x_j]\}$ and $M_j(f) = \sup\{f(x) : x \in [x_{j-1}, x_j]\}$, we have $m_j \le f(c_j) \le M_j$. Hence

$$L(f,P) = \sum_{j=1}^{n} m_j(f)\Delta x_j \le \sum_{j=1}^{n} f(c_j)\Delta x_j \le \sum_{j=1}^{n} M_j(f)\Delta x_j = U(f,P). \tag{22.11}$$

By equation (22.10), we can compute

$$\sum_{j=1}^{n} f(c_j)\Delta x_j = \sum_{j=1}^{n} (F(x_j) - F(x_{j-1}))$$

$$= F(x_n) - F(x_{n-1}) + F(x_{n-1}) - F(x_{n-2}) + \ldots + F(x_2) - F(x_1) + F(x_1) - F(x_0)$$

$$= F(x_n) - F(x_0) = F(b) - F(a),$$

since the sum, which we have written in opposite order, is a telescoping sum, so that all but the first and last terms cancel out. Substituting this result into equation (22.11) gives

$$L(f,P) \le F(b) - F(a) \le U(f,P),$$

which holds for all partitions P. Taking the supremum over all partitions on the left inequality and the infimum over all partitions on the right inequality, we obtain

$$L(f) \le F(b) - F(a) \le U(f).$$

Since $f \in \mathcal{R}([a,b])$, we have $L(f) = U(f)$, hence

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

■

The Fundamental Theorem of Calculus, Part I says that to compute $\int_a^b f(x)\,dx$, we just have to find an "antiderivative" F of f, and compute $F(b) - F(a)$. For many elementary functions, the antiderivative can be guessed by reversing derivative formulas. Then, for example, to find the the area A of the region under the graph of $f(x) = x^2$ and above the x-axis, for $0 \le x \le 1$, which we did in Example 22.0.11 using the definition, one can just note that the antiderivative of x^2 is $\frac{x^3}{3}$, and compute

$$A = \int_0^1 x^2\,dx = \left.\frac{x^3}{3}\right|_0^1 = \frac{1}{3} - \frac{0}{3} = \frac{1}{3},$$

where we have used the calculus notation $F|_a^b = F(b) - F(a)$.

For future reference, we make the observation that if f is a constant on $[a, b]$, say $f(x) = c$, then $\int_a^b f(x)\,dx = c(b-a)$. This simple fact can be obtained directly from the definition of the Riemann integral, but also follows because the function $f = c$ has antiderivative $F(x) = cx$, so that

$$\int_a^b c\,dx = (cx)|_a^b = cb - ca = c(b-a). \tag{22.12}$$

These examples raise the question of whether every reasonable function f has an antiderivative F. This question will be answered in the affirmative by the Fundamental Theorem of Calculus Part II below (Theorem 22.0.16).

At first glance, the geometric problems behind differentiation (finding the slope of the tangent line) and integration (finding the area under a curve) do not seem related, so the relation between them which is revealed by the Fundamental Theorem of Calculus, Part I, is profound. However, if we think of the derivative of a function as its instantaneous rate of change, the Fundamental Theorem of Calculus Part I has an intuitive explanation, as follows. Since $F'(x) = f(x)$ is the instantaneous rate of change of F at x, and if that rate of change doesn't vary too wildly, then over a small interval $[x_{j-1}, x_j]$, we can approximate $F'(x)$ by a constant value, say $F'(c_j) = f(c_j)$, for some $c_j \in [x_{j-1}, x_j]$. Then $f(c_j)\Delta x_j$ is approximately the rate of change of F over $[x_{j-1}, x_j]$ multiplied by the duration $x_j - x_{j-1} = \Delta x_j$ of the change, which would give the change $F(x_j) - F(x_{j-1})$ of F over the interval $[x_{j-1}, x_j]$. Then the sum $\sum_{j=1}^n f(c_j)\Delta x_j$ is the sum of these approximations to the changes over the small intervals, which is then an approximation to the total change $F(b) - F(a)$ of F over $[a, b]$, i.e.,

$$\sum_{j=1}^n f(c_j)\Delta x_j \approx F(b) - F(a). \tag{22.13}$$

As we take finer and finer approximations, the approximations converge to this total change $F(b) - F(a)$. Howeveer, we have $m_j(f) \approx f(c_j) \approx M_j(f)$, with more accuracy as the partition becomes finer and finer, so $\sum_{j=1}^n f(c_j)\Delta x_j \approx \sum_{j=1}^n m_j\Delta x_j \approx \sum_{j=1}^n M_j\Delta x_j$ converges to $\int_a^b f(x)\,dx$. Hence taking the limit on the left side of equation (22.13) we obtain that $\int_a^b f(x)\,dx$ is $F(b) - F(a)$, the net change of F over $[a, b]$. Thus computing $\int_a^b f(x)\,dx$ amounts to adding up all of the small changes of F on tiny intervals in $[a, b]$ to get the net change of F on $[a, b]$. Because of this interpretation, integration has a myriad of applications to real world problems that involve changing quantities (such as a density that changes continuously with position, coordinates of an object moving over time, force that changes with elevation, etc.).

The convergence of $\sum_{j=1}^n f(c_j)\Delta x_j$ to $\int_a^b f(x)\,dx$ as the partition becomes finer and finer explains the notation $\int_a^b f(x)\,dx$: in the integral, as compared to the sum, we replace the limits $1 \le j \le n$ by the limits a to b of the interval of integration, we replace the discrete sum $\sum$ by the smooth or continuous sum $\int$, we repace the sample values $f(c_j)$ by the general value $f(x)$, and we replace the interval length Δx_j by the "infinitesimal" length dx since Δx_j goes to 0 in the limit.

If we write the Fundamental Theorem of Calculus Part I in the form $\int_a^b F'(x)\,dx = F(b) - F(a)$, it states that the integral of the derivative F' comes back to F, albeit evaluated in the form $F(b) - F(a)$. Thus the integral cancels out the derivative, returning to the original function F. In that sense, integration is a left inverse of differentiation. In the Fundamental Theorem of Calculus, Part II, we will see that it is also a right inverse, if some sense.

We first establish some basic, useful properties of the Riemann integral. The next result describes the linearity of the Riemann integral. It shows that the class of Riemann integrable functions on an interval is a vector space. The proof does not use the fundamental theorem of calculus, and could have been done earlier in the presentation.

Proposition 22.0.13 *(Linearity of the Integral) Let $a, b \in \mathbb{R}$ with $a < b$. Suppose $f, g \in \mathcal{R}([a, b])$ and $c \in \mathbb{R}$. Then $cf \in \mathcal{R}([a, b])$, $f + g \in \mathcal{R}([a, b])$, and we have*

(a) $\int_a^b (cf)(x)\,dx = c\int_a^b f(x)\,dx$ (scalar homogeneity)
and
(b) $\int_a^b (f + g)(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$ (additivity).

Proof. For (a), first suppose $c \geq 0$. Let $P = \{x_0, x_1, \ldots, x_n\}$ be a partition of $[a, b]$. By the properties of suprema and infima, and the fact that $(cf)(x) = cf(x)$,

$$m_j(cf) = \inf\{(cf)(x) : x \in [x_{j-1}, j]\} = c\inf\{f(x) : x \in [x_{j-1}, j]\} = cm_j(f),$$

and similarly $M_j(cf) = cM_j(f)$, for each $j \in \{1, 2, \ldots, n\}$. Hence

$$L(cf, P) = \sum_{j=1}^{m} m_j(cf)\Delta x_j = c\sum_{j=1}^{m} m_j(f)\Delta x_j = cL(f, P)$$

and similarly $U(cf, P) = cU(f, P)$. Since this equality holds for all partitions P of $[a, b]$, and using the properties of suprema and infima again,,

$$L(cf) = \sup\{L(cf, P) : P \text{ is a partition of } [a, b]\} = \sup\{cL(f, P) : P \text{ is a partition of } [a, b]\}$$

$$= c\sup\{L(f, P) : P \text{ is a partition of } [a, b]\} = cL(f),$$

and similarly $U(cf) = cU(f)$. Since $f \in \mathcal{R}([a, b])$, we have $L(f) = U(f)$, hence

$$L(cf) = cL(f) = cU(f) = U(cf).$$

Hence $cf \in \mathcal{R}([a, b])$, and

$$\int_a^b cf(x)\,dx = L(cf) = cL(f) = c\int_a^b f(x)\,dx.$$

Now suppose $c < 0$. Then $c = -|c|$. Using the relations $\inf(-A) = -\sup A$ and $\sup(-A) = -\inf A$ for $A \subseteq \mathbb{R}$, where $-A = \{-a : a \in A\}$, we have

$$m_j(cf) = \inf\{(-|c|f)(x) : x \in [x_{j-1}, j]\}$$

$$= -\sup\{(|c|f)(x) : x \in [x_{j-1}, j]\} = -M_j(|c|f) = -|c|M_j(f) = cM_j(f),$$

and similarly $M_j(cf) = cm_j(f)$. Hence

$$L(cf, P) = \sum_{j=1}^{m} m_j(cf)\Delta x_j = \sum_{j=1}^{m} cM_j(f)\Delta x_j = c\sum_{j=1}^{m} M_j(f)\Delta x_j = cU(f, P),$$

and similarly $U(cf, P) = cL(f, P)$. Therefore

$$L(cf) = \sup\{L(cf, P) : P \text{ is a partition of } [a, b]\}$$

$$= \sup\{cU(f, P) : P \text{ is a partition of } [a, b]\}$$

$$= \sup\{-|c|U(f, P) : P \text{ is a partition of } [a, b]\}$$

$$= -\inf\{|c|U(f, P) : P \text{ is a partition of } [a, b]\}$$

$$= -|c| \inf\{U(f,P) : P \text{ is a partition of } [a,b]\} = cU(f),$$

and similarly $U(cf) = cL(f)$. Since $f \in \mathcal{R}([a,b])$, we have $L(f) = U(f)$, so

$$L(cf) = cU(f) = cL(f) = U(cf).$$

Therefore $cf \in \mathcal{R}([a,b])$, and $\int_a^b cf(x)\,dx = L(cf) = cU(f) = c\int_a^b f(x)\,dx$.

We leave the proof of (b) as an exercise. ■

The next result is the familiar fact from calculus that, for $a < c < b$, we can break the integral over $[a,b]$ up into the sum of the integrals over $[a,c]$ and $[c,b]$.

Proposition 22.0.14 *Let $a,b,c \in \mathbb{R}$ with $a < c < b$. Suppose $f \in \mathcal{R}([a,b])$. Then $f \in \mathcal{R}([a,c])$, $f \in \mathcal{R}([c,b])$, and*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx. \tag{22.14}$$

Proof. Let $\epsilon > 0$. Since $f \in \mathcal{R}([a,b])$, there exists a partition Q of $[a,b]$ such that $U(f,Q) - L(f,Q) < \epsilon$ (by Lemma 22.0.9). Let P be the partition obtained by adding the point c to the partition Q (if $c \in Q$, just let $P = Q$). (Alternatively, let R be the partition $R = \{a,c,b\}$ of $[a,b]$, and let $P = Q \cup R$ be the common refinement of Q and R.) Then $P = \{x_0, x_1, \ldots, x_k = c, \ldots, x_n\}$, where, as we have indicated, $c = x_k$ for some $k \in 1, 2, \ldots, n-1$. Then $P_1 = \{x_0, x_1, \ldots, x_k\}$ is a partition of $[a,c]$ and $P_2 = \{x_k, x_{k+1}, \ldots, x_n\}$ is a partition of $[c,b]$. Since P is a refinement of Q, Lemma 22.0.4 gives $L(f,Q) \le L(f,P) \le U(f,P) \le U(f,Q)$, and hence

$$U(f,P) - L(f,P) \le U(f,Q) - L(f,Q) < \epsilon.$$

Then

$$L(f,P) = \sum_{j=1}^{n} m_j(f)\Delta x_j = \sum_{j=1}^{k} m_j(f)\Delta x_j + \sum_{j=k+1}^{n} m_j(f)\Delta x_j = L(f,P_1) + L(f,P_2), \tag{22.15}$$

and similarly $U(f,P) = U(f,P_1) + U(f,P_2)$. Therefore

$$U(f,P_1) - L(f,P_1) + U(f,P_2) - L(f,P_2) = U(f,P) - L(f,P) < \epsilon. \tag{22.16}$$

Since $U(f,P_1) - L(f,P_1) \ge 0$ and $U(f,P_2) - L(f,P_2) \ge 0$ (by equation (22.2)), it follows that $U(f,P_1) - L(f,P_1) < \epsilon$ and $U(f,P_2) - L(f,P_2) < \epsilon$. Thus for arbitray $\epsilon > 0$ we have found a partition P_1 of $[a,c]$ such that $U(f,P_1) - L(f,P_1) < \epsilon$, so $f \in \mathcal{R}([a,c])$ by Lemma 22.0.9. Similarly, $f \in \mathcal{R}([c,b])$. Thus $\int_a^c f(x)\,dx$ and $\int_c^b f(x)\,dx$ exist.

For a given $\epsilon > 0$, let P, P_1, and P_2 be as in the first part of this proof. Since $\int_a^c f(x)\,dx = \inf\{U(f,S) : S \text{ is a partition of } [a,c]\}$ (since $f \in \mathcal{R}([a,c])$), we have $\int_a^c f(x)\,dx \le U(f,P_1)$, and similarly $\int_c^b f(x)\,dx \le U(f,P_2)$. By equations (22.16) and (22.15),

$$\int_a^c f(x)\,dx + \int_c^b f(x)\,dx \le U(f,P_1) + U(f,P_2)$$

$$< L(f,P_1) + L(f,P_2) + \epsilon = L(f,P) + \epsilon \le \int_a^b f(x)\,dx + \epsilon,$$

where the last inequality holds because $\int_a^b f(x)\,dx = \sup\{L(f,S) : S \text{ is a partition of } [a,b]\}$ (since $f \in \mathcal{R}([a,b])$). Since this equation holds for all $\epsilon > 0$, we obtain

$$\int_a^c f(x)\,dx + \int_c^b f(x)\,dx \le \int_a^b f(x)\,dx \tag{22.17}$$

(since, if $\int_a^c f(x)\,dx + \int_c^b f(x)\,dx > \int_a^b f(x)\,dx$, the inequality above fails for $0 < \epsilon < \int_a^c f(x)\,dx + \int_c^b f(x)\,dx - \int_a^b f(x)\,dx$).

Similarly, $\int_a^c f(x)\,dx \geq L(f,P_1)$ and $\int_c^b f(x)\,dx \geq L(f,P_2)$. Then

$$\int_a^c f(x)\,dx + \int_c^b f(x)\,dx \geq L(f,P_1) + L(f,P_2)$$

$$> U(f,P_1) + U(f,P_2) - \epsilon = U(f,P) - \epsilon \geq \int_a^b f(x)\,dx - \epsilon.$$

Since this equation holds for all $\epsilon > 0$, we get

$$\int_a^c f(x)\,dx + \int_c^b f(x)\,dx \geq \int_a^b f(x)\,dx.$$

This fact and inequality (22.17) imply that $\int_a^c f(x)\,dx + \int_c^b f(x)\,dx = \int_a^b f(x)\,dx$. ■

In calculus it is traditional to define $\int_a^a f(x)\,dx = 0$ for any $a \in \mathbb{R}$, and $\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$ when $a < b$. Then equation (22.14) holds for all $a, b, c \in \mathbb{R}$.

Next we require a couple of inequalities about integrals. The first is simple, and the second is one of the most commonly used inequalities in analysis.

Proposition 22.0.15 *Let $a, b \in \mathbb{R}$ with $a < b$.*

(a) Suppose $f, g \in \mathcal{R}([a,b])$ and $f(x) \leq g(x)$ for all $x \in [a,b]$. Then $\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx$.

(b) Suppose $f \in \mathcal{R}([a,b])$. Then $|f| \in \mathcal{R}([a,b])$ and

$$\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx. \tag{22.18}$$

Proof. One can give a direct proof of (a), but it is easier at this point to note that $-f \in \mathcal{R}([a,b])$ and hence $h = g - f \in \mathcal{R}([a,b])$ (using Proposition 22.0.13 with $c = -1$). By assumption, $h(x) = g(x) - f(x) \geq 0$ for all $x \in [a,b]$. Hence for any partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a,b]$, we have $m_j(h) = \inf\{h(x) : x \in [x_{j-1}, x_j]\} \geq 0$ since 0 is a lower bound for the set and the infimum is the greatest lower bound. Hence $L(f,P) = \sum_{j=1}^n m_j(h)\Delta x_j \geq 0$. Therefore $\int_a^b h(x)\,dx = L(h) \geq L(f,P) \geq 0$ (the equality is because $h \in \mathcal{R}([a,b])$ and the inequality is because $L(h)$ is the supremum of $L(h,P)$ over all partitions P). By Proposition 22.0.13,

$$\int_a^b g(x)\,dx - \int_a^b f(x)\,dx = \int_a^b g(x)\,dx + \int_a^b -f(x)\,dx = \int_a^b g(x) - f(x)\,dx = \int_a^b h(x)\,dx \geq 0,$$

hence $\int_a^b g(x)\,dx \geq \int_a^b f(x)\,dx$.

To prove (b), let $P = \{x_0, x_1, \ldots, x_n\}$ be any partition of $[a,b]$. Suppose $j \in \{1, 2, \ldots, n\}$ and $x, y \in [x_{j-1}, x_j]$. If $f(x) \leq f(y)$, then $m_j(f) \leq f(x) \leq f(y) \leq M_j(f)$, since $m_j(f)$ and $M_j(f)$ are ths infimum and supremum, respectively, of f on $[x_{j-1}, x_j]$. Hence $f(y) - f(x) \leq M_j(f) - m_j(f)$. If $f(y) \leq f(x)$, interchanging x and y gives $f(x) - f(y) \leq M_j(f) - m_j(f)$. Since $|f(y) - f(x)| = f(y) - f(x)$ if $f(x) \leq f(y)$ and $|f(y) - f(x)| = f(x) - f(y)$ if $f(y) \leq f(x)$, we conclude that

$$|f(y) - f(x)| \leq M_j(f) - m_j(f), \text{ for all } x, y \in [x_{j-1}, x_j].$$

Then by the triangle inquality,

$$|f(y)| \leq |f(y) - f(x)| + |f(x)| \leq M_j(f) - m_j(f) + |f(x)|, \text{ for all } x, y \in [x_{j-1}, x_j].$$

For each $x \in [x_{j-1}, x_j]$, we take the supremum over all $y \in [x_{j-1}, x_j]$ on the left side to obtain

$$M_j(|f|) = \sup\{|f(y)| : y \in [x_{j-1}, x_j]\} \leq M_j(f) - m_j(f) + |f(x)|, \text{ for all } x \in [x_{j-1}, x_j].$$

We rewrite this inequality as

$$M_j(|f|) - M_j(f) + m_j(f) \leq |f(x)|, \text{ for all } x \in [x_{j-1}, x_j].$$

Since the left side of this inequality is independent of x, we can take the infimum of the right side over all $x \in [x_{j-1}, x_j]$ to obtain

$$M_j(|f|) - M_j(f) + m_j(f) \leq m_j(|f|).$$

Reorganizing terms gives

$$M_j(|f|) - m_j(|f|) \leq M_j(f) - m_j(f), \tag{22.19}$$

for each $j \in \{1, 2, \ldots, n\}$.

Let $\epsilon > 0$. We will show $|f| \in \mathcal{R}([a,b])$ by showing that $|f|$ satisfies the criterion in Lemma 22.0.9. Since $f \in \mathcal{R}([a,b])$, by Lemma 22.0.9 there exists a partition $P = \{x_0, x_1, \dot{,}x_n\}$ such that $U(f,P) - L(f,P) < \epsilon$. Then by equation (22.19)

$$U(|f|,P) - L(|f|,P) = \sum_{j=1}^{n} M_j(|f|)\Delta x_j - \sum_{j=1}^{n} m_j(|f|)\Delta x_j = \sum_{j=1}^{n} \left(M_j(|f|) - m_j(|f|)\right)\Delta x_j$$

$$\leq \sum_{j=1}^{n} \left(M_j(f) - m_j(f)\right)\Delta x_j = \sum_{j=1}^{n} M_j(f)\Delta x_j - \sum_{j=1}^{n} m_j(f)\Delta x_j = U(f,P) - L(f,P) < \epsilon.$$

So by Lemma 22.0.9, $|f| \in \mathcal{R}([a,b])$.

To show inequality 22.18, note that $f(x) \leq |f(x)| = |f|(x)$ for all $x \in [a,b]$, hence by part (a),

$$\int_a^b f(x)\,dx \leq \int_a^b |f|(x)\,dx = \int_a^b |f(x)|\,dx.$$

Also, $-f(x) \leq |f(x)| = |f|(x)$ for all $x \in [a,b]$, so

$$-\int_a^b f(x)\,dx = \int_a^b -f(x)\,dx \leq \int_a^b |f|(x)\,dx = \int_a^b |f(x)|\,dx.$$

Since $\left|\int_a^b f(x)\,dx\right|$ is either $\int_a^b f(x)\,dx$ or $-\int_a^b f(x)\,dx$, the two preceding inequalities yield inequality (22.18). ■

If $f \in \mathcal{R}([a,b])$, then for each $x \in (a,b)$, we have $f \in \mathcal{R}([a,x])$, so $\int_a^x f(t)\,dt$ exists. The idea behind the Fundamental Theorem of Calculus, Part II, is to regard $\int_a^x f(t)\,dt$ as a function of x and investigate its differentiability properties.

Theorem 22.0.16 *(Fundamental Theorem of Calculus, Part II) Let $a, b \in \mathbb{R}$ with $a < b$, and suppose $f \in \mathcal{R}([a,b])$. For $x \in [a,b]$, let*

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

If $x_0 \in (a,b)$ and f is continuous at x_0, then F is differentiable at x_0 and $F'(x_0) = f(x_0)$.

Proof. Let $\epsilon > 0$. Since f is continuous at x_0, there exists $\delta > 0$ such that $|f(t) - f(x_0)| < \frac{\epsilon}{2}$ for all $t \in [a,b]$ such that $|t - x_0| < \delta$.

Suppose first that $x \in (x_0, x_0 + \delta) \cap [a,b]$. Then by Proposition 22.0.14,

$$F(x) - F(x_0) = \int_a^x f(t)\,dt - \int_a^{x_0} f(t)\,dt = \int_{x_0}^x f(t)\,dt.$$

Applying equation (22.12) to the constant $f(x_0)$, we have

$$\frac{1}{x - x_0}\int_{x_0}^x f(x_0)\,dt = \frac{1}{x - x_0} \cdot f(x_0)(x - x_0) = f(x_0).$$

Hence

$$\frac{F(x)-F(x_0)}{x-x_0} - f(x_0) = \frac{1}{x-x_0}\int_{x_0}^{x} f(t)\,dt - f(x_0)$$

$$= \frac{1}{x-x_0}\int_{x_0}^{x} f(t)\,dt - \frac{1}{x-x_0}\int_{x_0}^{x} f(x_0)\,dt = \frac{1}{x-x_0}\int_{x_0}^{x} (f(t)-f(x_0))\,dt,$$

using Proposition 22.0.13. Therefore, using Proposition 22.0.15 part (b),

$$\left|\frac{F(x)-F(x_0)}{x-x_0} - f(x_0)\right| = \left|\frac{1}{x-x_0}\int_{x_0}^{x} (f(t)-f(x_0))\,dt\right| \le \frac{1}{x-x_0}\int_{x_0}^{x} |f(t)-f(x_0)|\,dt.$$

But for $x_0 \le t \le x$, we have $|t-x_0| \le |x-x_0| < \delta$, hence $|f(t)-f(x_0)| < \frac{\epsilon}{2}$. So applying Proposition 22.0.15 part (a) with the constant function $\frac{\epsilon}{2}$, we obtain

$$\left|\frac{F(x)-F(x_0)}{x-x_0} - f(x_0)\right| \le \frac{1}{x-x_0}\int_{x_0}^{x} |f(t)-f(x_0)|\,dt$$

$$\le \frac{1}{x-x_0}\int_{x_0}^{x} \frac{\epsilon}{2}\,dt = \frac{1}{x-x_0}\cdot\frac{\epsilon}{2}(x-x_0) = \frac{\epsilon}{2} < \epsilon,$$

for all $x \in (x_0, x_0+\delta) \cap [a,b]$.

Now suppose $x \in (x_0-\delta, x_0) \cap [a,b]$, so $x_0 > x$. The argument is almost the same as the previous case, because

$$\frac{F(x)-F(x_0)}{x-x_0} = \frac{F(x_0)-F(x)}{x_0-x} = \frac{1}{x_0-x}\int_{x}^{x_0} f(t)\,dt.$$

So by similar steps as above,

$$\left|\frac{F(x)-F(x_0)}{x-x_0} - f(x_0)\right| = \left|\frac{1}{x_0-x}\int_{x}^{x_0} (f(t)-f(x_0))\,dt\right| \le \frac{1}{x_0-x}\int_{x}^{x_0} |f(t)-f(x_0)|\,dt$$

$$\le \frac{1}{x_0-x}\cdot\frac{\epsilon}{2}(x_0-x) = \frac{\epsilon}{2} < \epsilon.$$

Thus for all $x \in (x_0-\delta, x_0+\delta)\setminus\{x_0\}$, we have $\left|\frac{F(x)-F(x_0)}{x-x_0} - f(x_0)\right| < \epsilon$. Since $\epsilon > 0$ was arbitrary, by the definition of limit, we have that $\lim_{x\to x_0} \frac{F(x)-F(x_0)}{x-x_0}$ exists and equals $f(x_0)$. That is, F is differentiable at x_0 and $F'(x_0) = f(x_0)$. ■

If f is a continuous function on $[a,b]$, then the last result states that $F(x) = \int_a^x f(t)\,dt$ is differentiable and $F'(x) = f(x)$. Thus f has an antiderivative F, which resolves the question arising after the Fundamental Theorem of Calculus Part I as to whether a sufficiently nice function has an antiderivative. We may not be able to express the antiderivative simply in terms of the standard functions of calculus, but at least there does exist an antiderivative function.

If we use the Leibniz notation $\frac{d}{dx}$ to denote the derivative, i.e., $\frac{d}{dx}f = f'(x)$, then the Fundamental Theorem of Calculus Part II says that

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x),$$

which gives a sense in which applying the integral and then the derivative gives back the original function. Thus the integral is a right inverse of the derivative in this sense. This observation complements the Fundamental Theorem of Calculus Part I which we interpreted above as saying that the integral is a left inverse of the derivative.

The Fundamental Theorem of Calculus Part II can be used to give another proof, at least under the assumption that f is continuous on $[a,b]$, of the Fundamental Theorem of Calculus, Part I, which states that $\int_a^b f(t)\,dt = F(b)-F(a)$ if $F' = f$ on (a,b). To see this proof, define $G : [a,b] \to \mathbb{R}$ by

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

By the Fundamental Theorem of Calculus part II, $G'(x) = f(x) = F'(x)$ for all $x \in [a, b]$. By the linearity of the derivative, we have that $(F - G)'(x) = 0$ for all $x \in [a, b]$. By Corollary 21.0.14, it follows that $(F - G)(x) = C$, where C is some constant, for $x \in [a, b]$. That means that

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

If we substitute the value $x = a$ in this equation, we obtain $F(a) = C + \int_a^a f(t)\,dt = C + 0 = C$. Substituting $C = F(a)$, we obtain

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

Finally, letting $x = b$ and subtracting $F(a)$ from both sides of the equation gives

$$\int_a^b f(t)\,dt = F(b) - F(a).$$

Chapter 23

Sequences of Functions and Uniform Convergence

So far we have considered sequences of real numbers and their convergence, with problems like Example 1.0.1 in mind, where the value of $\sqrt{C}$ is found as the limit of a sequence of approximations. In many mathematical problems, the goal is to find not a number but a certain function, for example a function having particular properties (e.g., a continuous, nowhere differentiable function), or the solution of a differential equation. Often the solution is found as as a limit of an approximating sequence of functions. Example 1.0.2 suggests how this procedure can be used to find a solution of an ordinary differential equation. To make such a process precise, we must first define what we mean by the limit of a sequence of functions. It turns out that there are many possible definitions, and one may be useful in one problem and another in a different problem. Here we will discuss two notions: *pointwise* convergence and *uniform* convergence.

Definition 23.0.1 *Let $A \subseteq \mathbb{R}$ be a non-empty set. A sequence of functions (f_n) on A is a list, indexed by $\mathbb{N}$, such that for each $n \in \mathbb{N}$, $f_n : A \to \mathbb{R}$ is a function. For a function $f : A \to \mathbb{R}$, we say that f_n converges to f pointwise on A if the sequence of real numbers $(f_n(x))$ converges to $f(x)$, for all $x \in A$.*

When the set A is understood, we just say that f_n converges to f pointwise. Although pointwise convergence is the simplest and perhaps most natural sense of convergence of a sequence of functions, pointwise convergence is inadequate for many purposes. The reason is that we often hope to construct a function f with certain properties, such as continuity, differentiability, or Riemann integrability, as a limit of functions with these same properties. Often we don't know f explicitly, so we cannot check these properties directly for f. Instead, we hope that f inherits the needed property if the functions f_n have this property. Unfortunately, it is often the case that such properties do not pass to the limit under pointwise convergence. Here are some examples.

Example 23.0.2 *(Failure of continuity under pointwise convergence) For $n \in \mathbb{N}$, define $f_n : [0,1] \to \mathbb{R}$ by $f_n(x) = x^n$. Then each function f_n is continuous on $[0,1]$. If $0 \leq x < 1$, then*

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} x^n = 0.$$

However, for $x = 1$, we have $f_n(x) = f_n(1) = 1^n = 1$ for all n, so $\lim_{n\to\infty} f_n(1) = 1$. Hence, for $f : [0,1] \to \mathbb{R}$ defined by $f_n(x) = 0$ for $0 \leq x < 1$ and $f(1) = 1$, we have that f is the pointwise limit of (f_n) on $[0,1]$. However, f is not continuous, even though all of the f_n are continuous. That is, continuity is not necessarily inherited under pointwise convergence.

Example 23.0.3 *(The derivative of the limit may not equal the limit of the derivative) In Example 23.0.2, the functions x^n are all differentiable on $[0,1]$, but their pointwise limit is not continuous, hence not differentiable, at the point 1 (if one considers one-sided derivatives at the endpoints). However, one might hope that if the pointwise limit f is differentiable, then the sequence of derivatives f_n' would converge and that*

$f' = (\lim f_n)'$ would coincide with $\lim(f_n')$. However, that hope is not correct. For $n \in \mathbb{N}$, define $f_n\mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \frac{\sin(nx)}{\sqrt{n}}.$$

Then $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in \mathbb{R}$, so the limit function $f = 0$ is differentiable with $f'(x) = 0$ for all $x \in \mathbb{R}$. However,

$$f_n'(x) = \frac{n\cos(nx)}{\sqrt{n}} = \sqrt{n}\cos(nx)$$

is a divergent sequence, for all $x \in \mathbb{R}$; i.e., $\lim_{n\to\infty} f_n'(x)$ does not exist.

Example 23.0.4 *(The pointwise limit of Riemann integrable functions may not be Riemann integrable) Recall (see Proposition 12.0.13) that $\mathbb{Q}$, the set of rational numbers, is countably infinite, and hence so is $\mathbb{Q} \cap [0,1]$. So we can write*

$$\mathbb{Q} \cap [0,1] = \{r_1, r_2, \ldots, r_n, \ldots\}.$$

Define $f_n \to \mathbb{R}$ by $f_n(x) = 1$ for $x \in \{r_1, r_2, \ldots, r_n\}$ and $f_n(x) = 0$ for all other $x \in [0,1]$. Then each f_n has only finitely many discontinuities, and (it turns out - check) is Riemann integrable on $[0,1]$. If $x \in \mathbb{Q} \cap [0,1]$, then $x = r_j$ for some $j \in \mathbb{N}$, and hence $f_n(r_j) = 1$ for all $n \geq j$. Hence $\lim_{n\to\infty} f_n(x) = 1$ for all $x \in \mathbb{Q} \cap [0,1]$. However, if $x \in [0,1] \setminus \mathbb{Q}$, then $f_n(x) = 0$ for all $n \in \mathbb{N}$, and hence $\lim_{n\to\infty} f_n(x) = 0$. Thus the pointwise limit f of the sequence f_n is the Dirichlet function from Example 22.0.8. In that example, we showed that f is not Riemann integrable on $[0,1]$. So the Riemann integability of each sequence element does not guarantee the Riemann integrability of the pointwise limit.

Example 23.0.5 *(The integral of the limit may not equal the limit of the integrals) In the previous example, we say that even though each f_n may be Riemann integrable on an interval $[a,b]$, the pointwise limit f may not be Riemann integrable on that interval. However, if f is Riemann integrable on $[a,b]$, one might hope that the integral of f is the limit of the integrals of the f_n, i.e., that $\int_a^b f(x)\,dx = \int_a^b \lim_{n\to\infty} f_n(x)\,dx$ agrees with $\lim_{n\to\infty} \int_a^b f_n(x)\,dx$. However, that conclusion is not valid in general. For example, for $n \in \mathbb{N}$, define $f_n : [0,1] \to \mathbb{R}$ by $f_n(x) = n$ if $0 < x < \frac{1}{n}$, and $f_n(x) = 0$ for all other $x \in [0,1]$. Then $f_n \in \mathcal{R}([0,1])$ and*

$$\int_0^1 f_n(x)\,dx = \int_0^{1/n} n\,dx = \frac{1}{n} \cdot n = 1,$$

for all $n \in \mathbb{N}$, so $\lim_{n\to\infty} \int_0^1 f_n(x)\,dx = 1$. However, $f_n(0) = 0$ for all $n \in \mathbb{N}$, so $f(0) = \lim_{n\to\infty} f_n(0) = 0$, and, for all $x \in (0,1]$, we have $f_n(x) = 0$ for all $n > \frac{1}{x}$ (since then $x > \frac{1}{n}$), so $f(x) = \lim_{n\to\infty} f_n(x) = 0$. Therefore the pointwise limit f is just the function that is everywhere 0, so $\int_0^1 f(x)\,dx = 0$. So $\int_0^1 \lim_{n\to\infty} f_n(x)\,dx = 0 \neq 1 = \lim_{n\to\infty} \int_0^1 f_n(x)\,dx$.

Example 23.0.6 *($\lim f_n(x_n) \neq f(x)$) Suppose $A \subseteq \mathbb{R}$ is non-empty, (f_n) is a sequence of real-valued continuous functions on A which converges pointwise to $f : A \to \mathbb{R}$. Suppose (x_n) is a sequence of points of A, converging to a point $x \in A$. Is it necessarily true that $(f_n(x_n))$ converges and $\lim_{n\to\infty} f_n(x_n) = f(x)$? It seems reasonable that this statement should be true: x_n is close to x and $f_n(x)$ gets close to $f(x)$ as $n \to \infty$, for every $x \in A$. However, suppose $f_n : [0,1] \to \mathbb{R}$ is defined by $f_n(x) = x^n$, as in Example 23.0.2, where we saw that the pointwise limit of the sequence (f_n) is the function f which is 0 on $[0,1)$ and 1 at $x = 1$. Let $x_n = 1 - \frac{1}{n}$, for $n \in \mathbb{N}$. Then $x = \lim_{n\to\infty} x_n = 1$. Hence $f(1) = 1$. However, borrowing the result from calculus (which we have not proved yet, but we accept for now) that $\lim_{n\to\infty} \left(1 + \frac{t}{n}\right)^n = e^t$, with $t = -1$, we have*

$$\lim_{n\to\infty} f_n(x_n) = \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n = e^{-1} \neq 1 = f(1) = f(\lim_{n\to\infty} x_n).$$

So $\lim_{n\to\infty} f_n(x_n) \neq f(x)$.

These examples should convince us that pointwise convergence is not sufficient for many purposes in analysis. Fortunately, there is a stronger notion of convergence for a sequence of functions. We will see that this notion is sufficient for many purposes.

Definition 23.0.7 *Suppose $A \subseteq \mathbb{R}$ is non-empty, $f_n : A \to \mathbb{R}$ is a function for each $n \in \mathbb{N}$, and $f : A \to \mathbb{R}$ is a function. Then f_n converges to f uniformly on A if, for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \epsilon$ for all $n \in \mathbb{N}$ and all $x \in A$.*

This definition may appear to be the same as the definition of the convergence of $f_n(x)$ to $f(x)$ at each $x \in A$, i.e., pointwise convergence, just with the definition of convergence written out in terms of ϵ and N, but that is not the case. The difference is subtle but critical. In pointwise convergence, for each x and ϵ, there must be an associated $N \in \mathbb{N}$, which means that N can depend on both ϵ and x. In uniform convergence, after $\epsilon > 0$ is given, there must be an N which works for that ϵ and for all $x \in A$ simultaneously. That is, in uniform convergence, N must be independent on x, unlike the case of pointwise convergence.

It should be clear that uniform convergence implies pointwise convergence: if f_n converges to f uniformly on A, then f_n converges pointwise to f on A. The converse is not true in general, as the example from Example 23.0.2 shows.

Example 23.0.8 *Define $f_n : [0,1] \to \mathbb{R}$ by $f_n(x) = x^n$. In Example 23.0.2 we saw that f_n converges pointwise to the function $f : [0,1] \to \mathbb{R}$ defined by $f(x) = 0$ for $0 \leq x < 1$ and $f(1) = 1$. However, f_n does not converge uniformly to f. To see this fact, consider $\epsilon = \frac{1}{2}$. If f_n were to converge to f uniformly, then there would exist $N \in N$ such that for all $n > N$, we would have $|f_n(x) - f(x)| < \frac{1}{2}$ for all $x \in [0,1]$. In particular, then, for all $x \in [0,1)$, we have $f(x) = 0$, so we would have $x^{n_0} = |x^{n_0} - 0| < \frac{1}{2}$ for all $x \in [0,1)$, for some $n_0 \in \mathbb{N}$. However, the continuity of x^{n_0} at $x = 1$ guarantees that $\lim_{x\to 1} x^{n_0} = 1$, contradicting the statement that $x^{n_0} < \frac{1}{2}$ for all $x \in [0,1)$.*

Under unifrom convergence, the phenomena in Examples 23.0.2 - 23.0.6 do not occur. We start with continuity.

Theorem 23.0.9 *Suppose $A \subseteq \mathbb{R}$ and $x_0 \in A$. Suppose (f_n) is a sequence of real-valued functions on A which converge uniformly on A to a function $f : A \to \mathbb{R}$. If each f_n is continuous at x_0, then f is continuous at x_0.*

Proof. Let $\epsilon > 0$. Since f_n converges uniformly to f on A, there exists $N \in \mathbb{N}$ such that

$$|f_n(x) - f(x)| < \frac{\epsilon}{3} \text{ for all } n > N \text{ and all } x \in A. \tag{23.1}$$

Select $m > N$. Since f_m is continuous at x_0, there exists $\delta > 0$ such that

$$|f_m(x) - f_m(x_0)| < \frac{\epsilon}{3} \text{ for all } x \in A \text{ such that } |x - x_0| < \delta. \tag{23.2}$$

Hence if $|x - x_o| < \delta$ and $x \in A$, then

$$|f(x) - f(x_0)| = |f(x) - f_m(x) + f_m(x) - f_m(x_0) + f_m(x_0) - f(x_0)|$$

$$\leq |f(x) - f_m(x)| + |f_m(x) - f_m(x_0)| + |f_m(x_0) - f(x_0)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon,$$

where the estimates $|f(x) - f_m(x)| < \frac{\epsilon}{3}$ and $|f_m(x_0) - f(x_0)| < \frac{\epsilon}{3}$ hold by (23.1), and the estimate $|f_m(x) - f_m(x_0)| < \frac{\epsilon}{3}$ holds by (23.2). Hence f is continuous at x_0. ■

By Example 23.0.2, this result fails if we only assume that f_n converges pointwise to f on A. It is worth understanding where the proof of Theorem 23.0.9 breaks down in that case. One can still make the triangle inequality estimate, and one can still conclude that $|f_m(x_0) - f(x_0)| < \frac{\epsilon}{3}$ for m sufficiently large, by the pointwise convergence of f_n to f at the point x_0, and one can still obtain the estimate $|f_m(x) - f_m(x_0)| < \frac{\epsilon}{3}$ for $|x - x_0|$ sufficiently small, from the continuity of f_m at x_0. The problem comes from the first term $|f(x) - f_m(x)|$. One needs this term to be small, for a single m, for all x satisfying $|x - x_0| < \delta$. For a single x, one can find such an m, but without uniform convergence, one cannot guarantee the existence of an m that gives the estimate for all x satisfying $|x - x_0| < \delta$.

Corollary 23.0.10 *Suppose $A \subseteq \mathbb{R}$, and (f_n) is a sequence of real-valued functions on A which converge uniformly on A to a function $f : A \to \mathbb{R}$. If each f_n is continuous on A, then f is continuous on A.*

PROOF. The limit f is continuous at each $x_0 \in A$, by Theorem 23.0.9. ■

This corollary can be stated simply as: "The uniform limit of continuous functions is continuous." This is the positive result, under the assumption of uniform convergence, corresponding to Example 23.0.2. Regarding Example 23.0.3, the uniform convergence of a sequence of differentiable functions does not guarantee that the limit is differentiable, but under additional assumptions there is a positive result, as follows.

Theorem 23.0.11 *Let $a, b \in \mathbb{R}$ with $a < b$. Suppose that for each $n \in \mathbb{N}$, $f_n : (a, b) \to \mathbb{R}$ is a function such that f_n is differentiable on (a, b) and f_n' is continuous on (a, b). Suppose that there exists $f : (a, b) \to \mathbb{R}$ and $g : (a, b) \to \mathbb{R}$ such that:*

(i) f_n converges pointwise to f on (a, b)
and
(ii) f_n' converges uniformly to g on (a, b).

Then f is differentiable on (a, b) and $f' = g$ on (a, b).

We leave the proof as an exercise, with the hint to apply the fundamental theorem of calculus to each f_n.

Riemann integration behaves well under uniform convergence (compare to Examples 23.0.4 and 23.0.5). The proof of the following result is left as an exercise.

Theorem 23.0.12 *Let $a, b \in \mathbb{R}$ with $a < b$. Suppose that for each $n \in \mathbb{N}$, $f_n : [a, b] \to \mathbb{R}$ is a function such that $f_n \in \mathcal{R}([a, b])$. Suppose $f : [a, b] \to \mathbb{R}$ is a function and f_n converges uniformly to f on $[a, b]$. Then $f \in \mathcal{R}([a, b])$ and*

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

Finally, the following positive analogue of Example 23.0.6 holds.

Theorem 23.0.13 *Suppose $A \subseteq \mathbb{R}$ is non-empty. Suppose (f_n) is a sequence of real-valued functions on A that converges uniformly on A to a function $f : A \to \mathbb{R}$. Suppose (x_n) is a sequence of points of A that converges to some point $x \in A$. Then the sequence $(f_n(x_n))$ converges to $f(x)$.*

The proof of the last result is also left as an exercise.

The main point of this section is that uniform convergence is stronger and more useful than pointwise convergence, because the limit function often inherits good properties from the functions in the sequence, if the convergence is uniform. We mention, however, that there are many other notions of convergence of a sequence of functions that are useful. For example, in studying Fourier series it is important to consider "L^2-convergence." We say that a sequence of Riemann integrable functions (f_n) on an interval $[a, b]$ converge "in L^2" to a Riemann integrable function f on $[a, b]$ if

$$\lim_{n\to\infty} \int_a^b |f_n(x) - f(x)|^2\,dx = 0.$$

However, this topic is better understood in the more advanced context of Lebesgue integration.