Math 4100 Notes, Fall 2023

Pete L. Clark

Contents

A Little Review

The real numbers

The name of the course is **Real Analysis**, so let us begin with a check-in on the real numbers $\mathbb{R}$, perhaps the most important single mathematical object. Intuitively we view $\mathbb{R}$ as being the points on a number line, with the origin marked as 0 and with an orientation so that we may distinguish positive from negative. We may *represent* every real number via an infinite decimal expansion, and while this is certainly an excellent way to think about and work with real numbers, it works poorly as a definition.

The modern approach is to lean on a certain collection of axioms for $\mathbb{R}$:

I. The field axioms: $\mathbb{R}$ is endowed with two binary operations $+$ and $\cdot$, satisfying many familiar properties like commutativity, associativity and so forth.

II. The order axioms: $\mathbb{R}$ is endowed with a total order relation $\leq$.

III. The ordered field axioms, which give compatibility between the field operations and the order structure:
(OF1) For all $x \in \mathbb{R}$, exactly one of the following holds: $x = 0$, $x > 0$, $-x > 0$.
(OF2) For all $x, y, z \in \mathbb{R}$, if $x \leq y$ then $x + z \leq y + z$.
(OF3) For all $x, y \in \mathbb{R}$, if $x \geq 0$ and $y \geq 0$, then $x \cdot y \geq 0$.

A structure that satisfies all of the properties so far is called an **ordered field**. There are in fact an enormous number of ordered fields: the rational numbers, $\mathbb{Q}$, is one. A **subfield** of $\mathbb{R}$ is a subset $F \subseteq \mathbb{R}$ that contains 0 and 1 and is closed under the field operations: if $x, y \in F$ then $x + y, x - y, x \cdot y \in F$ and $\frac{x}{y} \in F$ if $y \neq 0$. If we take any subfield of $\mathbb{R}$ and restrict the relation $\leq$ to F, then we get an ordered field. This builds *a lot* (infinitely many, to say the least!) of subfields of $\mathbb{R}$. And there are also more exotic ordered fields that do not arise in this way, although you will probably never meet any in an undergraduate course (including this one).

The real numbers is characterized among ordered fields by satisfying:

IV. The completeness axiom, which can be stated in several equivalent forms.

But first let me nail down what "characterized" means: first of all, the real numbers satisfy these completeness axioms. Second, an ordered field F that satisfies one of these completeness axioms is "essentially" the real numbers, which means that one

can find a bijective function $f : F \to \mathbb{R}$ that preserves all of the structures: $+$, $\cdot$ and $\leq$. (More formally, f is an **isomorphism of ordered fields**.)

Now back to the completeness axioms. The most useful formulation is:

Dedekind's Completeness: If X is a subset of $\mathbb{R}$ that is nonempty and bounded above, then X has a least upper bound, or supremum, in $\mathbb{R}$.

Another version is **O'Connor's Completeness / Monotone Sequence Lemma**: Every bounded monotone sequence $\{x_n\}_{n=1}^{\infty}$ in $\mathbb{R}$ converges to a real number.

What do I mean by "versions" of completeness? I mean that it can be shown that if an ordered field satisfies Dedekind's Completeness than it also satisfies O'Connor's Completeness (this was an important result in Math 3100) and also conversely an ordered field that satisfies O'Connor's Completeness also satisfies Dedekind's Completeness (this was probably *not* covered in class in Math 3100 but see [**SS**]).

In any ordered field satisfying the completeness axioms, the following properties also hold:

Archimedean Property: For every real number x, there is a positive integer n with $n > x$.

Cauchy's Completeness: Every Cauchy sequence in $\mathbb{R}$ converges.

The second of these implications is a major result from a previous course: [**SS**, Thm. 2.6.11]. We will freely use it – in particular to give a generalization to Cauchy sequences in Euclidean space – in our course. The first of these implications may be less familiar. We leave it as Exercise 0.1.

The rational numbers $\mathbb{Q}$ are an example of an ordered field that do not satisfy Cauchy's completeness. It turns out that an ordered field satisfies the Archimedean property if and only if it is (isomorphic to) a subfield of $\mathbb{R}$, so non-Archimedean ordered fields are precisely the ones we called "exotic" above. Moreover:

PROPOSITION 0.1. *An ordered field that satisfies the Achimedean Property* and *Cauchy's completeness is Dedekind complete – and thus isomorphic to* $\mathbb{R}$.

PROOF. See [**SS**, Proposition 2.6.7 and Theorem 2.6.13b)]. □

Just a remark / reminder: one *does* need to show that there is a Dedekind complete ordered field that is unique up to isomorphism; that is, we still need to "construct the real numbers $\mathbb{R}$." At least, someone does. The first such construction was given by Dedekind in the late 1800's. The truth of it is that no such construction is particularly simple, so that one needs a certain amount of mathematical sophistication to understand it...at which point it seems to be a better use of any instructor's time to cover something else. So it is extremely rare to encounter the construction $\mathbb{R}$ in a course. This course will be no exception. But if by chance you do want to see a construction of $\mathbb{R}$, it is written up in [**HC**, Chapter 16].

Exercises.

Exercise 0.1. *Let $(F, +, \leq)$ be an ordered field that is Dedekind complete. Show that F is Archimedean.*
(Suggestion: proceeding by contrapositive, if F is not *Archimedean then there is $x \in F$ such that $n \leq x$ for every positive integer n, so the set $\mathbb{Z}^+$ of positive integers has an upper bound in F. Show: $\mathbb{Z}^+$ has no* least *upper bound in F.)*

Exercise 0.2. *Let $F \subsetneq \mathbb{R}$ be a proper subfield of $\mathbb{R}$. We make F into an ordered field by restricting the usual $\leq$ on $\mathbb{R}$ to F. Note that we must have $\mathbb{Q} \subseteq F$.*

a) *Show: F is an Archimedean ordered field.*
(In fact, every subfield of an Archimedean ordered field is an Archimedean ordered field: that is one way to go.)
b) *Show: F is* not *Dedekind complete.*
(Suggestion: Use the fact that every real number is the limit of a sequence of rational numbers.)

Problems.

Problem 0.1. *Let $(F, +, \leq)$ be an ordered field. Show that the following are equivalent:*

(i) *The field F is Archimedean.*
(ii) *Every sequence $\{x_n\}_{n=1}^{\infty}$ in F that is increasing and bounded is a Cauchy sequence in F.*

Problem 0.2. *Show: there is an ordered field that is Cauchy complete but not Dedekind complete: equivalently by Proposition 0.1, there is a non-Archimedean ordered field in which every Cauchy sequence is convergent.*

The extended real numbers

We define the **extended real numbers** $[-\infty, +\infty]$ to be the set of real numbers together with two additional symbols $+\infty$ and $-\infty$. We extend the ordering on $\mathbb{R}$ to the extended real numbers by putting $-\infty$ smaller than every other element and $+\infty$ larger than every other element.

Why do we do this? Here is the point: $\mathbb{R}$ with its usual ordering $\leq$ is **Dedekind complete**: every subset X of $\mathbb{R}$ that is nonempty and bounded above has a least upper bound. In this definition both of the words "nonempty" and "bounded above" are necessary: if X were unbounded above then...well, it does not have an upper bound in the real numbers: that is, for all $M \in \mathbb{R}$ there is $x \in X$ with $x > M$, so it certainly does not have a least upper bound. The case of $X = \varnothing$ is similar but somehow a bit more confusing. Here the problem is that *every* real number M if an upper bound for the empty set: since there are no elements in $\varnothing$, indeed every real number is at least as large as every element of $\varnothing$. So the least upper bound of $\varnothing$ would be the least real number...but there is no such thing, since for all $M \in \mathbb{R}$ we have $M - 1 < M$.

Each of these difficulties is remedied in the ordered set of extended real numbers $[-\infty, +\infty]$.

Proposition 0.2. *Let X be a subset of the extended real numbers $[-\infty, +\infty]$. Then X has a least upper bound in $[-\infty, +\infty]$.*

Proof.
Case 1: Suppose $X = \varnothing$. Then every element of $[-\infty, +\infty]$ is an upper bound for X, so the least upper bound of X is the least element of $[-\infty, +\infty]$, which is $-\infty$.
Case 2: Suppose $X = \{-\infty\}$. Then again every element of $[-\infty, \infty]$ is an upper bound for X, so the least upper bound is $-\infty$.
Case 3: Suppose $+\infty \in X$. Then $+\infty$ is the only upper bound of X, so it is the least upper bound.
Case 4: Suppose $+\infty$ is not an element of X, but $X \cap \mathbb{R}$ is unbounded above. Then no real number is an upper bound for X, so $+\infty$ is the only upper bound for X, hence the least upper bound.
Case 5: Suppose $X \supsetneq \{-\infty\}$ and X is bounded above. Then $X \cap \mathbb{R}$ has a least upper bound, which is a real number M. Since the only possible element of $X \setminus (X \cap \mathbb{R})$ is $-\infty$, which is certainly less than M, also M is an upper bound for X, so is the least upper bound. □

In a very similar way it can be shown that every subset of $[-\infty, +\infty]$ has a greatest lower bound, or infimum: this is left as an exercise. Thus if we work in the extended real numbers, suprema and infima always exist, which is very convenient. The terminology for this is that the set $[-\infty, \infty]$ is **complete** for its ordering $\leq$, whereas $\mathbb{R}$ was merely Dedekind complete.

However, note well that it is certainly *not* the case that $[-\infty, \infty]$ form a field in a way that is compatible with the ordering we have defined. Indeed, no ordered field has a largest or smallest element: for any element x in an ordered field, $x + 1$ is larger and $x - 1$ is smaller. It is natural to define *some* operations on extended real numbers, but some of the others cannot be usefully made for reasons relating to indeterminate forms. Namely, we will make the following definitions:

$$\forall x \in \mathbb{R}, x + (+\infty) = +\infty, \ x + (-\infty) = -\infty.$$

$$(+\infty) \cdot (+\infty) = +\infty, \ (+\infty) \cdot (-\infty) = -\infty, \ (-\infty) \cdot (-\infty) = +\infty.$$

$$\forall x \in (0, \infty), \ x \cdot (+\infty) = +\infty, \ x \cdot (-\infty) = -\infty.$$

$$\forall x \in (-\infty, 0), \ x \cdot (+\infty) = -\infty, \ x \cdot (-\infty) = +\infty.$$

These definitions apply to give extensions of familiar "limit laws" to the extended real numbers. For instance, if we have real sequences $\{\mathbf{x}_n\}$ and $\{\mathbf{y}_n\}$ such that $\mathbf{x}_n$ diverges to $+\infty$ and $\mathbf{y}_n$ converges to the real number M, then we have

$$\mathbf{x}_n + \mathbf{y}_n \to M + \infty = \infty.$$

The proof of this is an (ϵ, N)-argument that should be familiar from Math 3100; the point is that our convention gives the correct answer.

This should serve to explain why some of the possible field-theoretic operations with extended real numbers are *not defined*. For instance, we do not define $(+\infty)+(-\infty)$ because if $\{\mathbf{x}_n\}$ is a real sequence diverging to ∞ and $\{\mathbf{y}_n\}$ is a real sequence diverging to $-\infty$ then nothing can be said about the limiting behavior of $\mathbf{x}_n + \mathbf{y}_n$: it may converge, it may diverge to $+\infty$, it may diverge to $-\infty$, or it may do none of those things. For similar reasons we do not define $0 \cdot +\infty$, $\frac{\pm\infty}{+\infty}$ and so forth.

Exercises.

EXERCISE 0.3. *Let $X \subseteq [-\infty, +\infty]$. Show: X has a greatest lower bound m. Show that $m \in \mathbb{R}$ if and only if all of the following hold: (i) X is nonempty; (ii) $X \neq \{+\infty\}$; and (iii) X is bounded below by a real number.*

EXERCISE 0.4. *Let $X \subseteq Y \subseteq \mathbb{R}$.*

a) *Show:* $\sup X \leq \sup Y$.
b) *Show:* $\inf X \geq \inf Y$.

Problems.

PROBLEM 0.3. *Let $X \subseteq [-\infty, \infty]$. We define*

$$-X := \{-x \mid x \in X\} \subseteq [-\infty, \infty].$$

a) *Show:* $\inf X = -\sup(-X)$.
b) *Show:* $\sup X = -\inf(-X)$.

Problem 0.3 is an instance of a **reflection principle**: every statement involving suprema in $\mathbb{R}$ or $[-\infty, +\infty]$ will have a "reflected version" involving infima.

Inequalities

Whereas the ordered field $\mathbb{R}$ of real numbers is the most important mathematical object in this or any real analysis course, in real analysis we do not simply contemplate $\mathbb{R}$ "in stasis." Rather, real analysis is the study of various limiting processes and other concepts that can be defined in terms of limiting processes: sequential limits, functional limits, continuity, derivatives, integrals...This has been true (at least) since the work of Newton and Leibniz in the 17th century. Making $\mathbb{R}$ the foundational object of real analysis is a comparatively later development, achieved by work of Cauchy, Weierstrass and Dedekind throughout the 19th century. The reason that this works is that these mathematicians discovered that all these fundamental limiting processes – which had been present in mathematics for hundreds of years but with shrouded in mysterious language that even several contemporaries of Newton and Leibniz correctly pointed out was not really satisfiactory – could be rigorously defined in terms of inequalities.

Inequalities are really the main currency of real analysis, but they require some technique to work with. You learn to work with inequalities in both Math 3200 and Math 3100, and you will certainly get a chance to increase your skill in this course. Let us now recall the "first two tricks in the book" when it comes to working with inequalities:

First Trick: For $A, B \in \mathbb{R}$, we have $A \leq B \iff B - A \geq 0$.

This is clearly not a profundity: to get from the left hand side to the right hand side, flip the inequality around and add $-A$ to both sides, and to get from the right hand side to the left hand side, reverse that: i.e., flip the inequality around and add A to both sides. Nevertheless it is often much easier to show that $B - A \geq 0$. Why? Well, most often A and B are not really numbers like 3 and 7 but more abstract expressions: imagine for instance that they are functions of x: $A = A(x)$ and $B = B(x)$. Then the interpretation of the inequality $A(x) \leq B(x)$ is that the graph of the function $B(x)$ lies on or above the graph of the function $A(x)$. To

show this, we may need to understand both functions and how they relate to each other. On the other hand, the equivalent inequality $B(x) - A(x) \geq 0$ means that the graph of the one function $B(x) - A(x)$ lies on or above the y-axis. Somehow we have reduced a statement about two functions to a statement about one function.

Anyway, the First Trick is closely related to the

Second Trick: For $A \in \mathbb{R}$, we have $A \geq 0 \iff A = B^2$ for some $B \in \mathbb{R}$.

This is really saying that every non-negative real number has a real square root, which is a consequence of the Intermediate Value Theorem that you met in calculus and saw the proof of in Math 3100. However the Intermediate Value Theorem is probably not the point here, since in most applications of the Second Trick it is rather the implication $\Longleftarrow$ that we will use (and for what it's worth, this part holds in any ordered field).

To see what I mean, let's show that for all $x, y \in \mathbb{R}$ we have $2xy \leq x^2 + y^2$. By the First Trick, it is equivalent to show:

$$\forall x \in \mathbb{R}\ \forall y \in \mathbb{R},\ x^2 - 2xy + y^2 \geq 0.$$

Then the Second Trick is inviting us to write $x^2 - 2xy + y^2$ as the square of something else...not using the Intermediate Value Theorem but directly. Well, okay: indeed

$$x^2 - 2xy + y^2 = (x - y)^2 \geq 0,$$

and we're done. To get some appreciation for these two tricks, try to prove this inequality in some other way.

You might still think I'm messing around here, but I really am not. These two tricks will be used to prove the first nonobvious result in this course. Watch for it!

CHAPTER 1

Topology of Euclidean Space

1. Euclidean N-Space

Let $N \in \mathbb{Z}^+$. By $\mathbb{R}^N$ we mean the set of ordered N-tuples of real numbers

$$\mathbf{x} = (x_1, \ldots, x_N).$$

This is a familiar object from linear algebra, as a vector space over $\mathbb{R}$. This means that elements of $\mathbb{R}^N$ can be added to each other, and it also makes sense to "scale" an element $\mathbf{x}$ by a real number α:

$$\alpha(x_1, \ldots, x_N) \coloneqq (\alpha x_1, \ldots, \alpha x_n).$$

However, we are interested in $\mathbb{R}^N$ not just as a real vector space, but endowed with the **Euclidean norm**, which is a function from $\mathbb{R}^N$ to $[0, \infty)$, the non-negative real numbers. Specifically:

$$\forall\ \mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N,\ ||\mathbf{x}|| \coloneqq \sqrt{x_1^2 + \ldots + x_N^2}.$$

We recall a *very* basic fact: for elements $x_1, \ldots, x_N$ in any ordered field F, we have

$$x_1^2 + \ldots + x_N^2 \geq 0,$$

and

$$x_1^2 + \ldots + x_N^2 = 0 \iff x_1 = \ldots = x_N = 0.$$

That is: a sum of squares is never negative, and is 0 if and only if every term is 0.

From this we deduce:

$$\forall\ \mathbf{x} \in \mathbb{R}^N,\ \mathbf{x} = 0 \iff ||\mathbf{x}|| = 0.$$

Here is another easy property of the Euclidean norm:

Proposition 1.1. *For all $\mathbf{x} \in \mathbb{R}^N$ and all $\alpha \in \mathbb{R}$, we have*

$$||\alpha \mathbf{x}|| = |\alpha| ||\mathbf{x}||.$$

The proof of Proposition 1.1 is left as an exercise.

By **Euclidean N-space** I mean $\mathbb{R}^N$ equipped with its Euclidean norm. By the way, the Euclidean norm itself can be defined in terms of an **inner product** operation

$$\cdot : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R},$$

$$(x_1, \ldots, x_N) \cdot (y_1, \ldots, y_N) \coloneqq x_1 y_1 + \ldots + x_N y_N.$$

Then:

$$\forall\ \mathbf{x} \in \mathbb{R}^N,\ ||\mathbf{x}|| = \sqrt{\mathbf{x} \cdot \mathbf{x}}.$$

Inner products are *extremely* important in certain branches of analysis, but I think they will only make a brief appearance in this course.

We use the Euclidean norm to measure distance between points in $\mathbb{R}^N$, namely: for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, we define the **Euclidean distance**

$$d(\mathbf{x}, \mathbf{y}) := ||\mathbf{x} - \mathbf{y}||.$$

In mathematics, if we have a set X and a function $d : X \times X \to \mathbb{R}$, then to call d a "distance function" we usually require the following three properties:

(D1) (Positive Definiteness) For all $x, y \in X$ we have $d(x, y) \geq 0$, with equality if and only if $x = y$.
(D2) (Symmetry) For all $x, y \in X$ we have $d(x, y) = d(y, x)$.
(D3) (Triangle Inequality) For all $x, y, z \in X$ we have $d(x, z) \leq d(x, y) + d(y, z)$.

Shall we try to show that our Euclidean distance satisfies these three properties? It starts out easily:

(D1): For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, we have $d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}|| \geq 0$ because the norm of anything is at least 0, and moreover $||\mathbf{x} - \mathbf{y}|| = 0$ if and only if $\mathbf{x} - \mathbf{y} = 0$ if and only if $\mathbf{x} = \mathbf{y}$.

No problem!

(D2) For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, we have $d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}|| = ||-(\mathbf{y} - \mathbf{x})||$. Using Proposition 1.1 we have

$$||-(\mathbf{y} - \mathbf{x})|| = |-1|||\mathbf{y} - \mathbf{x}|| = ||\mathbf{y} - \mathbf{x}|| = d(\mathbf{y}, \mathbf{x}).$$

Again, no problem.

(D3) We want to show:

$$\forall\ \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^N,\ ||\mathbf{x} - \mathbf{z}|| \leq ||\mathbf{x} - \mathbf{y}|| + ||\mathbf{y} - \mathbf{z}||. \tag{1}$$

Hmm. Well, I notice that $\mathbf{x} - \mathbf{z} = (\mathbf{x} - \mathbf{y}) + (\mathbf{y} - \mathbf{z})$, so if we put

$$\mathbf{A} := \mathbf{x} - \mathbf{y},\ \mathbf{B} := \mathbf{y} - \mathbf{z},$$

then we have $\mathbf{A}, \mathbf{B} \in \mathbb{R}^N$ and we want to show

$$||\mathbf{A} + \mathbf{B}|| \leq ||\mathbf{A}|| + ||\mathbf{B}||.$$

In other words, we see that in order to show (D3) it suffices to show the slightly simpler statement:

$$\forall\ \mathbf{x}, \mathbf{y} \in \mathbb{R}^N, ||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||. \tag{2}$$

To show (2) we really need to do something, although there is more than one "something" that will work. The following approach is a good one in that brings nothing "extraneous" to bear. The main step is to establish the following closely related result.

THEOREM 1.2 (Cauchy-Schwarz in $\mathbb{R}^N$). *For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, we have*

$$|\mathbf{x} \cdot \mathbf{y}| \leq ||\mathbf{x}||||\mathbf{y}||.$$

PROOF. Write $\mathbf{x} = (x_1, \ldots, x_N)$, $\mathbf{y} = (y_1, \ldots, y_N)$. For non-negative real numbers X, Y we have $X \leq Y$ if and only if $X^2 \leq Y^2$, so it is equivalent to show

$$|\mathbf{x} \cdot \mathbf{y}|^2 \leq ||\mathbf{x}||^2 ||\mathbf{y}||^2.$$

Without any vector notation, what we want to show is:

$$(x_1 y_1 + \ldots + x_N y_N)^2 \leq (x_1^2 + \ldots + x_N^2)(y_1^2 + \ldots + y_N^2).$$

Put

$$L \coloneqq (x_1 y_1 + \ldots + x_N y_N)^2$$

and

$$R \coloneqq (x_1^2 + \ldots + x_N^2)(y_1^2 + \ldots + y_N^2),$$

so we want to show that $L \leq R$; it will certainly suffice to show $R - L \geq 0$. Now:

$$R = \sum_{i=1}^{N} x_i^2 y_i^2 + \sum_{1 \leq i \neq j \leq N} x_i^2 y_j^2 = \sum_i x_i^2 y_i^2 + \sum_{i<j} x_i^2 y_j^2 + \sum_{i<j} x_j^2 y_i^2,$$

while

$$L = \sum_{i=1}^{N} x_i^2 y_i^2 + \sum_{1 \leq i \neq j \leq N} x_i y_i x_j y_j = \sum_i x_i^2 y_i^2 + 2 \sum_{i<j} x_i y_i x_j y_j,$$

so

$$R - L = \sum_{i<j} x_i^2 y_j^2 - 2 \sum_{i<j} x_i y_j x_j y_i + \sum_{i<j} x_j^2 y_i^2 = \sum_{i<j} (x_i y_j - x_j y_i)^2 \geq 0. \qquad \square$$

Using Theorem 1.2, it easy to prove (2), especially if we allow ourselves to use simple properties of inner products from Exercise 1.2. Indeed, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$. We want to show that $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$. Again it suffices to show this after squaring both sides, so equivalently we want to show:

$$||\mathbf{x} + \mathbf{y}||^2 \leq (||\mathbf{x}|| + ||\mathbf{y}||)^2.$$

Now we have

$$||\mathbf{x} + \mathbf{y}||^2 = (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) = (\mathbf{x} \cdot \mathbf{x}) + (\mathbf{x} \cdot \mathbf{y}) + (\mathbf{y} \cdot \mathbf{x}) + (\mathbf{y} \cdot \mathbf{y})$$

$$= ||\mathbf{x}||^2 + 2(\mathbf{x} \cdot \mathbf{y}) + ||\mathbf{y}||^2 \leq ||\mathbf{x}||^2 + 2|\mathbf{x} \cdot \mathbf{y}| + ||\mathbf{y}||^2 \overset{\text{CS}}{\leq} ||\mathbf{x}||^2 + 2||\mathbf{x}||||\mathbf{y}|| + ||\mathbf{y}||^2$$

$$= (||\mathbf{x}|| + ||\mathbf{y}||)^2.$$

It is important to know when equality holds in Cauchy-Schwarz or (this is very closely related) in the Triangle Inequality.

COROLLARY 1.3. *Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$. The following are equivalent:*

(i) *The vectors $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent: that is, either $\mathbf{x} = 0$ or there is $\alpha \in \mathbb{R}$ such that $\mathbf{y} = \alpha \mathbf{x}$.*
(ii) *We have $|\mathbf{x} \cdot \mathbf{y}| = ||\mathbf{x}|| \cdot ||\mathbf{y}||$.*

Your are asked to prove Corollary 1.3 in Problem 1.1.

Exercises.

General Comment: Many exercises and problems will refer to $\mathbb{R}^N$. Here it should be understood that N is an *arbitrary* positive integer. That is, unless you are asked for an example, your solution should apply no matter what the value of N is.

Exercise 1.1. *Show: for all* $\mathbf{x} \in \mathbb{R}^N$ *and all* $\alpha \in \mathbb{R}$*, we have*

$$||\alpha \mathbf{x}|| = |\alpha| ||\mathbf{x}||.$$

Exercise 1.2. *Let* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^N$ *and* $\alpha \in \mathbb{R}$.

a) *Show:* $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$.
b) *Show:* $(\alpha \mathbf{x}) \cdot \mathbf{y} = \alpha(\mathbf{x} \cdot \mathbf{y})$.
c) *Show:* $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = (\mathbf{x} \cdot \mathbf{z}) + (\mathbf{y} \cdot \mathbf{z})$.

Exercise 1.3. *We showed that (2) implies (1). Show that conversely, (1) implies (2). Explicitly, suppose that:*

$$\forall\ \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^N,\ ||\mathbf{x} - \mathbf{z}|| \leq ||\mathbf{x} - \mathbf{y}|| + ||\mathbf{y} - \mathbf{z}||.$$

Show:

$$\forall\ \mathbf{x}, \mathbf{y} \in \mathbb{R}^N,\ ||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||.$$

Exercise 1.4 (Reverse Triangle Inequality). *Show: for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$*, we have*

$$\big| ||\mathbf{x}|| - ||\mathbf{y}|| \big| \leq ||\mathbf{x} - \mathbf{y}||.$$

Problems.

Problem 1.1. *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$*. Show that the following are equivalent:*

(i) *The vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *are linearly dependent: that is, either* $\mathbf{x} = 0$ *or there is* $\alpha \in \mathbb{R}$ *such that* $\mathbf{y} = \alpha \mathbf{x}$.
(ii) *We have* $|\mathbf{x} \cdot \mathbf{y}| = ||\mathbf{x}|| \cdot ||\mathbf{y}||$.

Problem 1.2. *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$.

a) *Suppose that* $||\mathbf{x}+\mathbf{y}|| = ||\mathbf{x}||+||\mathbf{y}||$*. Show:* $\mathbf{x}$ *and* $\mathbf{y}$ *are linearly dependent.*
b) *Find necessary and sufficient conditions for* $||\mathbf{x} + \mathbf{y}|| = ||\mathbf{x}|| + ||\mathbf{y}|$.

2. Sequences in $\mathbb{R}^N$

2.1. Sequences in a Set. Let X be any set. We have the notion of a sequence in X: informally, this is an infinite ordered list of elements of X:

$$x_1, x_2, \ldots, x_n, \ldots \quad \text{with } x_n \in X\ \forall n \in \mathbb{Z}^+.$$

This is formalized as a function $x_\bullet : \mathbb{Z}^+ \to X$; then we have $x_\bullet(n) = x_n$. For instance we could consider sequences in the set of real-or-made-up English words (such a thing consists of a finite string of letters from our alphabet, whether it is a valid English word or not), and then

$$b,\ bo,\ boo,\ booo,\ boooo \ldots \tag{3}$$

defines a sequence.

In this level of generality we can consider **subsequences**: to form a subsequence, we choose an infinite, strictly increasing sequence of positive integers

$$n_1 < n_2 < \ldots < n_k < \ldots$$

and then form the new sequence

$$x_{n_1}, x_{n_2}, \ldots, x_{n_k}, \ldots.$$

Again we can be a bit more formal: a strictly increasing sequence of positive integers corresponds to a strictly increasing function $n_\bullet : \mathbb{Z}^+ \to \mathbb{Z}^+$, and then to pass from the sequence $x_\bullet : \mathbb{Z}^+ \to X$ to the corresponding subsequence we form the composition of functions

$$x_\bullet \circ n_\bullet : k \mapsto n_k \mapsto x_{n_k}.$$

So for instance if we take $n_k = k^2$ for all k then in our above weird example we get the subsequence

$$b,\ booo, boooooooo, boooooooooooooo, \ldots. \tag{4}$$

But we're not really cooking with gas here. Rather we'd like a notion of *convergence* of sequences, and for this one needs some kind of extra structure on our set: for our weird sequence (3) above, if you asked me whether it converges I can only look at you quizzically: we just haven't set up enough for that question to be meaningful.

2.2. Sequences in $\mathbb{R}^N$. Let's motivate the definition for convergence of sequences in $\mathbb{R}^N$. First, for $N = 1$ we have seen this definition already: it is the single most important definition of Math 3100. If we have a sequence $\{x_n\}$ of real numbers, we say the sequence **converges** to a real number L if

$$\forall\ \epsilon > 0,\ \exists K \in \mathbb{Z}^+ \text{ such that } \forall n > K,\ |x_n - L| < \epsilon.$$

A sequence **converges** if it converges to some $L \in \mathbb{R}$; otherwise it **diverges**. One of the first things one shows is that if a sequence converges then its limit is *unique*.

Now let me rephrase this definition slightly. First, when $N = 1$ the Euclidean norm is precisely the absolute value, and thus $|x_n - L|$ is nothing else than the distance $d(x_n, L)$ between x_n and L...as is certainly familiar from Math 3100. Now I make the following observation:

• The sequence $\{x_n\}$ converges to $L \iff$ the sequence $\{d(x_n, L)\}$ converges to 0.

Indeed, if we write out the latter convergence statement, it is: for all $\epsilon > 0$, there is $N \in \mathbb{Z}^+$ such that for all $n > N$ we have $\big||x_n - L| - 0\big| < \epsilon$. But

$$\big||x_n - L| - 0\big| = |x_n - L|,$$

so this is the same as saying that $x_n \to L$.

Aha. So if $\{x_n\}$ is a sequence in $\mathbb{R}^N$ and $L \in \mathbb{R}^N$, we can (and do!) say that x_n converges to L – and write $x_n \to L$ – if $d(x_n, L) \to 0$. This is still an (ϵ, K) definition: spelling it out, we get that $x_n \to L$ means: for all $\epsilon > 0$, there is $K \in \mathbb{Z}^+$ such that for all $n > K$ we have $||x_n - L|| < \epsilon$.[1]

[1]So you see that I could have just said that we're replacing the absolute value $|\cdot|$ by the Euclidean norm $||\cdot||$. But I have my reasons for phrasing it in terms of the distance function, as will become clear at the end of the course.

EXAMPLE 1.1. *Consider the sequence* $\mathbf{x}_n = (\frac{1}{n^2}, \frac{n+3}{n+4})$ *in* $\mathbb{R}^2$. *We will show:*

$$\mathbf{x}_n \to (0,1).$$

From Math 3100 we know how to show that $\frac{1}{n^2} \to 0$ *and* $\frac{n+3}{n+4} \to 1$. *Let's put those together to show that* $\mathbf{x}_n \to (0,1)$. *Let* $\epsilon > 0$, *and put*

$$K := \left\lceil \frac{\sqrt{2}}{\epsilon} \right\rceil.$$

Step 1: If $n > K$ *we have*

$$\left| \frac{1}{n^2} - 0 \right| = \left| \frac{1}{n^2} \right| = \frac{1}{n^2} \leq \frac{1}{n} < \frac{1}{K} \leq \frac{1}{\lceil \frac{\sqrt{2}}{\epsilon} \rceil} < \frac{1}{\frac{\sqrt{2}}{\epsilon}} \leq \frac{\epsilon}{\sqrt{2}}.$$

Step 2: If $n > K$, *we also have*

$$\left| \frac{n+3}{n+4} - 1 \right| = \frac{1}{n+4} < \frac{1}{n} \leq \frac{\epsilon}{\sqrt{2}}.$$

Step 3:Thus, if $n > K$ *then*

$$||\mathbf{x}_n - (0,1)|| = \sqrt{(\frac{1}{n^2} - 0)^2 + (\frac{n+3}{n+4} - 1)^2} < \sqrt{\left(\frac{\epsilon}{\sqrt{2}}\right)^2 + \left(\frac{\epsilon}{\sqrt{2}}\right)^2} = \epsilon.$$

So $\mathbf{x}_n \to (0,1)$.

There is a general moral to extract here. We will get back to this shortly.

THEOREM 1.4 (Familiar Facts About Convergence). *Let* $\{\mathbf{x}_n\}$ *and* $\{\mathbf{y}_n\}$ *be sequences in* $\mathbb{R}^N$. *Suppose that* $\mathbf{x}_n \to \mathbf{L} \in \mathbb{R}^N$ *and* $\mathbf{y}_n \to \mathbf{M} \in \mathbb{R}^N$.

a) *Let* $\alpha \in \mathbb{R}$. *Then* $\alpha\mathbf{x}_n \to \alpha\mathbf{L}$.
b) *We have* $\mathbf{x}_n + \mathbf{y}_n \to \mathbf{L} + \mathbf{M}$.
c) *Every subsequence* $\{\mathbf{x}_{n_k}\}$ *of* $\mathbf{x}_n$ *also converges to* $\mathbf{L}$.
d) *If* $\mathbf{P} \in \mathbb{R}^N$ *is such that* $\mathbf{x}_n \to \mathbf{P}$, *then* $\mathbf{L} = \mathbf{P}$.

You are asked to prove each of these facts as exercises. Of course, this is mostly to get you to look back at the corresponding proofs for real sequences.

A subset $S \subseteq \mathbb{R}^N$ is **bounded** if there is $M \geq 0$ such that for all $x \in S$ we have $||x|| \leq M$. In other words, a subset is bounded if the distances of its elements from the origin are bounded above by a fixed real number. (Soon enough we will rephrase this by saying that S is contained in some closed ball centered at 0.)

We say that a sequence $\{\mathbf{x}_n\}$ in $\mathbb{R}^N$ is **bounded** if the set of terms $\{\mathbf{x}_n \mid n \in \mathbb{Z}^+\}$ is a bounded subset of $\mathbb{R}^N$. Here is one more familiar fact:

THEOREM 1.5. *Convergent sequences in* $\mathbb{R}^N$ *are bounded.*

PROOF. Suppose $\mathbf{x}_n \to L$. Then there is $K \in \mathbb{Z}^+$ such that for all $n > K$ we have $||\mathbf{x}_n - L|| \leq 1$. By the Reverse Triangle Inequality (Exercise 1.4), we get:

$$\forall n > K, \ \Big| ||\mathbf{x}_n|| - ||L|| \Big| \leq ||\mathbf{x}_n - L|| \leq 1,$$

so

$$\forall n > K, ||\mathbf{x}_n|| \leq ||L|| + 1.$$

Now put

$$M := \max(||x_1||, \ldots, ||x_K||, ||L|| + 1).$$

Then for all $n \in \mathbb{Z}^+$ we have $||\mathbf{x}_n|| \leq M$, so $\{\mathbf{x}_n\}$ is bounded. $\square$

There are some things that we did with real sequences that *do not* make sense for sequences in $\mathbb{R}^N$ for $N > 1$, namely:

• In $\mathbb{R}$ we can multiply sequences and show the analogue of Theorem 1.4 for products. In $\mathbb{R}^N$ we cannot in general multiply two vectors so as to get another vector. However, there are a few loopholes here:
(i) We can multiply vectors in $\mathbb{R}^2$. Indeed we can identify $\mathbb{R}^2$ with the complex numbers $\mathbb{C}$ and use the given multiplication.
(ii) We can multiply vectors in $\mathbb{R}^3$, using the cross product. This is a kind of weird multiplication operation (neither commutative nor associative), but nevertheless it exists and is often useful (though probably not for us in our course).
(iii) For all $N \in \mathbb{Z}^+$ we can multiply two elements of $\mathbb{R}^N$ to get an element of $\mathbb{R}$, using the scalar product.

It happens to be true that in all three cases, these products preserve convergence of sequences. The first two of these are explored in the exercises; we will prove the third a little later on.

• Whereas $\mathbb{R}$ comes equipped with an ordering, for $N > 1$ we do not have any (natural, useful) total ordering on $\mathbb{R}^N$. Thus the important notion of monotone sequence in $\mathbb{R}^N$ has no analogue in $\mathbb{R}^N$, although we could speak of monotonicity of the sequence of norms.

• In $\mathbb{R}$ we have the notion of diverging to $+\infty$ and also the notion of diverging to $-\infty$: a real sequence $\{x_n\}$ diverges to $+\infty$ if for all $M \in \mathbb{R}$ there is $N \in \mathbb{Z}^+$ such that for all $n > N$ we have $x_n > M$. Similarly, a real sequence $\{x_n\}$ diverges to $-\infty$ if for all $m \in \mathbb{R}$ there is $N \in \mathbb{Z}^+$ such that for all $n > N$ we have $x_n < m$. For $N > 1$ we have something similar but less precise. Namely, a sequence $\mathbf{x}$ in $\mathbb{R}^N$ **diverges to infinity** if the real sequence $||\mathbf{x}_N||$ diverges to ∞.

2.3. The Secret to Convergence in $\mathbb{R}^N$. Look back at Example 1.1 of a convergent sequence in $\mathbb{R}^2$:

$$\mathbf{x}_n = \left(\frac{1}{n^2}, \frac{n+3}{n+4}\right).$$

Put $x_n = \frac{1}{n^2}$ and $y_n = \frac{n+3}{n+4}$. Then the sequence of x-components converges to 0 and the sequence of y-components converges to 1; having established this it took us only one more line to show that $(x_n, y_n) \to (0, 1)$.

In fact this is a general phenomenon of convergence in $\mathbb{R}^N$! Namely, let $\{\mathbf{x}_n\}$ be a sequence in $\mathbb{R}^N$. For each $1 \leq i \leq N$, let $\mathbf{x}_{n,i}$ be the ith component of $\mathbf{x}_n$. Then the vector sequence $\{\mathbf{x}_n\}$ can be traded in for N different real sequences: $\{\mathbf{x}_{n,1}\}, \ldots, \{\mathbf{x}_{n,N}\}$. It turns out that the convergence of the vector sequence is equivalent to the convergence of all of the scalar sequences:

THEOREM 1.6. *Let $\{\mathbf{x}_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}^N$, and let $\mathbf{L} = (L_1, \ldots, L_N) \in \mathbb{R}^N$. Then the following are equivalent:*

(i) *The vector sequence $\mathbf{x}_n$ converges to $\mathbf{L}$.*

(ii) *For each* $1 \leq i \leq N$*, the real sequence* $\{\mathbf{x}_{n,i}\}$ *of* i*th components converges to* L_i*.*

PROOF. The key to this is the following relatively simple observation: let $\mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N$. There is at least one $1 \leq I \leq N$ such that

$$\forall 1 \leq i \leq N, \ |x_i| \leq |x_I|.$$

Fix such an I. Then for each $1 \leq i \leq N$ we have

$$|x_i| = \sqrt{x_i^2} \leq \sqrt{x_1^2 + \ldots + x_N^2} = ||\mathbf{x}|| \leq \sqrt{x_I^2 + \ldots + x_I^2} = \sqrt{N x_I^2} = \sqrt{N}|x_I|.$$

This shows: if $||\mathbf{x}||$ is small, then so is the absolute value of each coordinate of $\mathbf{x}$ – in fact, each is no larger than $||\mathbf{x}||$ – and conversely, if all of the absolute values of the coordinates are small, then $||\mathbf{x}||$ is also small: at most $\sqrt{N}$ times as large as the largest coordinate absolute value. These inequalities imply that for any sequence $\{\mathbf{x}_n\}$ in $\mathbb{R}^N$ we have $\mathbf{x}_n \to 0$ if and only if $\mathbf{x}_{n,i} \to 0$ for all $1 \leq i \leq N$. The general case follows from this special case applied to the sequence $\{\mathbf{x}_n - \mathbf{L}\}$. □

We extend the notion of **Cauchy sequence** to $\mathbb{R}^N$ in a straightforward way: a sequence $\{\mathbf{x}_n\}$ in $\mathbb{R}^N$ is Cauchy if for all $\epsilon > 0$, there is $K \in \mathbb{Z}^+$ such that for all $m, n \geq K$ we have $||\mathbf{x}_m - \mathbf{x}_n|| < \epsilon$. The same simple inequalities used in the proof of Theorem 1.6 also work to show:

THEOREM 1.7. *A sequence* $\{\mathbf{x}_n\}$ *in* $\mathbb{R}^N$ *is Cauchy if and only if for all* $1 \leq i \leq N$*, the real sequence* $\{\mathbf{x}_{n,i}\}$ *is Cauchy.*

We leave the details of this as an exercise. It follows that:

COROLLARY 1.8. *A sequence in* $\mathbb{R}^N$ *is convergent if and only if it is Cauchy.*

PROOF. We know the result for $N = 1$ from Math 3100. So by what we have just seen, the vector sequence is convergent if and only if each of its component scalar sequences is convergent if and only if each of its component scalar sequences is Cauchy if and only if the vector sequence is Cauchy. □

This has the same advantage of knowing the equivalence of Cauchy sequences and convergent sequences in $\mathbb{R}$: it allows us to decouple the question of convergence of a sequence from the question of knowing the limit of the sequence; often the former questions is much easier than the latter.

2.4. Bolzano-Weierstrass in $\mathbb{R}^N$. The celebrated Bolzano-Weierstrass Theorem says that every bounded real sequence has a convergent subsequence. This extends verbatim to sequences in $\mathbb{R}^N$, as we will now show. Let us first give an equivalent formulation of boundedness of subsets of Euclidean N-space. Suppose we are given real numbers $a_1 \leq b_1, a_2 \leq b_2, \ldots, a_N \leq b_N$. To this data we associate the set

$$B(a_1, b_1, \ldots, a_N, b_N) := \{\mathbf{x} = (x_1, \ldots, x_N) := \forall 1 \leq i \leq N, a_i \leq x_i \leq b_i\}.$$

A set $B(a_1, b_1, \ldots, a_N, b_N)$ is called a **closed box**. A closed box in $\mathbb{R}$ is simply a closed bounded interval. A closed box in $\mathbb{R}^2$ is a rectangle (together with its interior) whose edges are parallel to the coordinate axes. And so forth. Now:

LEMMA 1.9. *A subset* $X \subseteq \mathbb{R}^N$ *is bounded if and only if* X *is contained in some closed box. In particular, all closed boxes are bounded subsets of* $\mathbb{R}^N$*.*

We leave the proof of Lemma 1.9 as an exercise.

THEOREM 1.10 (Bolzano-Weierstrass in $\mathbb{R}^N$). *Every bounded sequence in $\mathbb{R}^N$ has a convergent subsequence.*

PROOF. Let $\{\mathbf{x}_n\}$ be a bounded sequence in $\mathbb{R}^N$. By Lemma1.9 there are real numbers $a_1 \leq b_1, \ldots, a_N \leq b_N$ such that every term $\mathbf{x}_n$ of the sequence lies in the box $B(a_1, b_1, \ldots, a_N, b_N)$.
Step 1: The sequence $\{\mathbf{x}_{n,1}\}$ of first coordinates lies in the interval $[a_1, b_1]$, so by Bolzano-Weierstrass in $\mathbb{R}$ it has a subsequence that converges to $L_1 \in \mathbb{R}$.

Interregnum: We now have a purely notational pitfall to avoid: we are going to be passing to subsequences quite a lot of times, so if we actually write this out using double index notation then in Step 2 we are going to get triple indices, in Step 3 quadruple indices, and so forth: it will be a terrible mess. So we will just *remember* that we passed to a subsequence so as to make the sequence of first coordinates converge.
Step 2: The sequence (which is actually a subsequence of our original sequence) $\mathbf{x}_{n,2}$ of second coordinates lies in the interval $[a_2, b_2]$, so by Bolzano-Weierstrass in $\mathbb{R}$ it has a subsequence that converge to $L_2 \in \mathbb{R}$. What happens with the sequence of first coordinates when we do this? Fortunately, if a sequence converges to a limit then every subsequence converges to the same limit, so passing to this second subsequence does not screw up what we did in Step 1: after two steps we have passed to a subsubsequence – which is still a subsequence! – of the original sequence so as to make each of the first two component real sequences converge.
Steps 3 to N: We move on to the bounded sequence of third components, apply Bolzano-Weierstrass again, and so forth. After N steps we have passed to a subsequence N times altogether to get a sequence in which each of the component sequences converge, hence by Theorem 1.6 the subsub.....subsequence converges. Passing from a sequence to a subsequence any finite number of times still yields a subsequence of the original sequence, so...we're done. $\square$

A point $L \in \mathbb{R}^N$ is a **partial limit** of a sequence $\{\mathbf{x}_n\}$ in $\mathbb{R}^N$ if there is some subsequence $\mathbf{x}_{n_k} \to L$. Thus Theorem 1.10 can be rephrased as: every bounded sequence in $\mathbb{R}^N$ has at least one partial limit.

When $N = 1$ we can go a little farther, saying that $+\infty$ is a partial limit of the real sequence $\{\mathbf{x}_n\}$ if some subsequence diverges to $+\infty$ and that $-\infty$ is a partial limit of $\{\mathbf{x}_n\}$ if some subsequence diverges to $-\infty$. By Exercise 1.6, a sequence has $+\infty$ as a partial limit if and only if it is unbounded above and a sequence has $-\infty$ as a partial limit if and only if it is unbounded below. Using Bolzano-Weierstrass, it follows that every real sequence has at least one partial limit in the extended real numbers $[-\infty, \infty]$.

Exercises.

EXERCISE 1.5. *Let $\{x_n\}$ be a real sequence.*

a) *Show that the following are equivalent:*
   (i) *We have $x_n \to +\infty$.*
   (ii) *Every subsequence $\{x_{n_k}\}$ diverges to $+\infty$.*
   (iii) *Every subsequence $\{x_{n_k}\}$ is unbounded above.*
b) *State and prove the analogue of part a) for sequences diverging to $-\infty$.*

EXERCISE 1.6. *Let $\{x_n\}$ be a real sequence.*

a) *Show that the following are equivalent:*
   (i) *There is a subsequence $\{x_{n_k}\}$ such that $x_{n_k} \to +\infty$.*
   (ii) *The sequence $\{x_n\}$ is unbounded above.*

b) *Show that the following are equivalent:*
   (i) *There is a subsequence $\{x_{n_k}\}$ such that $x_{n_k} \to -\infty$.*
   (ii) *The sequence $\{x_n\}$ is unbounded below.*

EXERCISE 1.7. *Let $\{\mathbf{x}_n\}$ be a sequence in $\mathbb{R}^N$ such that $\mathbf{x}_n \to L$, and let $\alpha \in \mathbb{R}$. Show: $\alpha\mathbf{x}_n \to \alpha L$.*

EXERCISE 1.8. *Let $\{\mathbf{x}_n\}$, $\{\mathbf{y}_n\}$ be sequences in $\mathbb{R}^N$. Suppose that $\mathbf{x}_n \to L$ and $\mathbf{y}_n \to M$. Show: $\mathbf{x}_n + \mathbf{y}_n \to L + M$.*

EXERCISE 1.9. *Let $\{\mathbf{x}_n\}$ be a sequence in $\mathbb{R}^N$. Show: if $\mathbf{x}_n \to \mathbf{L}$, then every subsequence $\{\mathbf{x}_{n_k}\}$ also converges to $\mathbf{L}$.*

EXERCISE 1.10. *Let $\{\mathbf{x}_n\}$ be a sequence in $\mathbb{R}^N$, and let $\mathbf{L}, \mathbf{P} \in \mathbb{R}^N$. Suppose that $\mathbf{x}_n \to \mathbf{L}$ and $\mathbf{x}_n \to \mathbf{P}$. Show: $\mathbf{L} = \mathbf{P}$.*

EXERCISE 1.11. *Show: every finite subset of $\mathbb{R}^N$ is bounded.*

EXERCISE 1.12. *Let $\{\mathbf{x}_n\}$ be a sequence in $\mathbb{R}^N$.*

a) *Show: if $\{\mathbf{x}_n\}$ is bounded, so is every subsequence.*
b) *Show: $\{\mathbf{x}_n\}$ is unbounded if and only if some subsequence of $\{\mathbf{x}_n\}$ diverges to ∞.*

EXERCISE 1.13. *A sequence $\{\mathbf{x}_n\}$ in $\mathbb{R}^N$ is Cauchy if and only if for all $1 \leq i \leq N$, the real sequence $\{\mathbf{x}_{n,i}\}$ is Cauchy.*

EXERCISE 1.14. *Suppose we are given real numbers $a_1 \leq b_1, a_2 \leq b_2, \ldots, a_N \leq b_N$. To this data we associate the set*

$$B(a_1, b_1, \ldots, a_N, b_N) := \{\mathbf{x} = (x_1, \ldots, x_N) := \forall 1 \leq i \leq N, a_i \leq x_i \leq b_i\}.$$

A set of the form $B(a_1, b_1, \ldots, a_N, B_N)$ is called a ***closed box****. Show: a subset $X \subseteq \mathbb{R}^N$ is bounded if and only if X is contained in some closed box.*

EXERCISE 1.15. *Let $1 \leq i \leq N$. We define the* ***coordinate projection map***

$$\pi_i : \mathbb{R}^N \to \mathbb{R},\ (x_1, \ldots, x_N) \mapsto x_i.$$

Show: a subset $X \subseteq \mathbb{R}^N$ is bounded if and only if for all $1 \leq i \leq N$, the subset $\pi_i(X)$ is a bounded subset of $\mathbb{R}$.

EXERCISE 1.16. *For a sequence $\{\mathbf{x}_n\}$ in $\mathbb{R}^N$, show the following are equivalent:*

(i) *The sequence diverges to ∞ (recall this means that $||\mathbf{x}_n|| \to +\infty$).*
(ii) *Every subsequence of $\mathbf{x}_n$ is unbounded.*
(iii) *The sequence $\{\mathbf{x}_n\}$ has no partial limit.*

COMMENT: Let $N \geq 2$. We could define the "extended Euclidean space"

$$\widetilde{\mathbb{R}^N} := \mathbb{R}^N \cup \{\infty\}$$

and say that ∞ is a partial limit of a sequence $\{\mathbf{x}_n\}$ in $\mathbb{R}^N$ if and only if some subsequence diverges to ∞. With this convention, combining Exercises 1.12 and 1.16 we would have that *every* sequence in $\mathbb{R}^N$ has at least one partial limit in $\widetilde{\mathbb{R}^N}$. We will not adopt this definition simply because we will not use it in this course.

EXERCISE 1.17. *By Bolzano-Weierstrass, every bounded sequence in $\mathbb{R}^N$ has at least one partial limit. Show that a bounded sequence converges if and only if it has* exactly *one partial limit.*

EXERCISE 1.18. *Find a sequence in $\mathbb{R}^N$ having* every $\mathbf{x} \in \mathbb{R}^N$ *as a partial limit.*

EXERCISE 1.19 (Geometric Pigeonhole Principle). *Let $A \subset \mathbb{R}^N$ be bounded, and let $\{\mathbf{x}_n\}_{n=1}^{\infty}$ be a sequence in A. Show: for all $\delta > 0$, there are distinct positive integers m and n such that $||\mathbf{x}_m - \mathbf{x}_n|| < \delta$.*
(Suggestion: put X inside a closed box B_1 and then "bisect" B_1 into 2^N subboxes each with half the side lengths of B_1. Since we have infinitely many terms of the sequence and only finitely many subboxes, infinitely many terms of the sequence must lie in at least one of the "stage 2 subboxes." Repeat this bisection until the subboxes are so small that any 2 points lying in the same subbox have distance less than δ.)

The following exercise uses a notation that is not officially defined until Section 1.4: for $\mathbf{x} \in \mathbb{R}^N$ and $r > 0$, we put

$$B^{\bullet}(\mathbf{x}, r) := \{\mathbf{y} \in \mathbb{R}^N \mid ||\mathbf{x} - \mathbf{y}|| \leq r\}.$$

EXERCISE 1.20. *A subset $A \subseteq \mathbb{R}^N$ is* ***totally bounded*** *if for all $\delta > 0$ there are finitely many points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^N$ such that*

$$A \subseteq \bigcup_{i=1}^{n} B^{\bullet}(\mathbf{x}_i, \delta).$$

a) *Show: every bounded subset of $\mathbb{R}^N$ is totally bounded.*
b) *Deduce the Geometric Pigeonhole (Exercise 1.19) from part a).*

Problems.

PROBLEM 1.3. *The set $\mathbb{C} = \{x + iy \mid x, y \in \mathbb{R}\}$ has a nice multiplication operation:*

$$(x + iy)(z + iw) = (xz - yw) + (xw + zy)i.$$

If we identify the vector (x, y) with the complex number $x + iy$, this gives a multiplication operation on $\mathbb{R}^2$:

$$(x, y) \cdot (z, w) := (xz - yw, xw + zy).$$

Show: if $\{\mathbf{x}_n\}$ and $\{\mathbf{y}_n\}$ are two sequences in $\mathbb{R}^2$ such that $\mathbf{x}_n \to L$ and $\mathbf{y}_n \to M$, then $\mathbf{x}_n \cdot \mathbf{y}_n \to L \cdot M$.

PROBLEM 1.4. *Let $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^3$, and let $\mathbf{x} \times \mathbf{y} \in \mathbb{R}^3$ be the cross product. Show: if $\mathbf{x}_N \to L$ and $\mathbf{y}_N \to M$ then $\mathbf{x}_n \times \mathbf{y}_n \to L \times M$.*

PROBLEM 1.5. *Let $\{x_n\}$ be a real sequence.*

a) *Show: $\{x_n\}$ has at least one partial limit in $[-\infty, \infty]$.*
b) *Show: $\{x_n\}$ converges if and only if it has a unique partial limit L, which is morever finite.*

The following problem uses notation that is not officially defined until Section 1.4: for $\mathbf{x} \in \mathbb{R}^N$ and $r > 0$, we put

$$B^{\circ}(\mathbf{x}, r) := \{\mathbf{y} \in \mathbb{R}^N \mid ||\mathbf{x} - \mathbf{y}|| < r\}.$$
$$B^{\bullet}(\mathbf{x}, r) := \{\mathbf{y} \in \mathbb{R}^N \mid ||\mathbf{x} - \mathbf{y}|| \leq r\}.$$

PROBLEM 1.6. *For a nonempty subset $A \subseteq \mathbb{R}^N$, let*

$$\operatorname{diam}(A) := \sup\{||\mathbf{x} - \mathbf{y}|| \mid \mathbf{x}, \mathbf{y} \in A\}.$$

a) *Suppose $\varnothing \subsetneq B \subseteq A$. Show:* $\operatorname{diam}(B) \leq \operatorname{diam}(A)$.
b) *Show that for all $R \geq 0$ and all $x \in \mathbb{R}^N$, we have*

$$\operatorname{diam}(B^\circ(x, R)) = \operatorname{diam}(B^\bullet(x, R)) = 2R.$$

c) *Show: A is bounded $\iff$* $\operatorname{diam}(A) < \infty$.

3. Sequential Limits Superior and Inferior

That a real sequence $\{x_n\}$ need not converge is a basic fact of life...but that dosen't stop it from being annoying sometimes. Already in calculus one sees the beginning of a workaround with the notion of sequences that diverge to ∞ or to $-\infty$: such sequences are certainly not convergent, but they exhibit a definite limiting behavior that can still often be used. But as we well know, a sequence $\{x_n\}$ may diverge "due to oscillation": in our language of partial limits, this means that the sequence has more than one partial limit in $[-\infty, \infty]$.

For a real sequence $\{x_n\}$, we define the **limit superior** $\overline{\lim}\, x_n$ to be the supremum in $[-\infty, \infty]$ of the set of partial limits of $\{\mathbf{x}_n\}$, i.e., the supremum of the set of subsequential limits. This supremum exists in $[-\infty, \infty]$ by Proposition 0.2. But in fact more is true:

THEOREM 1.11. *Let $\{x_n\}$ be a real sequence. There is a subsequence $\{x_{n_k}\}$ such that $x_{n_k} \to \overline{\lim}\, x_n$.*

PROOF. Case 1: Suppose $\overline{\lim}\, x_n = \infty$. Thus for all $M \in \mathbb{R}$ there is a subsequence $\{x_{n_k}\}$ that either converges to $L \geq M$ or diverges to $+\infty$. It follows that for all $M \in \mathbb{R}$ there are infinitely many terms of $\{xn_\}$ that are at least M. Thus the sequence is unbounded above, so $+\infty$ is a subsequential limit by Exercise 1.6.
Case 2: Suppose $\overline{\lim}\, x_n = -\infty$. This means that every subsequence of x_n diverges to $-\infty$, so by Exercise 1.5 we have $x_n \to -\infty$.
Case 3: Suppose $\overline{\lim}\, x_n = M \in \mathbb{R}$. Then for every $\epsilon > 0$ there is a subsequence converging to some L with $M - \epsilon \leq L \leq M$. In particular, for all $k \in \mathbb{Z}^+$ there are infinitely many terms of the sequence that are at least $M - \frac{1}{k}$, so we may choose n_1 to be such that $x_{n_1} \geq M - 1$, $n_2 > n_1$ such that $x_{n_2} \geq M - \frac{1}{2}$, and so forth. Then $x_{n_k} \to M$. $\square$

Thus $\overline{\lim}\, x_n$ is not just the supremum of the subsequential limits, it is actually the *maximum* of the subsequential limits.

In a similar way we define the **limit inferior** $\underline{\lim}\, x_n$ to be the infimum in $[-\infty, \infty]$ of the set of partial limits of $\{x_n\}$, i.e., the infimum of the set of subsequential limits. In Exercise 1.21 you are asked to show that the liminf is also attained as a subsequential limit.

The following result says in particular that "divergence due to oscillation" means precisely that the liminf is smaller than the limsup.

PROPOSITION 1.12. *Let $\{x_n\}$ be a real sequence.*

a) *The sequence converges if* $\underline{\lim}\, x_n$ *and* $\overline{\lim}\, x_n$ *are equal and finite, in which case the common value is the limit.*
b) *The sequence diverges to* $+\infty$ *if and only if* $\underline{\lim}\, x_n = \overline{\lim}\, x_n = +\infty$.
c) *The sequence diverges to* $-\infty$ *if and only if* $\underline{\lim}\, x_n = \overline{\lim}\, x_n = -\infty$.

PROOF. A sequence $\{x_n\}$ has a unique partial limit precisely when

$$\underline{\lim}\, x_n = \overline{\lim}\, x_n.$$

So part a) follows from Exercise 1.5 and parts b) and c) follow from Exercise 1.5. □

We will give two more characterizations of the limits superior and inferior. First:

PROPOSITION 1.13.
Let $\{x_n\}_{n=1}^{\infty}$ *be a real sequence. For* $N \in \mathbb{Z}^+$, *put* $X_N := \{x_n \mid n \geq N\}$.

a) *We have*

$$\overline{\lim}\, x_n = \lim_{N\to\infty} \sup X_N$$

b) *We have*

$$\underline{\lim}\, x_n = \lim_{N\to\infty} \inf X_N.$$

PROOF. We will prove part a) and leave part b) as an exercise.
a) Notice that X_1 is the set of all terms of the sequence, X_2 is the set of all terms of the sequences starting with the second term, and so forth, so $\{X_N\}_{N=1}^{\infty}$ is a nested sequence of sets. Exercise 0.4 gives

$$\sup X_1 \geq \sup X_2 \geq \ldots \geq \sup X_N \geq \ldots,$$

i.e., the sequence $\{\sup X_N\}$ is decreasing. Therefore it converges if it is bounded below and otherwise diverges to $-\infty$.
Case 1: Suppose that the sequence $\{x_n\}$ is unbounded above, meaning that $\overline{\lim}\, x_n = +\infty$. Whether a sequence is unbounded above is not affected by removing finitely many terms, so this implies that for all $N \in \mathbb{Z}^+$ the set X_N is unbounded above, so $\sup X_N = +\infty$ for all N, so $\overline{\lim}\, X_N = +\infty = \overline{\lim}\, x_n$ in this case.
Case 2: Suppose that $\{x_n\}$ is bounded above and and that $\{\sup X_N\}$ is bounded below by $m \in \mathbb{R}$ so that

$$m = \lim_{N\to\infty} \sup X_N.$$

Then for all $N \in \mathbb{Z}^+$ we have $m \leq \sup X_N$, so for all $\epsilon > 0$, then the number $m-\epsilon$ is not an upper bound for X_N: there is some $n \geq N$ such that $m-\epsilon < x_n$. Because of this we can choose a subsequence x_{n_k} such that $x_{n_k} \geq m-\epsilon$ for all k, so any partial limit of that subsequence must be at least $m-\epsilon$. Since ϵ was arbitrary, this means that $\overline{\lim}\, x_n \geq m$. If $m = +\infty$ then we must have $\overline{\lim}\, x_n = +\infty$ and we're done, so suppose that m is finite. Seeking a contradiction, suppose that $\overline{\lim}\, x_n > m$. Then there is a subsequence $\{x_{n_k}\}$ converging to some real number $M > m$ (recall that we have assumed that $\{x_n\}$ is bounded above). Choose m' such that

$$m < m' < M.$$

Then for all sufficiently large k we have $x_{n_k} > m'$, which implies that for all N, $\sup X_N \geq m'$, so

$$\lim_{N\to\infty} \sup X_N \geq m' > m,$$

a contradiction.
Case 3: Suppose that $\{x_n\}$ is bounded above and that $\{\sup X_N\}$ diverges to $-\infty$.

This means that for any $m \in \mathbb{R}$, for only finitely many $n \in \mathbb{Z}^+$ can we have $x_n \geq m$, which means that $x_n \to -\infty$, in which case

$$\overline{\lim}\, x_n = -\infty = \lim_{N \to \infty} \sup X_N. \qquad \square$$

Whereas limsups and liminfs are designed to deal with the case of oscillation in sequences, Proposition 1.12 reduces the computation of limsups and liminfs to the monotone case: the sequence $\{\sup X_N\}$ is always decreasing while the sequence $\{\inf X_N\}$ is always increasing. Thus inside the extended real numbers, the limsup is computed as a "minimax" – we maximize each X_N and then take the minimum of the corresponding sequence – while the liminf is computed as a "maximin" – we minimize each X_N and then take the maximum of the corresponding sequence. Dealing with an oscillatory quantity by first maximizing then minimizing will be a recurring theme in our course.

Next we want to give a "creeping" interpretation of limsups in the bounded case:

PROPOSITION 1.14. *Let $\{x_n\}$ be a bounded real sequence. Then $\overline{\lim}\, \{x_n\}$ is the unique real number M with the following property: for all $\epsilon > 0$, there are infinitely many $n \in \mathbb{Z}^+$ such that $x_n > M - \epsilon$ and there are only finitely many $n \in \mathbb{Z}^+$ such that $x_n > M + \epsilon$.*

PROOF. Let M be a real number such that for all $\epsilon > 0$, there are infinitely many $n \in \mathbb{Z}^+$ such that $x_n > M - \epsilon$ and there are only finitely many $n \in \mathbb{Z}^+$ such that $x_n > M + \epsilon$. The former condition implies that the sequence $\{\mathbf{x}_n\}$ has a partial limit that is at least $M - \epsilon$ and the latter condition implies that the sequence has no partial limit that is at least $M + \epsilon$. The first condition means that M can be no larger than $\overline{\lim}\, \mathbf{x}_n$ and the second condition means that M can be no smaller than $\overline{\lim}\, \mathbf{x}_n$, so if such an M exists we must have $M = \overline{\lim}\, \mathbf{x}_n$. Very similar reasoning shows that $\overline{\lim}\, \mathbf{x}_n$ has both of these properties. $\square$

Thus for a bounded real sequence $\{x_n\}$, as one ascends the real line the limit superior is the threshold at which one goes from having infinitely many terms of the sequene above us to only finitely many terms of the sequence above us. However the ϵ in the statement is necessary: for instance, if a real sequence converges to 0 then we know that for any $\epsilon > 0$ all but finitely many of its terms lie in the intervel $[-\epsilon, \epsilon]$. We don't know however how many terms are greater than zero or even greater than equal to zero: in fact, for any subset T of $\mathbb{Z}^+$ there is a sequence $x_n \to 0$ such that x_n is negative if and only if $n \in T$.

Exercises.

EXERCISE 1.21. *Let $\{x_n\}$ be a real sequence. Show: there is a subsequence $\{x_{n_k}\}$ such that $x_{n_k} \to \underline{\lim}\, x_n$.*

EXERCISE 1.22. *State and prove an analgoue of Proposition 1.14 for the $\underline{\lim}$ of a bounded sequence.*

Problems.

PROBLEM 1.7. *Let $\{x_n\}$ be a real sequence. Show:*

$$\overline{\lim}\, x_n = -\underline{\lim}\, (-x_n) \text{ and } \underline{\lim}\, x_n = -\overline{\lim}\, (-x_n).$$

PROBLEM 1.8. *Let $\{x_n\}$ and $\{y_n\}$ be bounded real sequences.*

a) *Show:* $\overline{\lim}\ (x_n + y_n) \leq \overline{\lim}\ x_n + \overline{\lim}\ y_n$.
b) *Show:* $\overline{\lim}\ x_n + \underline{\lim}\ y_n \leq \overline{\lim}\ (x_n + y_n)$.
(Hint: $x_n = (x_n + y_n) + (-y_n)$.*)*

PROBLEM 1.9. *Prove part b) of Proposition 1.13.*

4. Topology of $\mathbb{R}^N$

4.1. Open and Closed Sets. Let $\mathbf{x} \in \mathbb{R}^N$ and $r > 0$. We define the **open ball centered at x with radius** r to be

$$B^\circ(\mathbf{x}, r) := \{\mathbf{y} \in \mathbb{R}^N \mid ||\mathbf{x} - \mathbf{y}|| < r\}$$

and the **closed ball centered at x with radius** r to be

$$B^\bullet(\mathbf{x}, r) := \{\mathbf{y} \in \mathbb{R}^N \mid ||\mathbf{x} - \mathbf{y}|| \leq r\}.$$

Thus $B^\circ(\mathbf{x}, r)$ consists of all points of $\mathbb{R}^N$ whose distance from $\mathbf{x}$ – called the **center** of the ball – is less than r, and the same goes for $B^\bullet(\mathbf{x}, r)$ except that now the distance is less than *or equal to* r.

The latter concept has actually arisen already: to see this, notice that our definition of a subset of $\mathbb{R}^N$ being bounded is precisely that it is contained in $B^\bullet(0, M)$ for some M. Notice also that open and closed balls are always bounded sets: indeed, for any $\mathbf{x} \in \mathbb{R}^N$ and $r > 0$, the Triangle Inequality gives:

$$B^\circ(\mathbf{x}, r) \subseteq B^\bullet(\mathbf{x}, r) \subseteq B^\bullet(0, ||\mathbf{x}|| + r).$$

In $\mathbb{R}^1$ balls are not very interesting: you are asked to show as an exercise that a subset of $\mathbb{R}$ is an open ball if and only if it is a bounded open interval and that a subset of $\mathbb{R}$ is a closed ball if and only if it is a bounded closed interval. (This is actually one reason why we want to work in $\mathbb{R}^N$ at the beginning of the course: the topological concepts we want to deal with are not trivial when we restrict to the one variable case, but they are "geometrically degenerate" in a way that may hamper intuition.)

We say that a subset $U \subseteq \mathbb{R}^N$ is **open** if for every $\mathbf{x} \in U$, there is $\epsilon > 0$ such that $B^\circ(\mathbf{x}, \epsilon) \subseteq U$. In other words, a set is open if whenever it contains a point $\mathbf{x}$ it also contains all points of $\mathbb{R}^N$ that are sufficiently close to $\mathbf{x}$.

The terminology suggests than an open ball should itself be an open set, but we had better prove that.

PROPOSITION 1.15. *Every open ball is an open subset of* $\mathbb{R}^N$.

PROOF. Consider the open ball $B^\circ(\mathbf{x}, r)$ and a point $\mathbf{y}$ in it. We need to find a (smaller!) ball centered at $\mathbf{y}$ that is entirely contained in the first ball. The question really is: how do we choose $\epsilon > 0$ such that

$$B^\circ(\mathbf{y}, \epsilon) \subseteq B^\circ(\mathbf{x}, r)?$$

Let's try to work it out: if $\mathbf{z} \in B^\circ(\mathbf{y}, \epsilon)$, then

$$d(\mathbf{y}, \mathbf{z})) = ||\mathbf{z} - \mathbf{y}|| < \epsilon,$$

so by the Triangle Inequality we have

$$d(\mathbf{x}, \mathbf{z}) < \epsilon + d(\mathbf{x}, \mathbf{y}).$$

So the point $\mathbf{z}$ will lie in $B^\circ(\mathbf{x}, r)$ provided that $\epsilon + d(\mathbf{x}, \mathbf{y}) \leq r$. Thus we can take

$$\epsilon = r - d(\mathbf{x}, \mathbf{y}),$$

which is indeed positive: since $\mathbf{y} \in B^\circ(\mathbf{x}, r)$, we have $d(\mathbf{x}, \mathbf{y}) < r$. □

Non-examples are as helpful as understanding new concepts as examples, so:

PROPOSITION 1.16. *No closed ball is an open subset of* $\mathbb{R}^N$.

PROOF. Consider $B := B^\bullet(\mathbf{x}, r)$. Let $p := \mathbf{x} + (r, 0, \ldots, 0)$ be the rightmost point on the ball. Then any open ball $B^\circ(p, \epsilon)$ contains the point $p + (\frac{\epsilon}{2}, 0, \ldots, 0) = \mathbf{x} + (r + \frac{\epsilon}{2}, 0, \ldots, 0)$. This point has distance $r + \frac{\epsilon}{2}$ from $\mathbf{x}$ so does not lie in B. □

Let A be a subset of $\mathbb{R}^N$. We say that $L \in \mathbb{R}^N$ is a **limit point** of A if there is a sequence $\{\mathbf{x}_n\}$ in A such that $\mathbf{x}_n \to L$.

Every point $L \in A$ is a a limit point of A, because we can take the constant sequence $L, L, L, \ldots$. (If this feels like cheating...good! You are probably grasping for the related concept of **accumulation point**, which is coming up soon.)

EXAMPLE 1.2. *We claim that every point of the closed ball $B^\bullet(\mathbf{x}, r)$ is a limit point of the corresponding open ball $B^\circ(\mathbf{x}, r)$. Indeed, we need only look at points $p \in B^\bullet(\mathbf{x}, r) \setminus B^\circ(\mathbf{x}, r)$, i.e., points p whose distance from $\mathbf{x}$ is exactly r. Then take*

$$\mathbf{x}_n = \mathbf{x} + (1 - \frac{1}{n})(p - \mathbf{x}).$$

Then

$$d(\mathbf{x}_n, x) = (1 - \frac{1}{n})||p - \mathbf{x}|| = (1 - \frac{1}{n})r < r,$$

so $\mathbf{x}_n \in B^\circ(\mathbf{x}, r)$. And $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x} + (p - \mathbf{x}) = p$.

This example motivates the second key definition of this section: a subset A of $\mathbb{R}^N$ is **closed** if every limit point of A is an element of A. Another way of saying this is that A is *closed* under taking limits of convergent sequences.

A basic fact in Math 3100 is that limits of sequences preserve non-strict inequalities: that is, if we every term of a convergent sequence is at least a, then the limit is also at least a, and if every term of a convergent sequence is at most b, then the limit is also at most b. This means precisely that the closed interval $[a, b]$ is a closed subset of $\mathbb{R}$. Recalling that these are precisely the closed balls in $\mathbb{R}$, we get that every closed ball in $\mathbb{R}^1$ is a closed subset of $\mathbb{R}^1$.

We would like to extend this to $\mathbb{R}^N$: let's try. Consider a closed ball $B^\bullet(\mathbf{x}, r)$ in $\mathbb{R}^N$. Seeking a contradiction, suppose that there is some $L \in \mathbb{R}^N \setminus B^\bullet(\mathbf{x}, r)$ and a sequence $\{\mathbf{x}_n\}$ in $B^\bullet(\mathbf{x}, r)$ such that $\mathbf{x}_N \to L$. Let

$$d := d(\mathbf{x}, L)$$

be the distance from the limit point to the center of the ball. Our assumption is that $d > r$. Take $\epsilon := d - r$. I claim that

$$B^\circ(L, \epsilon) \cap B^\bullet(\mathbf{x}, r) = \varnothing.$$

Indeed, if $y \in B^\circ(L, \epsilon) \cap B^\bullet(\mathbf{x}, r)$ then

$$d = d(\mathbf{x}, L) \leq d(\mathbf{x}, y) + d(y, L) < r + \epsilon = d.$$

That's a contradiction. But if we had a sequence in $B^\bullet(\mathbf{x}, r)$ converging to L then sufficiently large terms of the sequence will give elements of $B^\bullet(\mathbf{x}, r)$ that are less than ϵ away from L, so there is no such sequence. Therefore closed balls are closed.

If we look back this proof, we really showed that for every point L of the complement $\mathbb{R}^N \setminus B^\bullet(\mathbf{x}, r)$, there is an open ball centered at L and contained in the complement. In other words, we showed that $B^\bullet(\mathbf{x}, r)$ is closed essentially by showing that its complement was open. This is true in general, very important, and not so difficult to prove.

THEOREM 1.17. *A subset $A \subseteq \mathbb{R}^N$ is closed if and only if its complement $\mathbb{R}^N \setminus A$ is open.*

You are asked to prove Theorem 1.17 as an exercise.

4.2. Continuous Functions. Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is continuous at a point $c \in \mathbb{R}$ if: for all $\epsilon > 0$, there is $\delta > 0$ such that for all $x \in \mathbb{R}$, if $|x - c| < \delta$ then $|f(x) - f(c)| < \epsilon$. A function $f : \mathbb{R} \to \mathbb{R}$ is continuous if it is continuous at every $c \in \mathbb{R}$.

If we have a function f defined not on all of $\mathbb{R}$ but only on some subset A, then we used the same definition as above but with one reasonable change: $f : A \to \mathbb{R}$ is continuous at $c \in A$ if for all $\epsilon > 0$ there is $\delta > 0$ such that for all $x \in A$, if $|x - c| < \delta$ then $|f(x) - f(c)| < \epsilon$.

If we observe that $|f(x) - f(c)| < \epsilon$ means that $d(f(x), f(c)) < \epsilon$ and $|x - c| < \delta$ means $d(x, c) < \delta$, it should be pretty clear how to generalize this definition to maps between Euclidean spaces. Again, let's do it in two steps. First suppose that M, N are positive integers and we have

$$f : \mathbb{R}^N \to \mathbb{R}^M.$$

We say that f is continuous at $c \in \mathbb{R}^N$ if for all $\epsilon > 0$, there is $\delta > 0$ such that for all $\mathbf{x} \in \mathbb{R}^N$, if $d(\mathbf{x}, c) = ||\mathbf{x} - c|| < \delta$ then $d(f(x), f(c)) = ||f(\mathbf{x}) - f(c)|| < \epsilon$. We say f is continuous if it is continuous at every $c \in \mathbb{R}^N$.

And again it is no problem to make a more general definition: if A is a subset of $\mathbb{R}^N$ and $f : A \to \mathbb{R}^M$ is a function, then f is continuous at $c \in \mathbb{R}^N$ if for all $\epsilon > 0$ there is $\delta > 0$ such that for all $\mathbf{x} \in A$, if $d(\mathbf{x}, c) = ||\mathbf{x} - c|| < \delta$ then $||f(\mathbf{x}) - f(c)|| < \epsilon$.

Let us rephrase this definition in terms of open balls. A function $f : A \to \mathbb{R}^N$ is continuous at $\mathbf{c} \in A$ if for all $\epsilon > 0$, there is some $\delta > 0$ such that f maps $A \cap B^\circ(\mathbf{c}, \delta)$ into $B^\circ(f(\mathbf{c}), \epsilon)$.

The following is an extension of an important result from Math 3100: continuous functions are characterized by their preservation of limits of convergent sequences.

THEOREM 1.18. *Let $X \subseteq \mathbb{R}^N$, and let $f : X \to \mathbb{R}^M$ be a function. Let $\mathbf{c} \in X$. The following are equivalent:*

(i) *f is continuous at $\mathbf{c}$.*
(ii) *For every sequence $\{\mathbf{x}_n\}$ in X such that $\mathbf{x}_n \to \mathbf{c}$, we have $f(\mathbf{x}_n) \to f(\mathbf{c})$.*

PROOF. (Compare this to [**SS**, Theorem 2.7.5]: it's virtually identical.)
(i) $\implies$ (ii): Fix $\epsilon > 0$. Because f is continuous at $\mathbf{c}$, there is $\delta > 0$ such that for all $\mathbf{x} \in X$ with $||\mathbf{x} - \mathbf{c}|| < \delta$, we have $||f(\mathbf{x}) - f(\mathbf{c})|| < \epsilon$. Because $\mathbf{x}_n \to c$, there is $K \in \mathbb{Z}^+$ such that for all $n > K$ we have $||\mathbf{x}_n - \mathbf{c}|| < \delta$. Thus for all $n > K$ we have $||f(\mathbf{x}_n) - f(\mathbf{c})|| < \epsilon$.
(ii) $\implies$ (i): We will prove the contrapositive: suppose f is *not* continuous at $\mathbf{c}$. Then there is $\epsilon > 0$ such that for all $\delta > 0$ there is $\mathbf{x} \in X$ with $||\mathbf{x} - \mathbf{c}|| < \delta$ and $||f(\mathbf{x}) - f(\mathbf{c})| \geq \epsilon$. For $n \in \mathbb{Z}^+$, taking $\delta = \frac{1}{n}$ gives $\mathbf{x}_n \in X$ such that $||\mathbf{x}_n - \mathbf{c}|| < \frac{1}{n}$ and $||f(\mathbf{x}_n) - f(\mathbf{c})|| \geq \epsilon$. Thus $\mathbf{x}_n \to \mathbf{c}$, but $f(\mathbf{x}_n)$ *does not converge* to $f(\mathbf{c})$. $\square$

4.3. New Continuous Functions From Old. Let us now discuss some ways of building new continuous functions out of old continuous functions. We can start with the real ground floor:

PROPOSITION 1.19. *Let $X \subset \mathbb{R}^N$, and let $f : X \to \mathbb{R}^M$ be a constant function: for all $\mathbf{x}, \mathbf{y} \in X$ we have $f(\mathbf{x}) = f(\mathbf{y})$. Then f is continuous.*

PROOF. Indeed for any $\epsilon > 0$ we may take *any* positive value of δ we like, since in fact for *any* $\mathbf{x}, \mathbf{y} \in X$ we have $||f(\mathbf{x}) - f(\mathbf{y})|| = ||0|| = 0 < \epsilon$. $\square$

After constant functions, perhaps the simplest functions $f : \mathbb{R}^N \to \mathbb{R}$ are the **coordinate functions** or **coordinate projections**: for $1 \leq i \leq N$, put

$$\pi_i : \mathbb{R}^N \to \mathbb{R} \text{ by } (x_1, \ldots, x_N) \mapsto x_i.$$

This is pretty fancy/careful notation. In practice we will often speak of "the function x_i". It is quite easy to see that these functions are continuous: indeed, let $\mathbf{x} \in \mathbb{R}^N$, and fix $\epsilon > 0$. Then for $\mathbf{y} \in \mathbb{R}^N$, we have

$$|x_i - y_i| \leq \sqrt{(x_1 - y_1)^2 + \ldots + (x_N - y_N)^2} = ||\mathbf{x} - \mathbf{y}||,$$

so if $||\mathbf{x} - \mathbf{y}|| < \epsilon$ then also $|x_i - y_i| < \epsilon$, so we may take $\delta = \epsilon$.

Here is one use of the coordinate projections: let $X \subseteq \mathbb{R}^N$ and let $f : X \to \mathbb{R}^M$ be a function. Then for all $x \in X$, we have

$$f(x) = (\pi_1(f(x)), \ldots, \pi_M(f(x)).$$

The notation may momentarily obscure this unprofound identity: we are just reassembling the components of the vector-valued function f. Now we have:

PROPOSITION 1.20. *For a function $f : X \subseteq \mathbb{R}^N \to \mathbb{R}^M$ and $\mathbf{x} \in X$, the following are equivalent:*

(i) *The function f is continuous at $\mathbf{x}$.*
(ii) *Each of the functions $f_1, \ldots, f_M$ is continuous at $\mathbf{x}$.*

The proof uses the same idea as Theorem 1.6 – a vector has small norm if and only if each of its components has small absolute value – and is left as an exercise.

PROPOSITION 1.21. *Let $f_1, \ldots, f_M : X \subseteq \mathbb{R}^N \to \mathbb{R}$, and let $\mathbf{x} \in X$. If each of $f_1, \ldots, f_M$ are continuous at $\mathbf{x}$, then so are $\sum_{i=1}^M f_i$ and $\prod_{i=1}^M f_i$.*

PROOF. Let's use Theorem 1.18: let $\{\mathbf{x}_n\}$ be a sequence in X that converges to $\mathbf{x}$. Since each $f_1, \ldots, f_M$ is continuous at $\mathbf{x}$ we have $f_i(\mathbf{x}_n) \to f_i(\mathbf{x})$ as sequences in $\mathbb{R}$. By the extension Theorem 1.4b) from 2 sequences to M sequences (a completely routine induction argument does this) we know that $f_1(\mathbf{x}_n) + \ldots + f_M(\mathbf{x}_n) \to$

$f_1(\mathbf{x}) + \ldots + f_M(\mathbf{x})$, and applying Theorem 1.18 once more we get that $\sum_{i=1}^M f_i$ is continuous at $\mathbf{x}$. The argument for $\prod_{i=1}^M f_i$ except we use the fact that for real sequences we have $x_n \to L$ and $y_n \to M$ implies $x_n y_n \to LM$ [**SS**, Theorem 2.5.4b)] (and again, its evident extension from 2 sequences to M sequences). □

A function $f : \mathbb{R}^N \to \mathbb{R}$ is a **polynomial** if it is built up out of constant functions and coordinate functions by (finitely!) repeated addition and multiplication. Thus for instance $xyz + 17y^5 - \pi x^2 y^2 z^2$ is a polynomial function. It follows from Propositions 1.19, 1.20 and 1.21 that polynomial functions are continuous. In particular:

COROLLARY 1.22. *The inner product map* $\mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$, $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x} \cdot \mathbf{y}$ *is continuous.*

PROOF. We may identify $\mathbb{R}^N \times \mathbb{R}^N$ with $\mathbb{R}^{2N}$ and then the inner product map is $(x_1, \ldots, x_{2N}) \mapsto x_1 x_{N+1} + x_2 x_{N+2} + \ldots + x_N x_{2N}$. This is a polynomial function, so it is continuous. □

PROPOSITION 1.23. *Let* $M, N, P \in \mathbb{Z}^+$. *Let* $X \subseteq \mathbb{R}^N$ *and* $Y \subseteq \mathbb{R}^M$. *Let* $f : X \to \mathbb{R}^M$ *and* $g : Y \to \mathbb{R}^P$ *be functions. Suppose that* $f(X) \subseteq Y$, *so that the composition* $g \circ f$ *is defined.*

a) *Let* $\mathbf{x} \in X$. *If* f *is continuous at* $\mathbf{x}$ *and* g *is continuous at* $f(\mathbf{x})$, *then* $g \circ f$ *is continuous at* $\mathbf{x}$.
b) *If* f *and* g *are both continuous, so is* $g \circ f$.

PROOF. a) Let $\epsilon > 0$. Since g is continuous at $f(\mathbf{x})$ there is $D > 0$ such that if $\mathbf{w} \in Y$ is such that $||\mathbf{w} - f(\mathbf{x})|| < D$, then $||g(\mathbf{w}) - g(f(\mathbf{x}))|| < \epsilon$. Since f is continuous at $\mathbf{x}$, there is $\delta > 0$ such that if $\mathbf{z} \in X$ is such that $||\mathbf{z} - \mathbf{x}| < \delta$, then $||f(\mathbf{z}) - f(\mathbf{x})|| < D$. So altogether, if $\mathbf{z} \in X$ is such that $||\mathbf{z} - \mathbf{x}|| < \delta$, then $||f(\mathbf{z}) - f(\mathbf{x})|| < D$, so $||g(f(\mathbf{z})) - g(f(\mathbf{x}))|| < \epsilon$, so $g \circ f$ is continuous at $\mathbf{x}$.
b) This follows immediately. □

If we assume as known that the function $\sqrt{x} : [0, \infty) \to \mathbb{R}$ is continuous, then we can also prove that the norm function

$$|| \cdot || : \mathbb{R}^N \to \mathbb{R}$$

is continuous: indeed, it is the composition of the polynomial function $x_1^2 + \ldots x_N^2$ with the square root function. Similarly, the Euclidean distance function

$$d : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R},\ (x_1, \ldots, x_N, y_1, \ldots, y_N) \mapsto \sqrt{(x_1 - y_1)^2 + \ldots + (x_N - y_N)^2}$$

is the composition of the polynomial function $(\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})$ with the square root function, hence is continuous.

As one more application of these ideas, we will prove:

PROPOSITION 1.24. *The addition function* $+ : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N$ *is continuous.*

PROOF. Again, we may identify $\mathbb{R}^N \times \mathbb{R}^N$ with $\mathbb{R}^{2N}$, and then we are trying to show that the function

$$(x_1, \ldots, x_N, x_{N+1}, \ldots, x_{2N}) \mapsto (x_1 + x_{N+1}, \ldots, x_N + x_{2N})$$

is continuous. By Proposition 1.20 it's enough to show that each component is continuous. But the ith component is $x_i + x_{N+i}$, which is a polynomial. □

4.4. Sequential Compactness in $\mathbb{R}^N$. A subset $X \subseteq \mathbb{R}^N$ is **sequentially compact** if every sequence in X has a subsequence that converges to some $L \in X$.

PROPOSITION 1.25. *Let $X \subset \mathbb{R}^N$ be a sequentially compact subset, and let $f : X \to \mathbb{R}^M$ be a continuous function. Then the image $f(X)$ is sequentially compact.*

PROOF. Let $\{\mathbf{y}_n\}$ be a sequence in $f(X)$. By definition of the image, every element of $f(X)$ is of the form $f(x)$ for some $x \in X$, so for each $n \in \mathbb{Z}^+$ we may choose $\mathbf{x}_n \in X$ such that $f(\mathbf{x}_n) = \mathbf{y}_n$. Because X is sequentially compact, there is some subsequence $\{\mathbf{x}_{n_k}\}$ that converges to an element $\mathbf{x}$ of X. By Theorem 1.18 we have

$$y_{n_k} = f(\mathbf{x}_{n_k}) \to f(\mathbf{x}).$$

Since $f(\mathbf{x}) \in f(X)$, this shows that $f(X)$ is sequentially compact. □

The following is actually quite a big theorem.

THEOREM 1.26. *A subset of $\mathbb{R}^N$ is sequentially compact if and only if it is closed and bounded.*

PROOF. Step 1: We show that sequentially compact sets are both closed and bounded. We do this contrapositively.

First suppose that X is not closed. Then there is a sequence $\{\mathbf{x}_n\}$ in X that converges to an element $L \in \mathbb{R}^N \setminus X$. Because every subsequence of a convergent sequence converges *to the same limit*, whatever subsequence we take will still be convergent but the limit will lie outside of X, so X is not sequentially compact.

Now suppose that X is not bounded. We will produce a sequence in X no subsequence of which is convergent. Indeed, since X is not bounded, for all $n \in \mathbb{Z}^+$ there is $\mathbf{x}_n \in X$ with $||x_n|| \geq n$. Such a sequence is unbounded, hence divergent. Moreover, passing to a subsequence $\{x_{n_k}\}$ is no help: $||x_{n_k}|| \geq n_k \geq k$, so every subsequence is unbounded. (In other words, this sequence diverges to ∞, hence so does every subsequence.) So X is not sequentially compact.

Step 2: Suppose X is closed and bounded. Let $\{\mathbf{x}_n\}$ be a sequence in X. Since X is bounded, by Bolzano-Weierstrass, there is a subsequence that converges to some $L \in \mathbb{R}^N$. Since X is closed, we have $L \in X$. So X is sequentially compact. □

At this point you're probably thinking: "Hey, I'm not impressed with sequential compactness because it turns out to be a fancy way to say closed and bounded." Let me try to debunk this. First, even if you want to think of it that way, we have learned something very important about closed and bounded subsets of Euclidean spaces. Namely, putting together the last two results, we (immediately!) get:

COROLLARY 1.27. *Let $X \subseteq \mathbb{R}^N$ be closed and bounded, and let $f : X \to \mathbb{R}^M$ be continuous. Then the image $f(X)$ is a closed and bounded subset of $\mathbb{R}^M$.*

On the other hand, this does not work for either closedness or boundedness alone.

EXAMPLE 1.3.

a) *Consider $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \frac{1}{x^2+1}$. Then f is continuous and $f(\mathbb{R}) = (0,1)$, so f takes the closed set $\mathbb{R}$ to the not-closed set $(0,1)$.*
b) *Consider $f : (0,1) \to \mathbb{R}$ by $f(x) = \frac{1}{x}$. Then f is cotinuous and $f((0,1)) = (1,\infty)$, so f takes the bounded set $(0,1)$ to the unbounded set $(1,\infty)$.*

The actual answer is a bit more complicated though. All of the concepts that we have introduced in the course so far can in fact be studied in much more generality: namely in any metric space (we will learn a bit about metric spaces at the end of the course). If we have a subset X of any metric space, then it will turn out that if it is sequentially compact then it must also be closed and bounded, but in a general metric space a closed, bounded subset does not need to be sequentially compact. To get a glimpse of this, imagine we were working in $\mathbb{Q}$ instead of $\mathbb{R}$, with all the rest of the definitions being the same. Then

$$[0,2]_{\mathbb{Q}} := \{x \in \mathbb{Q} \mid 0 \leq x \leq 2\}$$

is a closed, bounded subset of $\mathbb{Q}$, but it is *not* sequentially compact: there is a sequence in $[0,2]_{\mathbb{Q}}$ that converges to the irrational real number $\sqrt{2}$, hence so does every subsequence, hence no subsequence converges to an element *of* $[0,2]_{\mathbb{Q}}$.

Just as in Math 3100 we used Bolzano-Weierstrass in $\mathbb{R}$ to show that Cauchy sequences in $\mathbb{R}$ must converge, pretty much the same argument will show that in a sequantially compact metric space, every Cauchy sequence must converge. So sequential compactness has something to do with *completeness*, but it is even stronger, since Cauchy sequences in $\mathbb{R}^N$ converge but $\mathbb{R}^N$ is not sequentially compact.

Coming back to earth: from Corollary 1.27 we deduce:

Corollary 1.28 (Multivariable Extreme Value Theorem). *Let X be a subset of $\mathbb{R}^N$ that is nonempty, closed and bounded, and let $f : X \to \mathbb{R}$ be a continuous function. Then f assumes its maximum and minimum values.*

Proof. By the previous corollary, $f(X)$ is a subset of $\mathbb{R}$ that is nonempty, closed and bounded. By Exercise 1.10 it follows that $\sup f(X)$ and $\inf f(X)$ both lie in $f(X)$, so $f(X)$ has a largest and smallest element. $\square$

Exercises.

Exercise 1.23.

a) *Show: a subset of $\mathbb{R}$ is an open ball if and only if it is a bounded open interval (a,b).*
b) *Show: a subset of $\mathbb{R}$ is a closed ball if and only if it is a bounded closed interval $[a,b]$.*

Exercise 1.24. *Let I be a nonempty set, and let $\{U_i\}_{i\in I}$ be an indexed family of open subsets of $\mathbb{R}^N$.*

a) *Show: $\bigcup_{i\in I} U_i$ is also an open subset of $\mathbb{R}^N$.*
b) *Show: if I is finite, then $\bigcap_{i\in I} U_i$ is also an open subset of $\mathbb{R}^N$.*
c) *Give an example in which I is infinite and $\bigcap_{i\in I} U_i$ is* not *an open subset of $\mathbb{R}^N$.*

Exercise 1.25. *Show: A subset $A \subseteq \mathbb{R}^N$ is closed if and only if its complement $\mathbb{R}^N \setminus A$ is open.*

Exercise 1.26. *Let I be a nonempty set, and let $\{A_i\}_{i\in I}$ be an indexed family of closed subsets of $\mathbb{R}^N$.*

a) *Show: $\bigcap_{i\in I} A_i$ is also a closed subset of $\mathbb{R}^N$.*
b) *Show: if I is finite, then $\bigcup_{i\in I} A_i$ is also a closed subset of $\mathbb{R}^N$.*

c) *Give an example in which I is infinite and $\bigcup_{i \in I} A_i$ is* not *an open subset of $\mathbb{R}^N$.*

(Comment: if you remember DeMorgan's Laws, you can immediately deduce this from Exercises 1.24 and 1.25. But if not, you can still solve this exercise directly, and even if you, you might want to try it that way as well.)

EXERCISE 1.27. *Let $X \subseteq \mathbb{R}^N$, and let $f : X \to R^M$. For $x \in X$, we may write $f(x)$ as $(f_1(x), \ldots, f_M(x))$; this defines functions $f_1, \ldots, f_M : X \to \mathbb{R}$.*

a) *For $1 \leq i \leq M$, let $\pi_i : \mathbb{R}^M \to R$ be the coordinate projections of Exercise 1.15. Show: for all $1 \leq i \leq M$, we have $f_i(x) = \pi_i \circ f$.*
b) *Let $\mathbf{x} \in X$. Show: f is continuous at $\mathbf{x}$ if and only if f_i is continuous at $\mathbf{x}$ for all $1 \leq i \leq M$.*

EXERCISE 1.28. *Show that the scalar multiplication operation $\alpha \cdot \mathbf{x} \mapsto \alpha x$ defines a continuous function $\mathbb{R} \times \mathbb{R}^N \to \mathbb{R}^N$.*

EXERCISE 1.29. *Consider the following functions $f, g : \mathbb{R} \to \mathbb{R}$:*

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

$$g(x) = \begin{cases} x & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ irrational} \end{cases}.$$

a) *Show: for all $c \in \mathbb{R}$, f is* not *continuous at c.*
b) *Show: g is continuous at 0 and at no other point of $\mathbb{R}$.*

EXERCISE 1.30. *For a subset $A \subseteq \mathbb{R}^N$, we define the* ***interior*** *A° to be the set of $\mathbf{x} \in A$ for which we have $B^\circ(\mathbf{x}, \delta) \subseteq A$ for some $\delta > 0$.*

a) *Show that A° is the largest open subset of A: that is, show (i) A° is an open subset of A and (ii) if U is an open subset of A then $U \subseteq A^\circ$.*
b) *Show: A is open $\iff A = A^\circ$.*

Problems.

PROBLEM 1.10. *Let $A \subseteq \mathbb{R}$ be a nonempty subset.*

a) *Suppose that A is bounded above. Show that the supremum $\sup(A)$ is a limit point of A.*
b) *Suppose that A is bounded below. Show that the infimum $\inf(A)$ is a limit point of A.*
c) *Deduce: if A is closed and bounded, then A has a maximum element (i.e., an element larger than any other element of A and a minimum element (i.e., an element smaller than any other element of A).*

PROBLEM 1.11. *For a sequence $\{\mathbf{x}_n\}$ in $\mathbb{R}^N$, let $\mathcal{L}(\mathbf{x}_n)$ be the set of all partial limits of the sequence.*

a) *Show: $\mathcal{L}(\mathbf{x}_n)$ is a closed subset of $\mathbb{R}^N$.*
b) *Suppose that the sequence $\{\mathbf{x}_n\}$ is injective (i.e., for all $m \neq n$ we have $\mathbf{x}_m \neq \mathbf{x}_n$). Let $X := \mathbf{x}_\bullet(\mathbb{Z}^+) = \{\mathbf{x}_n \mid n \in \mathbb{Z}^+\}$ be the set of terms of the sequence. Show that $\mathcal{L}(\mathbf{x}_n)$ is the set of accumulation points of X.*

PROBLEM 1.12. *Show: for every closed subset $X \subseteq \mathbb{R}^N$ there is a sequence $\{\mathbf{x}_n\}$ in $\mathbb{R}^N$ such that the set $\mathcal{L}(\mathbf{x}_n)$ of partial limits is X.*

PROBLEM 1.13. *Define* ***Thomae's function*** $T : \mathbb{R} \to \mathbb{R}$ *as follows:*
• *We put* $T(0) = 1$.
• *If* x *is a nonzero rational number, we may write* $x = \frac{m}{n}$ *for* $m \in \mathbb{Z}$ *and* $n \in \mathbb{Z}^+$ *with no common factor. Then we put* $T(\frac{m}{n}) = \frac{1}{n}$.
• *If* x *is an irrational number, we put* $T(x) = 0$.
Show: for all $x \in \mathbb{R}$*, the function* T *is continuous at* x *if and only if* x *is irrational.*

5. Uniform Continuity

Let $A \subseteq \mathbb{R}^N$. A function $f : A \to \mathbb{R}^M$ is **uniformly continuous** if for all $\epsilon > 0$ there is $\delta > 0$ such that for all $x, y \in A$, if $d(x, y) = ||x - y|| < \delta$, then $d(f(x), f(y)) = ||f(x) - f(y)|| < \epsilon$.

The point of this definition is that ordinary continuity applies to one point at a domain at a time, so for each fixed $\epsilon > 0$, the δ that works for one point may not work for another point. Uniform continuity means precisely that we may choose the same δ to work for all points at once. Thus uniformly continuous functions are continuous. The converse is not always true.

EXAMPLE 1.4. *Consider the continuous function* $f : \mathbb{R} \to \mathbb{R}$ *by* $f(x) = x^2$. *We claim that* f *is* not *uniformly continuous. Indeed, for a positive integer* n*, take* $x \in \mathbb{R}$ *and* $y = x + \delta$. *Then*

$$|f(x) - f(y)| = |(x + \delta)^2 - x^2| = |2x\delta + \delta^2|.$$

No matter how small δ *is, this quantity will still be large if* $|x|$ *is sufficiently large, so in fact for no* $\epsilon > 0$ *is there a* $\delta > 0$ *such that* $|x-y| < \delta$ *implies* $|f(x)-f(y)| < \epsilon$.

EXAMPLE 1.5. *Consider the continuous function* $f : (0, 1) \to \mathbb{R}$ *by* $f(x) = \frac{1}{x}$. *We claim that* f *is not uniformly continuous. Take* $x \in (0, 1)$ *and* $y = x + \delta$. *Then*

$$|f(x) - f(y)| = \left|\frac{1}{x} - \frac{1}{x + \delta}\right| = \left|\frac{\delta}{x(x + \delta)}\right|.$$

For each fixed δ*, as* $x \to 0$ *the above expression tends to* ∞*, so for no* $\epsilon > 0$ *is there a* $\delta > 0$ *such that* $|x - y| < \delta$ *implies* $|f(x) - f(y)| < \epsilon$.

In the first example the domain is closed but not bounded. In the second example the domain is bounded but not closed.

Here is a sequential characterization of uniform continuity:

PROPOSITION 1.29. *Let* $X \subseteq \mathbb{R}^N$ *be a subset, and let* $f : X \to \mathbb{R}^M$ *be a function. The following are equivalent:*

(i) *f is uniformly continuous.*
(ii) *For all pairs of sequences* $\{\mathbf{x}_n\}$ *and* $\{\mathbf{y}_n\}$ *in* X *such that* $d(\mathbf{x}_n, \mathbf{y}_n) \to 0$*, we have* $d(f(\mathbf{x}_n), f(\mathbf{y}_n)) \to 0$.

PROOF. First suppose that (i) fails: then there is some $\epsilon > 0$ such that for all $\delta > 0$ there are $\mathbf{x}_\delta, \mathbf{y}_\delta \in X$ such that $||\mathbf{x}_\delta - \mathbf{y}_\delta|| < \delta$ and $||f(\mathbf{x}_\delta) - f(\mathbf{y}_\delta)|| \geq \epsilon$. In particular, for each $n \in \mathbb{Z}^+$ this holds for $\delta = \frac{1}{n}$. Let's write $\mathbf{x}_n$ and $\mathbf{y}_n$ in place of $\mathbf{x}_{\frac{1}{n}}$ and $\mathbf{y}_{\frac{1}{n}}$: then for all $n \in \mathbb{Z}^+$ we have $d(\mathbf{x}_n, \mathbf{y}_n) < \frac{1}{n}$ and $d(f(\mathbf{x}_n), f(\mathbf{y}_n)) \geq \epsilon$. In particular $d(f(\mathbf{x}_n), f(\mathbf{y}_n))$ fails to converge to 0, so condition (ii) fails.

Now suppose that condition (ii) fails: then we have sequences $\{\mathbf{x}_n\}$ and $\{\mathbf{y}_n\}$

in X such that $\mathbf{x}_n - \mathbf{y}_n$ converges to 0 but $f(\mathbf{x}_n) - f(\mathbf{y}_n)$ fails to converge to 0. The latter means that there is some $\epsilon > 0$ and infinitely many positive integers n such that $||f(\mathbf{x}_n) - f(\mathbf{y}_n)|| \geq \epsilon$. This infinite set of positive integers defines subsequences $\mathbf{x}_{n_k}$ and $\mathbf{y}_{n_k}$. Because passing to a subsequence preserves convergence we have

$$\lim_{k\to\infty} \mathbf{x}_{n_k} - \mathbf{y}_{n_k} \to 0,$$

and now $||f(\mathbf{x}_{n_k}) - f(\mathbf{y}_{n_k})|| \geq \epsilon$ for all positive integers k. Beause $\mathbf{x}_{n_k} - \mathbf{y}_{n_k}$ converges to 0, for all $\delta > 0$ there is $k \in \mathbb{Z}^+$ such that $||\mathbf{x}_{n_k} - \mathbf{y}_{n_k}|| < \delta$, and we still have $||f(\mathbf{x}_{n_k}) - f(\mathbf{y}_{n_k})|| \geq \epsilon$. So condition (i) fails. □

LEMMA 1.30. *Let $\{\mathbf{x}_n\}$ and $\{\mathbf{y}_n\}$ be sequences in $\mathbb{R}^N$ with $d(\mathbf{x}_n, \mathbf{y}_n) \to 0$. If $\mathbf{x}_n \to L \in \mathbb{R}^N$, then also $\mathbf{y}_n \to L$.*

You are asked to prove Lemma 1.30 in Exercise 1.33.

THEOREM 1.31 (Uniform Continuity Theorem). *Let $X \subseteq \mathbb{R}^N$ be sequentially compact. Then every continuous function $f : X \to \mathbb{R}^M$ is uniformly continuous.*

PROOF. Seeking a contradiction, we suppose that f is *not* uniformly continuous. Then by Proposition 1.29 there are sequences $\{\mathbf{x}_n\}$ and $\{\mathbf{y}_n\}$ in X with $d(\mathbf{x}_n, \mathbf{y}_n) \to 0$ and $d(f(\mathbf{x}_n), f(\mathbf{y}_n))$ *not* converging to 0. As we saw in the proof of Proposition 1.29, this means that there is $\epsilon > 0$ and subsequences $\{\mathbf{x}_{n_k}\}$ and $\{\mathbf{y}_{n_k}\}$ such that $d(\mathbf{x}_{n_k}, \mathbf{y}_{n_k}) \geq \epsilon$ for all $k \in \mathbb{Z}^+$. Thus if X is *not* uniformly continuous then there are sequences $\{\mathbf{x}_n\}$ and $\{\mathbf{y}_n\}$ in X such that $d(\mathbf{x}_n, \mathbf{y}_n) \to 0$ and $\epsilon > 0$ such that $d(f(\mathbf{x}_n), f(\mathbf{y}_n)) \geq \epsilon$ for all $n \in \mathbb{Z}^+$.

We will use the sequential compactness of X to get a contradiction. Indeed, since X is sequential compact, there is a subsequence $\{\mathbf{x}_{n_k}\}$ that converges to some element of X, say L. Since $\mathbf{x}_n - \mathbf{y}_n \to 0$, also $\mathbf{x}_{n_k} - \mathbf{y}_{n_k} \to 0$, so by Lemma 1.30 the sequence $\{\mathbf{y}_{n_k}\}$ also converges to L. Because f is continuous, we have

$$f(\mathbf{x}_{n_k}) \to f(L) \text{ and } f(\mathbf{y}_{n_k}) \to f(L),$$

from which it follows that $f(\mathbf{x}_{n_k}) - f(\mathbf{y}_{n_k}) \to 0$ and thus for all sufficiently large k we have $d(\mathbf{x}_{n_k}, \mathbf{y}_{n_k}) < \epsilon$. Contradiction! □

So here is a question that is so much more than fair: **why** uniform continuity? What we have established up to this point is that uniform continuity is a variant of continuity that is in general subtly stronger, still has a sequential characterization, and that the two concepts coincide on closed, bounded subsets of Euclidean space. But...what's the point?

One thing that makes the study of theoretical mathematics challenging is that key definitions emerge after years (centuries, here) of work on specific problems. When the mathematics is presented however it is much more efficient to present the definitions first and the application later on. Indeed, later on in this course we will absolutely want to know that every continuous function $f : [a, b] \to \mathbb{R}$ is uniformly continuous: this will be the key to showing that every such function is Riemann integrable. However, I would like to show an application of uniform continuity *now*, so in the next section we consider the extension problem for continuous functions.

Exercises.

EXERCISE 1.31. *Use Proposition 1.29 to show that the function of Example 1.4 is not uniformly continuous.*

EXERCISE 1.32. *Use Proposition 1.29 to show that the function of Example 1.5 is not uniformly continuous.*

EXERCISE 1.33. *Let* $\{\mathbf{x}_n\}$ *and* $\{\mathbf{y}_n\}$ *be sequences in* $\mathbb{R}^N$ *with* $d(\mathbf{x}_n, \mathbf{y}_n) \to 0$. *If* $\mathbf{x}_n \to L \in \mathbb{R}^N$, *show that also* $\mathbf{y}_n \to L$.

EXERCISE 1.34. *State and prove an analogue of Exercise 1.27 for uniform continuity.*

EXERCISE 1.35. *Let* $M, N, P \in \mathbb{Z}^+$. *Let* $X \subseteq \mathbb{R}^N$ *and let* $Y \subseteq \mathbb{R}^M$. *Let* $f : X \to R^M$ *and let* $g : Y \to \mathbb{R}^P$ *be functions. Suppose that* $f(X) \subseteq Y$, *so that the composition* $g \circ f$ *is defined. Show: if* f *and* g *are both uniformly continuous, so is* $g \circ f$.

EXERCISE 1.36. *Let* $f : \mathbb{R} \to \mathbb{R}$ *be a polynomial function, say*

$$f(x) = a_n x^n + \ldots + a_1 x + a_0$$

with $n \in \mathbb{Z}^{\geq 0}$ *and* $a_n \neq 0$.

a) *Suppose that* f *is a linear function: i.e., that* $n \leq 1$. *Show:* f *is uniformly continuous.*
b) *Suppose that* $n \geq 2$. *Show:* f *is* not *uniformly continuous.*
c) *Can you generalize this to polynomials* $f : \mathbb{R}^N \to \mathbb{R}$?

Problems.

PROBLEM 1.14. *A function* $f : \mathbb{R} \to \mathbb{R}$ *is called* ***periodic*** *if there is* $a \in \mathbb{R} \setminus \{0\}$ *such that for all* $x \in \mathbb{R}$ *we have* $f(x+a) = f(x)$. *Show: a function that is continuous and periodic is uniformly continuous.*

PROBLEM 1.15. *We say that a function* $f : \mathbb{R}^N \to \mathbb{R}^M$ ***vanishes at infinity*** *if for all* $\epsilon > 0$ *there is* $R > 0$ *such that for all* $x \in \mathbb{R}^N$, *if* $||x|| > R$ *then* $||f(x)|| < \epsilon$. *Show: if* f *is continuous and vanishes at infinity, then* f *is uniformly continuous.*

PROBLEM 1.16. *Let* $X \subseteq \mathbb{R}^N$ *and let* $f : X \to \mathbb{R}^M$ *be uniformly continuous. Show: if* X *is bounded, then* $f(X)$ *is bounded.*

6. Accumulation Points and Continuous Extensions

Suppose X is a subset of $\mathbb{R}^N$ and $f : \mathbb{R}^N \to \mathbb{R}^M$ is a continuous function. It is natural to ask: can f be extended to a continuous function on all of $\mathbb{R}^N$?

EXAMPLE 1.6. *Let* $f : [a, b] \to \mathbb{R}$ *be continuous. Then* f *extends continuously to all of* $\mathbb{R}$: *indeed, we can put* $f(x) = f(a)$ *for all* $x < a$ *and* $f(x) = f(b)$ *for all* $x > b$. *This works!*

EXAMPLE 1.7. *The function* $f(x) = \frac{1}{x}$ *is a continuous function on* $(0, 1]$ *that does* not *extend continuously to all of* $\mathbb{R}$. *In the language of calculus, we would say that* $\lim_{x \to 0^+} f(x) = \infty$, *which prevents such an extension. This is correct, but here is an explanation using the language and concepts we have been developing:* f *is continuous at a point* $\mathbf{x}$ *if for all* $\epsilon > 0$ f *maps some ball* $B^\circ(\mathbf{x}, \delta)$ *into the ball* $B^\circ(f(\mathbf{x}), \epsilon)$. *In particular* f *must be bounded in some small ball around* $\mathbf{x}$. *Since*

$f((0,\delta)) = (\frac{1}{\delta}, \infty)$, no matter how we define f at 0, the function will be unbounded in any δ-ball around 0, so it does not have any continuous extension to $[0,1]$.

One moral one can extract from this is:

PROPOSITION 1.32. *Let $f : X \subseteq \mathbb{R}^N \to \mathbb{R}^M$. If f has a continuous extension to $\mathbb{R}^N$, then for all bounded subsets $Y \subseteq X$, the image $f(Y)$ is bounded.*

PROOF. It is enough to see that for $f : \mathbb{R}^N \to \mathbb{R}^M$, if $Y \subseteq \mathbb{R}^N$ is bounded then so is $f(Y)$. If Y is bounded, then it is contained in a closed ball B, which is closed and bounded, so $f(B)$ is closed and bounded, so $f(Y) \subseteq f(B)$ is bounded. □

This criterion is however not sufficient.

EXAMPLE 1.8. *The function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ defined by $f(x) = \sin(1/x)$ is not only bounded on every bounded subset; it is just bounded. Nevertheless, it does not extend continuously to $\mathbb{R}$, as is left as an exercise.*

Let us worry about extending a continuous function one point at a time. There are two cases of this; one is trivial, and the other is not.

EXAMPLE 1.9. *Let $f : (0,1) \to \mathbb{R}$ be a continuous function. Suppose we want to extend f to a continuous function on $(0,1) \cup \{2\}$. There is precisely no problem here: for any $L \in \mathbb{R}$ we can put $f(2) := L$, and the function $f : (0,1) \cup \{2\}$ will be continuous. Why? Because for any $\epsilon > 0$, take $\delta = 1$: we need to check that if $|x-2| < 1$ then $|f(x) - L| < \epsilon$. But the only $x \in (0,1) \cup \{2\}$ with $|x-2| < 1$ is $x = 2$ itself, and $|f(2) - L| = 0$.*

This example motivates the following definition: let A be a subset of $\mathbb{R}^N$. An **isolated point** of A is a point $L \in A$ such that for some $\delta > 0$ we have $B^\circ(L,\delta) \cap A = \{L\}$. In other words, a point of A is isolated if for some $\delta > 0$ the only point of A that is within δ of L is L itself.

PROPOSITION 1.33. *Let $A \subseteq \mathbb{R}^N$, and let $L \in A$ be an isolated point. Then* every *function $f : A \to \mathbb{R}^M$ is continuous at L.*

PROOF. This is the same argument as in Example 1.9: if $\delta > 0$ is such that $B^\circ(L,\delta) \cap A = \{L\}$, then for any $\epsilon > 0$, we have that for all $\mathbf{x} \in A$, $d(\mathbf{x}, L) < \delta \implies d(f(\mathbf{x}), f(L) < \epsilon$....because the only $\mathbf{x}$ that satisfies the first inequality is $\mathbf{x} = L$! □

In terms of the extension problem, this means: if L is an isolated point of A, then every continuous function f on $A \setminus \{L\}$ extends continuously to A, and we can do so by defining $f(L)$ to be whatever we want!

Okay, that was indeed a trivial case. Let's move on to the other case, which involves a variant of the notion of limit point that was alluded to before. By an **injective sequence** in a set X, we mean a sequence $\{\mathbf{x}_n\}$ in X for which the defining function $\mathbf{x}_\bullet : \mathbb{Z}^+ \to X$ is injective. In plainer language, an injective sequence is a sequence in which every term is a different element of X. Now for a subset $A \subseteq \mathbb{R}^N$, an **accumulation point** is a point $L \in \mathbb{R}^N$ for which there is an injective sequence $\{\mathbf{x}_n\}$ in A converging to L.

Compare with the definition of a limit point: the only difference is that we have

added the word "injective." Thus every accumulation point of A is a limit point of A. The converse is not true in general: for instance if A is finite there are no injective sequences in A, so A has no accumulation points, but as always, every element of A is a limit point of A. In general:

PROPOSITION 1.34. *Let $A \subseteq \mathbb{R}^N$, and let $L \in \mathbb{R}^N$ be a limit point of A. Then exactly one of the following holds:*

(i) *L is an accumulation point of A.*
(ii) *L is an isolated point of A.*

PROOF. Step 1: If L is an isolated point of A, then a sequence $\{\mathbf{x}_n\}$ in A converges to L if and only if we have $\mathbf{x}_n = L$ for all sufficiently large n. We leave this as an exercise (Exercise 1.39). From this it follows that if L is an isolated point of A then L is not the limit of any injective sequence in A, so L is not an accumulation point of A. Thus we have shown that conditions (i) and (ii) are mutually exclusive.
Step 2: Suppose that L is a limit point of A that is not an isolated point of A. This means that *either $L \notin A$ or $L \in A$* but for all $\delta > 0$ there is $\mathbf{x}_\delta \in A$ with $0 < ||\mathbf{x}_\delta - L|| < \delta$. In each of these two cases we will produce an injective sequence in A that converges to L.
Case 1: $L \notin A$. Because L is a limit point of A there is a sequence $\{\mathbf{x}_n\}$ in A that converges to A. The problem is that is may not be injective: i.e., terms may repeat. However, any element of $p \in A$ can show up only finitely many times in the sequence: indeed, since $p \in A$ and $L \notin A$, we have $p \neq L$, so $d = d(p, L) > 0$, and because the sequence converges to L, we have $d(\mathbf{x}_n, L) < d$ for all sufficiently large n. Therefore we can form a subsequence simply by omitting every term that is a repetition: i.e., for which the same element of A has alerady occurred earlier in the sequence. This builds an injective subsequence, which must still converge to L.
Case 2: $L \in A$. Because we have elements of A arbitrarily close to L but different from L, we can build a sequence the nth term of which has distance less than $\frac{1}{n}$ from L and is also closer than any previous term. In other words, let $\mathbf{x}_1$ be an element of $A \setminus \{L\}$ with $d(\mathbf{x}_1, L) < 1$. Let $\mathbf{x}_2$ be an element of $A \setminus \{L\}$ with $d(\mathbf{x}_2, L) < \min(\frac{1}{2}, d(\mathbf{x}_1, L))$. Let x_3 be an element of $A \setminus \{L\}$ with $d(\mathbf{x}_3, L) < \min(\frac{1}{3}, d(\mathbf{x}_2, L))$. And so forth. This gives an injective sequence in A converging to L. □

Note that an isolated point of A is necessarily a point of A, but an accumulation point of A may or may not be a point of A. For instance, every point of an open or closed ball is an accumulation point.

Since a set $X \subseteq \mathbb{R}^N$ is closed if it contains all its limit points, but every limit point of X is either an element of X or an accumulation point (again, both are possible!), it follows that a set is closed if and only if it contains its accumulation points.

So now let's consider the nontrivial case of the "one point extension problem": let $A \subseteq \mathbb{R}^N$, let $L \in \mathbb{R}^N \setminus A$ be an accumulation point of A, and let $f : A \to \mathbb{R}^M$ be a continuous function. The question is whether we can extend f to a continuous function on $A \cup \{L\}$. First we observe that there is at most one way to do this: indeed, suppose that $g : A \cup \{L\} \to \mathbb{R}^M$ is a continuous extension of f. Because L is an accumulation point of A, there is a sequence $\{\mathbf{x}_n\}$ in A such that $\mathbf{x}_n \to L$.

By Theorem 1.18 we have

$$g(L) = g(\lim_{n\to\infty} \mathbf{x}_n) = \lim_{n\to\infty} f(\mathbf{x}_n).$$

This tells us how to define $g(L)$, so its value must indeed be unique.

THEOREM 1.35. *Let $A \subseteq \mathbb{R}^N$, let $L \in \mathbb{R}^N \setminus A$ be an accumulation point of A, and let $f : A \to \mathbb{R}^M$ be a continuous function. Suppose there is some $r > 0$ such that*

$$f|_{B^\bullet(L,r)\cap A} : B^\bullet(L,r) \cap A \to \mathbb{R}^M$$

is uniformly continuous. Then f admits a continuous extension to $A \cup \{L\}$.

PROOF. Step 1: Above we assumed that the continuous extension g existed and gave a formula for it: namely, choose a sequence $\{\mathbf{x}_n\}$ in A such that $\mathbf{x}_n \to L$; then $g(L) = \lim_{n\to\infty} f(\mathbf{x}_n)$. So we want to *define*

$$f(L) := \lim_{n\to\infty} f(\mathbf{x}_n).$$

In fact, so as to make use of the assumed uniform continuity, we want the sequence $\mathbf{x}_n$ to lie in $B^\bullet(L,r)$. Because the original sequence converges to L, we can attain this just by removing finitely many terms, so let's do so. Now we need to show first that this limit actually exists and second that it does not depend upon the sequence $\{\mathbf{x}_n\}$ we chose.
Step 1a): Because we know that Cauchy sequences in $\mathbb{R}^M$ converge, it is enough to show that $\{f(\mathbf{x}_n)\}$ is Cauchy. For this, we know that the sequence $\{\mathbf{x}_n\}$ is convergent in $\mathbb{R}^N$, so it is Cauchy. Happily, it is easy to show that *uniformly* continuous maps send Cauchy sequences to Cauchy sequences: let $\epsilon > 0$. Because of the uniform continuity of f, there is $\delta > 0$ such that for all $y, z \in B^\bullet(L,r) \cap A$, we have $d(y,z) < \delta \implies d(f(y), f(z)) \leq \epsilon$. Since $\{\mathbf{x}_n\}$ lies in $B^\bullet(L,r)$ and is Cauchy, there is $K \in \mathbb{Z}^+$ such that if $m, n \geq K$ then $d(\mathbf{x}_m, \mathbf{x}_n) \leq \delta$, and thus

$$\forall m, n \geq K,\ d(f(\mathbf{x}_m), f(\mathbf{x}_n)) < \epsilon.$$

This shows that the sequence $f(\mathbf{x}_n)$ converges.
Step 1b): Let $\{\mathbf{y}_n\}$ be another sequence in $B^\bullet(L,r) \cap A$ such that $\mathbf{y}_n \to L$. Then $d(\mathbf{x}_n, \mathbf{y}_n) \to 0$, so by Proposition 1.29, we have $d(f(\mathbf{x}_n), f(\mathbf{y}_n)) \to 0$. Since both sequences are convergent, it follows from Lemma 1.30 their limits are equal.
Step 2: It remains to show that our extended function is continuous at L. But in fact this follows from our Sequential Characterization of Uniform Continuity, since we have just shown that if if $\mathbf{x}_n \to L$ then $f(\mathbf{x}_n) \to f(L)$. Strictly speaking, we showed this only for sequences each of whose terms lie in $B^\bullet(L,r) \cap A$, but again any sequence that converges to L becomes such a sequence after removing finitely many terms; so any such sequence converges after removing finitely many of its terms...so any such sequence converges. □

Let me quickly discuss some further developments of these ideas.

For a subset A of $\mathbb{R}^N$, we can define its **closure** $\overline{A}$ to be A together with all of its limit points (equivalently, with all of its accumulation points). As the name implies, $\overline{A}$ is then a closed set (this is not completely obvious: it comes down to showing a limit point of limit points of A is still a limit point of A). In fact $\overline{A}$ is the smallest closed set containing A. It follows from our discussion that every continuous function had *at most one* continuous extension to $\overline{A}$. Such a continuous

extension need not exist, but it will if f is *uniformly continuous* on A. But in fact this condition is a little too strong, and the precise result is the following.

THEOREM 1.36. *Let $A \subseteq \mathbb{R}^N$, and let $f : A \to \mathbb{R}^M$ be continuous. The following are equivalent:*

(i) *f admits a continuous extension to $\overline{A}$.*
(ii) *The restriction of f to each bounded subset of A is uniformly continuous.*

We are not so terribly far away from a proof of this important result; it will be developed in some exercises.

A subset $X \subset \mathbb{R}^N$ is called **dense** if its closure is all of $\mathbb{R}^N$. This means: for every $\mathbf{y} \in \mathbb{R}^N$ and every $\epsilon > 0$, there is $\mathbf{x} \in X$ with $d(\mathbf{x}, \mathbf{y}) < \epsilon$. For instance $\mathbb{Q}$ is dense in $\mathbb{R}$. Theorem 1.36 therefore shows that if you have a continuous function on a dense subset of $\mathbb{R}^N$ then it extends continuously to all of $\mathbb{R}^N$ if and only if it is uniformly continuous on each bounded subset. As an example, consider an exponential function a^x. If you think about it, we can make good sense of a^x using the methods of precalculus when x is any rational number, but what does $a^{\sqrt{2}}$ mean? In order to make sense of it we need to use some limiting process. One way to define a^x as a function on all of $\mathbb{R}$ is to show that $a^x : \mathbb{Q} \to \mathbb{R}$ is uniformly continuous on bounded subsets. (It is not uniformly continuous on all of $\mathbb{Q}$.)

What if A is not dense, so $\overline{A} \subsetneq R^N$ is closed? It turns out that if $X \subseteq \mathbb{R}^N$ is a closed subset and $f : X \to \mathbb{R}^M$ is continuous, then there is always a continuous extension of f to all of $\mathbb{R}^N$: in fact there are always lots and lots of such extensions. This is a special case of an important result called the **Tietze Extension Theorem**, which you might learn about in Math 4200: see [**GT**, Theorem 2.89].

Before we depart this topic, let us observe that we have essentially rediscovered the notion of **limit**. Namely, let $X \subseteq \mathbb{R}^N$, let $\mathbf{c}$ be an accumulation point of X, and let $f : X \setminus \{\mathbf{c}\} \to \mathbb{R}^M$ be a function. Then we define

$$\lim_{\mathbf{x} \to \mathbf{c}} f(x) = L$$

to mean: if we extend f to X by setting $f(\mathbf{c}) := L$, then f is continuous at $\mathbf{c}$. Again, the value L is then the common value $\lim_{n\to\infty} f(\mathbf{x}_n)$ for all sequences $\{\mathbf{x}_n\}$ in $A \setminus \{L\}$ that converge to c, so it is uniquely determined, if it exists. It is immediate to see that the limit is L iff: forall $\epsilon > 0$ there is $\delta > 0$ such that for all $\mathbf{x} \in A \setminus \{\mathbf{c}\}$, we have $d(\mathbf{x}, \mathbf{c}) < \delta \implies d(f(\mathbf{x}), L) < \epsilon$.

Exercises.

EXERCISE 1.37. *Let $A \subseteq \mathbb{R}^N$ and let $\mathbf{c}$ be an isolated point of A. Show: every function $f : A \to \mathbb{R}^M$ is continuous at $\mathbf{c}$.*

EXERCISE 1.38. *Show that the function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ given by $f(x) = \sin(\frac{1}{x})$ has no continuous extension to 0.*

EXERCISE 1.39. *Let $A \subseteq \mathbb{R}^N$, and let $L \in A$ be an isolated point. Let $\{\mathbf{x}_n\}$ be a sequence in A. Show that $\mathbf{x}_n \to L$ if and only if there is $K \in \mathbb{Z}^+$ such that $\mathbf{x}_n = L$ for all $n > K$.*

EXERCISE 1.40.

a) *Use the Bolzano-Weierstrass Theorem to show that if $A \subseteq \mathbb{R}^N$ is infinite and bounded, then A has an accumulation point in $\mathbb{R}^N$.*
b) *Suppose that every infinite bounded subset of $\mathbb{R}^N$ has an accumulation point in $\mathbb{R}^N$. Deduce the Bolzano-Weierstrass Theorem in $\mathbb{R}^N$.*

Exercise 1.41. *Let $X \subseteq \mathbb{R}^N$. Show that the following are equivalent:*
(i) *X has an accumulation point in $\mathbb{R}^N$.*
(ii) *There is some bounded subset $B \subseteq \mathbb{R}^N$ such that $X \cap B$ is infinite.*

Exercise 1.42. *Let $A \subseteq \mathbb{R}^N$ be bounded. Let $f : A \to \mathbb{R}^M$ be continuous.*
a) *Show that the following are equivalent:*
   (i) *f is uniformly continuous.*
   (ii) *f admits a continuous extension to $\overline{A}$.*
b) *Show that under the equivalent condiitons of part a), the continuous extension of f to $g : \overline{A} \to \mathbb{R}^M$ is unique and uniformly continuous.*

Exercise 1.43. *Let $A \subseteq \mathbb{R}^N$, and let $f : A \to \mathbb{R}^M$.*
a) *Show that f is continuous if and only if its restriction to each bounded subset of f is continuous.*
b) *Show: f admits a continuous extension to $\overline{A}$ if and only if its restriction to each bounded subset of A is uniformly continuous.*

The remaining exercises make use of the following definitions: let $X \subseteq \mathbb{R}^N$. We say that X is **discrete** if every point of X is an isolated point. We say that X is **uniformly discrete** if there is $\delta > 0$ such that for all $x_1, x_2 \in X$, if $||x_1 - x_2|| < \delta$ then $x_1 = x_2$.

Exercise 1.44.
a) *Show: if X is uniformly discrete, then X is discrete.*
b) *Show $\{\frac{1}{n} \mid n \in \mathbb{Z}^+\}$ is a subset of $\mathbb{R}$ that is discrete, not uniformly discrete and not closed.*
c) *Show: if X is uniformly discrete, then X is closed.*
d) *Find a closed subset $X \subseteq \mathbb{R}$ that is discrete but not uniformly discrete.*

Problems.

Problem 1.17. *Let $A \subseteq \mathbb{R}^N$. Let $\overline{A}$ be the union of A and the accumulation points of A.*
a) *Show: $\overline{A}$ is closed.*
b) *Show: $\overline{A}$ is the intersection of all closed subsets of $\mathbb{R}^N$ containing A.*
c) *Show: A is bounded if and only if $\overline{A}$ is sequentially compact.*

Problem 1.18. *Let $A \subseteq \mathbb{R}^N$. The **boundary** ∂A of A is the set of all $\mathbf{x} \in \mathbb{R}^N$ such that for all $\delta > 0$ the open ball $B^\circ(\mathbf{x}, \delta)$ contains a point of A and also contains a point of $\mathbb{R}^N \setminus A$.*
a) *Compute (with proof) the boundaries of open and closed balls.*
b) *Show: $\partial A = \overline{A} \cap \overline{\mathbb{R}^N \setminus A}$. Deduce: ∂A is closed.*
c) *Show: $(\partial A) \cap A = A \setminus A^\circ$.*
d) *Show: $\overline{A} = A \cup \partial A$.*

Problem 1.19. *Let $A \subseteq \mathbb{R}^N$, and let $f : A \to \mathbb{R}^M$ be a function. Show that the following are equivalent:*

(i) *The function f is continuous.*
(ii) *For all open subsets V of $\mathbb{R}^M$, the inverse image*

$$f^{-1}(V) := \{\mathbf{x} \in A \mid f(\mathbf{x}) \in V\}$$

is of the form $U \cap A$ for some open subset U of $\mathbb{R}^N$.
(iii) *For all closed subsets V of $\mathbb{R}^M$, the inverse image $f^{-1}(V)$ is of the form $C \cap A$ for some closed subset C of $\mathbb{R}^N$.*

PROBLEM 1.20. *Let $X \to \mathbb{R}^N$ be a subset.*

a) *Show that the following are equivalent:*
   (i) *X is discrete.*
   (ii) *Every function $f : X \to \mathbb{R}^M$ is continuous.*
   (iii) *Every function $f : X \to \mathbb{R}$ is continuous.*
b) *Show that the following are equivalent:*
   (i) *X is uniformly discrete.*
   (ii) *Every function $f : X \to \mathbb{R}^M$ is uniformly continuous.*
   (iii) *Every function $f : X \to \mathbb{R}$ is uniformly continuous.*

PROBLEM 1.21. *Let $X \subseteq \mathbb{R}^N$. Show that the following are equivalent:*

(i) *Every continuous function $f : X \to \mathbb{R}^M$ is uniformly continuous.*
(ii) *Every continuous function $f : X \to \mathbb{R}$ is uniformly continuous.*
(iii) *X is either sequentially compact or uniformly discrete.*

7. Functional Limits

Let $X \subseteq \mathbb{R}^N$ be a nonempty subset, and let $c \in X$ be a nonisolated point. For a function $f : X \setminus \{c\} \to \mathbb{R}^M$, recall that we say that $\lim_{x\to c} f(x) = L$ if definining $f(c) := L$ makes f continuous at c. Spelling out, this means: for all $\epsilon > 0$, there is $\delta > 0$ such that for all $x \in X$, if $0 < ||x - c|| < \delta$ then $||f(x) - L|| < \epsilon$. If the limit exists, then its value is unique.

The following is a variation on the fact that compositions of continuous functions are continuous.[2]

PROPOSITION 1.37. *Let $X \subseteq \mathbb{R}^N$ and $Y \subseteq \mathbb{R}^M$, and let $f : X \setminus \{c\} \to \mathbb{R}^M$ and $g : Y \to \mathbb{R}^P$. Suppose that $f(X) \subseteq Y$. Let $c \in X$ be a nonisolated point. Then: if $\lim_{x\to c} f(x) = L$ and $\lim_{y\to L} g(y) = M$. Then*

$$\lim_{x\to c} g(f(x)) = M = g(\lim_{x\to c} f(x)).$$

PROOF. Define f at c by $f(c) := L$; then f is continuous at c. There is a sequence $\{x_n\}$ in X converging to c, hence $f(x_n)$ converges to $f(c) = L$, so L is a limit point of Y. If $L \notin Y$, we put $g(L) := M$; if $L \in Y$, then we *redefine* $g(L) := M$. Either way this makes g continuous at M. By Proposition 1.23, the composition $g \circ f$ is then continuous at c, so $\lim_{x\to c} g(f(x)) = g(f(c)) = g(L) = M$. □

[2]The annoying complicatedness of the statement is again a sign that continuity is the fundamental concept on which limits should be based, not the other way around.

Exercises.

EXERCISE 1.45 (Squeeze Theorem). *Let $X \subseteq \mathbb{R}^N$, and let $c \in X^\circ$. Let $f : X \setminus \{c\} \to \mathbb{R}$ be a function. Let $\delta > 0$ be such that $B^\circ(c, \delta) \subseteq X$, and suppose there are functions*

$$m, M : B^\circ(c, \delta) \setminus \{c\} \to \mathbb{R}$$

such that

$$\forall x \in B^\circ(c, \delta) \setminus \{c\},\ m(x) \leq f(x) \leq M(x)$$

and

$$\lim_{x \to c} m(x) = L = \lim_{x \to c} M(x).$$

Show: $\lim_{x \to c} f(x) = L$.

In the above statement of the Squeeze Theorem, it is in fact not critical that c be an interior point of X: we could have worked with any accumulation point c of X and in place of $B^\circ(c, \delta)$ used $B^\circ(c, \delta) \cap X$. We just wanted a relatively clean statement.

EXERCISE 1.46. *Let $X \subseteq \mathbb{R}$, and let a be an accumulation point of X. We say that the* ***right-handed limit*** $\lim_{x \to a^+} f(x) = L$ *exists if a is still an accumulation point of $X \cap [a, \infty)$ and upon restricting the domain from X to $X \cap [a, \infty)$, the limit exists and is equal to L. We say that the* ***left-handed limit*** $\lim_{x \to a^-} f(x) = L$ *exists if a is still an accumulation point of $X \cap (-\infty, a]$ and upon restricting the domain from X to $X \cap (-\infty, a]$), the limit exists and is equal to L. Show:* $\lim_{x \to a} f(x)$ *exists if and only if* $\lim_{x \to a^+} f(x)$ *and* $\lim_{x \to a^-} f(x)$ *both exists and are equal, in which case the common value is* $\lim_{x \to a} f(x)$.

Problems.

PROBLEM 1.22 (Monotone Functions Have Simple Disontinuities). *Let $I \subseteq \mathbb{R}$ be an interval, and let $c \in I$. We say that f has a* ***simple discontinuity*** *at c if each of the one-sided limits* $\lim_{x \to c^+} f(x)$ *and* $\lim_{x \to c^-} f(x)$ *exist but f is* not *continuous at c: this means that either* $\lim_{x \to c^+} \neq f(c)$ *or* $\lim_{x \to c^-} f(c)$ *(or both).*

Let $f : I \to \mathbb{R}$ be a monotone function: that is, f is increasing or decreasing.

a) *Suppose c is not an endpoint of I. Show that* $\lim_{x \to c^+} f(x)$ *and* $\lim_{x \to c^-} f(x)$ *both exist. Indeed, if f is increasing, show:*

$$\lim_{x \to c^+} f(x) = \inf\{f(x) \mid x > c\} \text{ and } \lim_{x \to c^-} f(x) = \sup\{f(x) \mid x < c\},$$

while if f is decreasing, show:

$$\lim_{x \to c^+} f(x) = \sup\{f(x) \mid x > c\} \text{ and } \lim_{x \to c^-} f(x) = \inf\{f(x) \mid x < c\}.$$

b) *Suppose that I has a left endpoint a. Show:* $\lim_{x \to a^+} f(x)$ *exists.*

c) *Suppose that I has a right endpoint b. Show:* $\lim_{x \to a^-} f(x)$ *exists.*

CHAPTER 2

The Riemann Integral

1. Abstract Integrals and the Fundamental Theorem of Calculus

We now begin our study of "the integral calculus." The basic idea here is as follows: for a function $f : [a, b] \to \mathbb{R}$ we wish to associate a real number $\int_a^b f$, the **definite integral**. When f is non-negative, our *intutition* is that $\int_a^b f$ should represent the area under the curve $y = f(x)$ — more precisely the area of the region bounded above by $y = f(x)$, below by $y = 0$, on the left by $x = a$ and on the right by $x = b$. For general functions f, the integral $\int_a^b f$ is supposed to represent the **signed area** — more on this later.

The above sentiment is roughly analogous to the intuition that a continuous function is one whose graph is a "nice, unbroken" curve. Namely, it is a geometric idea that must be analytically formalized, and whose analytic formalization requires further ideas. The above gives a precise description of a subset of the plane associated to $f : [a, b] \to [0, \infty)$, namely the set

$$S_f := \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b \text{ and } 0 \leq y \leq f(x)\}.$$

It is easy to see that S_f is bounded if and only if f is bounded (Exercise 2.1). So if we knew how to assign an area to every bounded subset of $\mathbb{R}^2$, then this would work as a definition. The issue is that this "assigning areas" problem is itself a very challenging one: the part of mathematics that deals with this in a satisfactory way is called **measure theory**, which is part of *graduate* real analysis.

So our main task here is to define a new limiting process telling us how to assign the real number $\int_a^b f$ to the function $f : [a, b] \to \mathbb{R}$. Just as for all previous limiting processes (limits of sequences and series, functional limits at a point, continuity, differentiability) the limit need not exist for all functions, and indeed there are some functions $f : [a, b] \to \mathbb{R}$ for which $\int_a^b f$ is not defined. (This is true both for the particular limiting processes that we will study but also, for certain choices of f, for *any* reasonable limiting process.) Just as we call a function differentiable if the limiting process defining the limit exists, we will call a function **integrable** if it lies in the class of functions for which the limiting process works to assign a number $\int_a^b f$. (This is not yet a definition since we haven't said what the process is!)

Before we plunge into the details of a particular limiting process, it will be helpful to consider some properties that we want our integral to study. If the integral is supposed to be a signed area, it should surely satisfy the following properties:

(I1) If $f = C$ is constant, then $\int_a^b C = C(b - a)$.

Indeed, when $C > 0$ the set S_f is just a rectangle with base length $b - a$ and with height C: that's an area that we know. When $C < 0$, the set S_f is a rectangle with the same description, but now it is bounded *above* by the x-axis and *below* by $y = C$, so our convention is that this counts as "negative area." When $C = 0$, the set S_f is just the line segment $[a, b]$, which indeed should have area 0.

Comment: Until further notice, we will "explain" our properties only for non-negative functions f. This case is simpler and easier to explain. Once we sufficiently develop the theory we will be able to understand how to recover the general case from this (essentially we add a sufficiently large constant to make f non-negative).

(I2) If $f_1, f_2 : [a, b] \to \mathbb{R}$ satisfy $f_1 \leq f_2$ — that is, for all $x \in [a, b]$ we have $f_1(x) \leq f_2(x)$ — then $\int_a^b f_1 \leq \int_a^b f_2$.

Under our running "explanatory assumption" that f_1 and f_2 are non-negative, if $f_1 \leq f_2$ then S_{f_1} is a subset of S_{f_2}, and certainly the area of a subset should be less than or equal to the area of the entire set.

(I3) If $f : [a, b] \to \mathbb{R}$ and $a \leq c \leq b$, then $\int_a^b f = \int_a^c f + \int_c^b f$.

To explain this, again under the additional assumption that $f \geq 0$, we will add to our notation by writing $S_{f,[a,b]}$ for what we above wrote as S_f, taking the interval $[a, b]$ as known. Then we have

$$S_{f,[a,b]} = S_{f,[a,c]} \cup S_{f,[c,b]}$$

and $S_{f,[a,c]} \cap S_{f,[c,b]}$ is just the vertical line segment from $(c, 0)$ to $(c, f(c))$, which should have area 0. The way we think areas should work is that the area of the union should be the sum of the areas minus the area of the intersection, so this explains (I3).

Again, let me emphasize: I am not *proving* (I1), (I2) and (I3). I couldn't possibly do that until I tell you what $\int_a^b f$ means. I am just writing down some desired consequences of any reasonable definition of $\int_a^b f$. Or, if you like, we are writing down **axioms** that our integration process should satisfy.

In fact, I do like – I find the axiomatic approach to be a clean way to come at this problem. To make it work completely, I want to add one more ingredient: what is the "domain." Namely, suppose we are given a subset $\mathcal{R}[a, b]$ of the set of all functions $f : [a, b] \to \mathbb{R}$ that we call the **integrable functions**.

(There is a little fine print here: first of all, we actually mean to define $\mathcal{R}[a, b]$ for each pair of real numbers (a, b) with $a \leq b$. Second of all, if $a \leq c \leq b$ and $f \in \mathcal{R}[a, b]$, we want $f|_{[a,c]} : [a, c] \to \mathbb{R}$ to lie in $\mathcal{R}[a, c]$ and $f|_{[c,b]} : [c, b] \to \mathbb{R}$ to lie in $\mathcal{R}[c, b]$. This is necessary to make sense of Axiom (I3), for instance.)

Having done this, an **integral** is, for each $a \leq b$, a function

$$\int : \mathcal{R}[a, b] \to \mathbb{R},\ f \mapsto \int_a^b f$$

that should satisfy the above axioms. This means that we want constant functions to be integrable and that we require (I2) and (I3) to hold for functions $f_1, f_2, f \in \mathcal{R}[a,b]$.

However, we need one more thing in order to be sure we are doing something nontrivial. That is...we need to say something about $\mathcal{R}[a,b]$, the set of integrable functions. The only functions that our axioms ensure lie in $\mathcal{R}[a,b]$ are the constant functions. So we could take $\mathcal{R}[a,b]$ to consist of constant functions and then we are only talking about signed areas of rectangles. One step away would be to take $\mathcal{R}[a,b]$ to be all polynomial functions. In this case, verfiying the axioms corresponds roughly to the amount of understanding posssessed by a B-level calculus student: we just need to know to reverse the power rule for differentiation.

So let us sneak in one more axiom to ensure that there is some content here:

(I0) For all real numbers $a < b$ we have that:
(I0(a)) Every continuous function $f : [a,b] \to \mathbb{R}$ lies in $\mathcal{R}[a,b]$; and
(I0(b)) Every function $f \in \mathcal{R}[a,b]$ is bounded.

Concerning this last axiom: we start with part b) and *do not* give a justification but rather admit that it is there to simplify the situation. However we observe that parts a) and b) are compatible because of the Extreme Value Theorem: every continuous function is bounded. Therefore because of axiom (IO) we can — at the least – integrate every continuous function $f : [a,b] \to \mathbb{R}$. Such an integral is guaranteed to have real content: because of the close connection to the area problem, such an integral gives a rigorous mathematical meaning to "the area under a non-negative continuous curve $y = f(x)$."

Now something remarkable happens: if we *assume* that we have an integral $\int : \mathcal{R}[a,b] \to \mathbb{R}$ satisfying axioms (I0) through (I3), then without knowing anything about how this function is actually defined, we can use it to prove the Fundamental Theorem of Calculus!

THEOREM 2.1 (Fundamental Theorem of Calculus). *Let $\int : \mathcal{R}[a,b] \to \mathbb{R}$ satisfy (IO), (I1), (I2) and (I3). Let $f \in \mathcal{R}[a,b]$. For $x \in [a,b]$, we define*

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

Then:

a) *The function $\mathcal{F} : [a,b] \to \mathbb{R}$ is continuous.*
b) *If f is continuous at c, then $\mathcal{F}$ is differentiable at c, and $\mathcal{F}'(c) = f(c)$.*
c) *If f is continuous and $F : [a,b] \to \mathbb{R}$ is any antiderivative of f – i.e., $F' = f$ – then*

$$\int_a^b f = F(b) - F(a).$$

PROOF. a) By (I0(b)), there is $M > 0$ such that $|f(x)| \leq M$ for all $x \in [a,b]$. Let $\epsilon > 0$, and take $\delta := \frac{\epsilon}{M}$. For any $a \leq c \leq d \leq b$, because $-M \leq f \leq M$,

applying (I2) and (I1) we get

$$-M(d-c) = \int_c^d (-M) \leq \int_c^d f \leq \int_c^d M = M(d-c),$$

so using (I3) we get

$$\left|\mathcal{F}(d) - \mathcal{F}(c)\right| = \left|\int_a^d f - \int_a^c f\right| = \left|\int_c^d f\right| \leq M(d-c), \tag{5}$$

which shows that $\mathcal{F}$ is uniformly continuous with $\delta = \frac{\epsilon}{M}$.
b) Since f is continuous at c, for all $\epsilon > 0$, there is δ such that $|x-c| < \delta$ implies

$$f(c) - \epsilon < f(x) < f(c) + \epsilon.$$

Thus:

$$f(c) - \epsilon = \frac{\int_c^x (f(c)-\epsilon)}{x-c} \leq \frac{\int_c^x f}{x-c} \leq \frac{\int_c^x (f(c)+\epsilon)}{x-c} = f(c) + \epsilon,$$

which we may rewrite as

$$\left|\frac{\mathcal{F}(x) - \mathcal{F}(c)}{x-c} - f(c)\right| = \left|\frac{\int_c^x f}{x-c} - f(c)\right| \leq \epsilon,$$

which shows that

$$\mathcal{F}'(c) = \lim_{x \to c} \frac{\mathcal{F}(x) - \mathcal{F}(c)}{x-c} = f(c).$$

c) Suppose f is continuous. By part b), we know that $\mathcal{F}(x) = \int_a^x f$ is an antiderivative of f. By Exercise 4.4 we know that antiderivatives are unique up to the addition of a constant, which means that if F is any antiderivative of f there is $C \in \mathbb{R}$ such that

$$\forall x \in [a,b],\ F(x) = \mathcal{F}(x) + C,$$

and thus,

$$\begin{aligned} F(b) - F(a) &= (\mathcal{F}(b) + C) - (\mathcal{F}(a) + C) \\ &= \mathcal{F}(b) - \mathcal{F}(a) = \int_a^b f - \int_a^a f = \int_a^b f; \end{aligned}$$

above we used Exercise 2.2 to get $\int_a^a f = 0$. □

We now have several important remarks to make.

First, as discussed above, any integral $\int : \mathcal{R}[a,b] \to \mathbb{R}$ restricts to an integral $\int : \mathcal{C}[a,b] \to \mathbb{R}$ on the set of all continuous functions $f : [a,b] \to \mathbb{R}$. But part c) of the Fundamental Theorem of Calculus tells us that in this case there is no need for axiomatics: the integral of any continuous function is *necessarily* given as $F(b) - F(a)$ for any antiderivative F of f. In other words, the function $\int : \mathcal{C}[a,b] \to \mathbb{R}$ is **unique**.

Second: I must observe that the proof of Theorem 2.1 was...quite easy. Admittedly the statement was a bit technical, but the proof of each part took only a few lines. Our proofs that our fancy-looking function $\mathcal{F}$ is always continuous and is differentiable when f is continuous each came out right away: earlier in our course we worked harder to prove the continuity/differentiability of very *specific* functions.

Why is the proof of the Fundamental Theorem of Calculus so easy? This is a question I thought a lot about the first time I taught undergraduate real analysis, in

2004 at McGill University. The proof of FTC is usually given in freshman calculus courses, but the theory of the Riemann integral is much more intricate (um, wait for it; you'll see). How is it possible that the theory is hard but its main theorem is easy?

The answer is that as we've stated it, the Fundamental Theorem of Calculus is *not* the crux that we might think it is. Do you see why? The answer is that our statement of the Fundamental Theorem *assumes* that we have an integral satisfying the axioms (I0) through (I3) and defined on the class of continuous functions (so (I0) is satisfied). After we prove the theorem, it turns out that on the class of continuous functions, this integral is *unique*. But how do we know that the integral exists? Answer: we don't, yet. That's where the real work lies.

Third: Theorem 2.1 has the following very important consequence:

COROLLARY 2.2. *Each continuous function $f : [a, b] \to \mathbb{R}$ has an antiderivative.*

Indeed, the Fundamental Theorem supplies us with the particular antiderivative $\mathcal{F}(x) = \int_a^x f$. (I emphasize that at the moment we know this conditionally on the assumption that the integral exists.) Once again we know, as a consequence of the Mean Value Theorem, that antiderivatives are unique up to an additive constant. As we saw in the proof, we have $\mathcal{F}(a) = \int_a^a f = 0$, so that tells us which antiderivative we're getting: the unique one that is 0 at the left endpoint.

It may be interesting to ask *how much* of the content of the Fundamental Theorem of Calculus is carried by Corollary 2.2: that is, suppose that we know, somehow, that every continuous function has an antiderivative. Can we then use this to show the existence of an integral on $\mathcal{C}[a, b]$? The answer is **yes**: if F is antiderivative of f, then you can show directly that $\int_a^b f := F(b) - F(a)$ defines an integral $\int : \mathcal{C}[a, b] \to \mathbb{R}$. This is an amusing exercise: Exercise 2.1. On the other hand, although there are several ways to go about constructing this integral $\int : \mathcal{C}[a, b] \to \mathbb{R}$ that we have been talking about, I believe that I do not know any way to prove Corollary 2.2 that does not involve constructing the integral in some way and then differentiating $\int_a^x f$ to get $f(x)$.

Let me now give a small preview of what's coming next: we will define a certain **process** that can be applied to any function $f : [a, b] \to \mathbb{R}$. This process returns two different extended real numbers – i.e., either real numbers, ∞ of $-\infty$. These are called the **upper Darboux integral** $\overline{\int}_a^b f$ and the **lower Darboux integral** $\underline{\int}_a^b f$. It will turn out that in all cases we have

$$\underline{\int}_a^b f \leq \overline{\int}_a^b f.$$

We say that the function f is **Darboux integrable** if the two are equal and the common value is a real number (and not $\pm\infty$).

We will study the Darboux integration process and show that it satisfies all our axioms: that is, if we define $\mathcal{R}_D[a, b]$ to be the set of Darboux integrable functions, then these functions satisfy (I1), (I2), (I3), and most importantly, (I0): every Darboux integrable function is bounded (indeed boundedness is equivalent to the upper

and lower integrals both being finite) and every continuous function $f : [a, b] \to \mathbb{R}$. It is this last statement that carries most of the content of the Fundamental Theorem of Calculus. We will also show some further useful properties of the Darboux integral: for instance, we will see that $\mathcal{R}_D[a, b]$ is a vector space over the real numbers and the integral $\int : \mathcal{R}_D[a, b] \to \mathbb{R}$ is a linear map.

At this point, we will know that $\mathcal{R}_D[a, b]$ contains all the continuous functions, and it will not be hard to see that it contains many other functions as well – e.g. all bounded functions that are either monotone or have finitely many discontinuities. So it is natural to ask: can we determine exactly which functions are Darboux integrable?

Leaving that question hang in the air for now, here is a very different question: why have we not said "Riemann" yet? After all, in calculus one speaks of the Riemann integral and after all that is the title of this chapter. Well, what we called the "Darboux integral" above is what many would call the Rieman integral. However we have a distinction to make: Riemann himself defined a different *process* from Darboux's: in other words, Riemann's actual technical definition of the limit is different from Darboux's. Rather we should say that Darboux's definition is different from Riemann's, since Riemann's came first: Darboux's is actually easier to understand and easier to work with in many respects. The main advantage of Riemann's definition is that it is indeed a (rather complicated!) limit of **Riemann sums**, which means that certain sequential limits can be evaluated by interpreting them as Riemann sums of a Riemann integrable function.

What is the relationship between the integrals of Darboux and Riemann? Although their descriptions are different, we have already shown that as functions $\mathcal{C}[a, b] \to \mathbb{R}$ they must be equal, i.e., the real number $\int_a^b$ assigned to each continuous $f : [a, b] \to \mathbb{R}$ must be the same, because both satisfy the axioms and there is a unique integral on the continuous functions satisfying the axioms. In fact their relationship is closer still: if we let $\mathcal{R}_R[a, b]$ denote the set of Riemann integrable functions, then in fact

$$\mathcal{R}_D[a, b] = \mathcal{R}_R[a, b]$$

— that is, a function is Riemann integrable if and only if it is Darboux integrable – and moreover when a function $f : [a, b] \to \mathbb{R}$ is integrable according to either definition the assigned values $\int_a^b f$ agree. So at the end of the day, although Riemann and Darboux are different *processes*, they yield exactly the same *integral*. In other words, they are ultimately two different descriptions of the same thing.

Exercises.

Exercise 2.1. *Let $f : [a, b] \to [0, \infty)$ be a function. Show that the subset*

$$S_f := \{(x, y) \in \mathbb{R}^2 \mid a \le x \le b \text{ and } 0 \le y \le f(x)\}$$

is bounded if and only if f is bounded.

Exercise 2.2. *Show that the axioms (I1), (I2) and (I3) imply that for any integrable $f : [a, b] \to \mathbb{R}$ and any $c \in [a, b]$, we have $\int_c^c f = 0$.*

Problems.

PROBLEM 2.1. *Suppose that you happen to know that every continuous function has an antiderivative. Show that, defining, for every continuous function $f : [a, b] \to \mathbb{R}$, $\int_a^b f$ to be $F(b) - F(a)$ where F is any antiderivative of f, defines an integral $\int : \mathcal{C}[a, b] \to \mathbb{R}$: in other words, check that the axioms (I1) through (I3) are satisfied.*

2. Darboux's Riemann Integral

In this section we present Darboux's approach to the Riemann integral. Throughout this section $a < b$ are real numbers.

2.1. Upper and lower sums, upper and lower integrals. Partitions: A **partition** of $[a, b]$ is a finite subset $\mathcal{P}$ of $[a, b]$ containing a and b. Thus we may write $\mathcal{P}$ as $\{x_0, x_1, \ldots, x_{n-1}, x_n\}$ with $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$. Notice that the positive integer n is one less than the number of elements of $\mathcal{P}$; we think of a partition $\mathcal{P}$ as subdividing the interval $[a, b]$ into subintervals $[a, x_1], [x_1, x_2], \ldots, [x_{n-1}, b]$, and thus n is the number of subintervals into which we subdivided $[a, b]$. The telescoping sum

$$\sum_{i=0}^{n-1} (x_{i+1} - x_i) = (x_1 - a) + (x_2 - x_1) + \ldots + (b - x_{n-1}) = b - a$$

shows that the length of the interval $[a, b]$ is the sum of the lengths of the subintervals into which we divided it using $\mathcal{P}$.

Because $[a, b]$ is infinite, there are certainly infinitely many partitions of it. We introduce a relation among them: we say that a partition $\mathcal{P}_2$ of $[a, b]$ **refines** a partition $\mathcal{P}_1$ of $[a, b]$ if $\mathcal{P}_1 \subseteq \mathcal{P}_2$: thus, $\mathcal{P}_2$ contains all the points of $\mathcal{P}_1$ and (if $\mathcal{P}_2 \neq \mathcal{P}_1$) some others. We can think of $\mathcal{P}_2$ as being obtained from $\mathcal{P}_1$ by repeatedly choosing one of the subintervals $[x_i, x_{i+1}]$ given by $\mathcal{P}_1$ and subdividing it by adding an addition point $z \in (x_i, x_{i+1})$. This refinement relation is a partial ordering on the set of partitions of $[a, b]$: this just means that every partition refines itself; if each of two partitions refines the other than they are equal; and if $\mathcal{P}_3$ refines $\mathcal{P}_2$ and $\mathcal{P}_2$ refines $\mathcal{P}_1$ then $\mathcal{P}_3$ refinee $\mathcal{P}_1$.

Now let $f : [a, b] \to \mathbb{R}$ be a *bounded* function. To every partition $\mathcal{P} = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ of $[a, b]$, we will define an **upper sum** $L(f, \mathcal{P}) \in \mathbb{R}$ and a **lower sum** $U(f, \mathcal{P}) \in \mathbb{R}$. To this we first define:

- For all $0 \leq i \leq n - 1$, let $M_i(f)$ be the supremum of $f([x_i, x_{i+1}])$, and
- For all $0 \leq i \leq n - 1$, let $m_i(f)$ be the infimum of $f[(x_i, x_{i+1}])$.

Now we put

$$U(f, \mathcal{P}) := \sum_{i=0}^{n-1} M_i(f)\,(x_{i+1} - x_i)$$

and

$$L(f, \mathcal{P}) := \sum_{i=0}^{n-1} m_i(f)\,(x_{i+1} - x_i)\,.$$

Some remarks are in order.

REMARK. a) *Since for any nonempty subset X of $\mathbb{R}$ we have $\inf X \leq \sup X$, for any $f : [a, b] \to \mathbb{R}$ and all $0 \leq i \leq n-1$ we have $m_i(f) \leq M_i(f)$, from which it follows that*

$$L(f, \mathcal{P}) \leq U(f, \mathcal{P}).$$

b) *For all $0 \leq i \leq n-1$, we have $M_i(f) \in \mathbb{R}$ because f is bounded above. Suppose on the other hand that f were not bounded above. Then by Exercise 2.3b), there is at least one i such that f is not bounded above on the subinterval $[x_i, x_{i+1}]$, so the supremum of $f([x_i, x_{i+1}])$ is ∞. It is a standard convention the arithmetic of extended real numbers that $\infty + \infty = \infty$ and for $a \in (0, \infty]$ we have $a \cdot \infty = \infty$. Using these conventions we find that if f is unbounded above we can make sense of the upper sum $U(f, \mathcal{P})$: it will always be ∞.*

c) *Similarly, for all $0 \leq i \leq n-1$, we have $m_i(f) \in \mathbb{R}$ because f is bounded below. If were unbounded below then by the same reasoning as part a) we find that we can make sense of the lower sum $L(f, \mathcal{P})$ but it will always be $-\infty$. Thus for* any *function f we have $U(f, \mathcal{P}) \in \mathbb{R} \cup \{\infty\}$ and $L(f, \mathcal{P}) \in \mathbb{R} \cup \{-\infty\}$.*

d) *Suppose that $f : [a, b] \to \mathbb{R}$ is continuous. Then f is bounded, and for all $1 \leq i \leq n$, by the Extreme Value Theorem we get that $m_i(f)$ is the* ***minimum*** *of f on $[x_i, x_{i+1}]$ and $M_i(f)$ is the* ***maximum*** *of f on $[x_i, x_{i+1}]$. Though we will not define Riemann sums until the next section – the extra complication of choosing a "sample point" in each subinterval is part of what Darboux's approach manages to avoid – nevertheless we remark now that when f is continuous the upper sum and lower sum are both Riemann sums for f.*

EXAMPLE 2.1. *Consider $f : [0, 1] \to \mathbb{R}$ by $f(x) = x^2$.*

a) *Suppose we take the smallest possible partition: $\mathcal{P}_1 = \{0, 1\}$. The minimum of f on $[0, 1]$ is 0 and the maximum of f is 1, so*

$$L(x^2, \mathcal{P}_1) = 0 < 1 = U(x^2, \mathcal{P}_1).$$

We can interpret this geometrically: consider the $S_{x^2} = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1,\ 0 \leq y \leq x^2\}$, whose area we are trying to define via some limiting process. This set contains the line segment $S_0 = [0, 1] \times \{0\}$, that has area 0, and it is contained in the unit square $S_1 = [0, 1] \times [0, 1]$, that has area 1. So although we haven't defined the integral yet, the idea is that we have learned from $\mathcal{P}_1$ is that we want $\int_0^1 x^2$ to be some real number such that

$$L(x^2, \mathcal{P}_1) = 0 < \int_0^1 x^2 \leq 1 = U(x^2, \mathcal{P}_1).$$

b) *Even a vague memory of definite integrals from calculus should suggest that we try something else: for $n \in \mathbb{Z}^+$ let*

$$\mathcal{P}_n := \{0 < \frac{1}{n} < \frac{2}{n} < \ldots \frac{n-1}{n} < 1\}$$

be the partition that subdivides $[0, 1]$ into n equally spaced subintervals. Because $f(x) = x^2$ is increasing, the supremum it takes on any subinterval

$[c,d]$ of $[0,1]$ is just $f(d)$ and the infimum it takes on any subinterval $[c,d]$ of $[0,1]$ is just $f(c)$. So:

$$U(x^2,\mathcal{P}_n)=\sum_{i=0}^{n-1}M_i(x^2)\left(\frac{i+1}{n}-\frac{i}{n}\right)=\sum_{i=0}^{n-1}(\frac{i+1}{n})^2\cdot\frac{1}{n}$$
$$=\frac{1}{n^3}\sum_{i=0}^{n-1}(i+1)^2=\frac{1}{n^3}\sum_{i=1}^{n}i^2.$$

Oh, thank goodness that in a previous course (Math 3200) we practiced induction with sums like these: we happen to remember that

$$\forall n\in\mathbb{Z}^+,\sum_{i=1}^{n}i^2=\frac{n(n+1)(2n+1)}{6},$$

so we get

$$U(x^2,\mathcal{P}_n)=\frac{n(n+1)(2n+1)}{6n^3}.$$

The computation for the lower sums is very similar: for all i we have $m_i(x^2)=\left(\frac{i}{n}\right)^2$, which leads to

$$L(x^2,\mathcal{P}_n)=\frac{1}{n^3}\sum_{i=0}^{n-1}i^2=\frac{1}{n^3}\sum_{i=1}^{n-1}i^2=\frac{(n-1)n(2n-1)}{6n^3}.$$

Now we observe that

$$\lim_{n\to\infty}U(x^2,\mathcal{P}_n)=\frac{1}{3}=\lim_{x\to\infty}L(x^2,\mathcal{P}_n).$$

So we found a sequence of partitions along which the lower sums converged to $\frac{1}{3}$ and along which the upper sums also converged to $\frac{1}{3}$. This makes us strongly suspect that we want $\int_0^1 x^2=\frac{1}{3}$. In calculus we would probably be happy to take either one of these limits *as sufficient to give the answer, but now we are trying to find our way to a principled definition of an integrable function. We can reason as follows: for* any *partition $\mathcal{P}$, we can interpret $U(f,\mathcal{P})$ as the area enclosed by a piecewise constant function that is always greater than or equal to f and we can interpret $L(f,\mathcal{P})$ as the area enclosed by a piecewise constant function that is always less than or equal to f, so we should have*

$$\forall \text{ partitions } \mathcal{P} \text{ of } [a,b],\ L(f,\mathcal{P})\le\int_a^b f\le U(f,\mathcal{P}). \tag{6}$$

So in our case we want $\int_a^b f$ to satisfy

$$\forall n\in\mathbb{Z}^+,\ \frac{(n-1)n(2n-1)}{6n^3}\le\int_0^1 x^2\le\frac{n(n+1)(2n+1)}{6n^3}.$$

Limits of sequences preserve lax inequalities ($\le$ and $\ge$, not $<$ and $>$), so

$$\frac{1}{3}=\lim_{n\to\infty}L(f,\mathcal{P})\le\int_0^1 x^2\le\lim_{n\to\infty}U(f,\mathcal{P})=\frac{1}{3}.$$

This tells us that $\int_0^1 x^2=\frac{1}{3}$!

In other words, our one idea about $\int_a^b f$ is that it should lie in between

$L(f,\mathcal{P})$ and $U(f,\mathcal{P})$ for any partition $\mathcal{P}$ of $[a,b]$. In this particular example, just by looking at the sequence of partitions $\{\mathcal{P}_n\}_{n=1}^{\infty}$ we see that the only real number that could possibly satisfy this is $\frac{1}{3}$.

This leads us to our first definition of Darboux integrability: a function $f : [a,b] \to \mathbb{R}$ is **Darboux integrable** if there is exactly one real number S such that for all partitions $\mathcal{P}$ of $[a,b]$ we have

$$L(f,\mathcal{P}) \leq S \leq U(f,\mathcal{P});$$

for a Darboux integrable function f we put $\int_a^b f$ to be this unique real number S.

PROPOSITION 2.3. *Let $f : [a,b] \to \mathbb{R}$ be a function. If f is Darboux integrable, then it is bounded.*

PROOF. We will show the contrapositive: suppose f is unbounded; we claim that f is *not* Darboux integrable.
Case 1: If f is unbounded both above and below then for all partitions $\mathcal{P}$ of $[a,b]$ we have $U(f,\mathcal{P}) = \infty$ and $L(f,\mathcal{P}) = -\infty$. So every real number lies in between every lower sum and upper sum: thus the uniqueness of I fails.
Case 2: Suppose f is unbounded above but bounded below. Then for all partitions $\mathcal{P}$ of $[a,b]$ we have that $U(f,\mathcal{P}) = \infty$ but $L(f,\mathcal{P}) \in \mathbb{R}$. Every real number is at most ∞, so in order to be Darboux integrable there would have to be a unique real number I greater than or equal to $L(f,\mathcal{P})$ for all partitions $\mathcal{P}$. In other words, the set $\{L(f,\mathcal{P}) \mid \mathcal{P}$ is a partition of [a,b]$\}$ would need to have a *unique* upper bound. That's not how upper bounds in $\mathbb{R}$ work: if $I \in \mathbb{R}$ is an upper bound for *any* subset X of $\mathbb{R}$ then so is $I+1$, so X cannot have a unique, finite upper bound.
Case 3: If f is bounded above but unbounded below, the reasoning of Case 2 applies: there is no *unique* real number less than or equal $U(f,\mathcal{P})$ for all $\mathcal{P}$. $\square$

However, this definition of Darboux integrability is not so easy to work with: one can see this by observing that in Example 2.1 we did *not* show that x^2 is Darboux integrable on $[0,1]$: all we showed was that *if it is*, then the integral is $\frac{1}{3}$.

The awkwardness in our definition of Darboux integrability is characteristic of many definitions in theoretical mathematics: the definition involves a universal quantifier over an infinite set and for each element of that set asserts something nontrivial. Here that set is the set of all partitions of $[a,b]$. In our above example, showing Darboux integrability apparently asks us to compute $U(f,\mathcal{P})$ and $L(f,\mathcal{P})$ for *every* partition of $[0,1]$ and check that $L(f,\mathcal{P}) \leq \frac{1}{3} \leq U(f,\mathcal{P})$. Are we really supposed to perform *infinitely many* computations to check that x^2 is integrable?!?

No, not really. This definition is too hard to check directly, so we need a result that tells us that it is sufficient to do something easier. The result that we are going for here is as follows: a bounded function $f : [a,b] \to \mathbb{R}$ is Darboux integrable if and only if: for all $\epsilon > 0$, there is a partition $\mathcal{P}$ of $[a,b]$ such that $U(f,\mathcal{P}) - L(f,\mathcal{P}) < \epsilon$. Once we establish this, we don't need to look at *all* partitions; we just need to exhibit a *sequence* of partitions along which the gap between the upper and lower sums tends to 0. That is much easier: in Example 2.1, the sequence $\{\mathcal{P}_n\}$ works. Because $\lim_{n\to\infty} U(f,\mathcal{P}_n) = \frac{1}{3} = \lim_{n\to\infty} L(f,\mathcal{P}_n)$, it follows that $\lim_{n\to\infty} U(f,\mathcal{P}_n) - L(f,\mathcal{P}_n) = 0$, so for any $\epsilon > 0$, just taking $\mathcal{P}_n$ for large enough n does what we want.

In order to show this, we need a few preliminaries. They involve refinements of partitions, which you may notice that we defined but have not yet used for anything whatsoever. Well, now is the time.

LEMMA 2.4. *Let $f : [a, b] \to \mathbb{R}$ be a function.*

a) *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be partitions of $[a, b]$, with $\mathcal{P}_2$ refining $\mathcal{P}_1$ (that is, $\mathcal{P}_2 \supseteq \mathcal{P}_1$). Then we have*

$$L(f, \mathcal{P}_1) \leq L(f, \mathcal{P}_2) \leq U(f, \mathcal{P}_2) \leq U(f, \mathcal{P}_1).$$

b) *Let $\mathcal{P}$ and $\mathcal{Q}$ be any partitions of $[a, b]$. Then we have*

$$L(f, \mathcal{P}) \leq U(f, \mathcal{Q}).$$

PROOF. We get from $\mathcal{P}_1$ to $\mathcal{P}_2$ by adding finitely many more points. So it suffices to treat the case in which $\mathcal{P}_2$ is obtained from $\mathcal{P}_1$ by adding a single additional point $c \in (x_i, x_{i+1})$ for some $0 \leq i \leq n-1$ and then show that $L(f, \mathcal{P}_2) \leq L(f, \mathcal{P}_1)$ and $U(f, \mathcal{P}_2) \leq U(f, \mathcal{P}_1)$. (Notice that we already know the middle inequality $L(f, \mathcal{P}_2) \leq U(f, \mathcal{P}_2)$; it is just there to make everything look nice.) This is actually quite easy: most of the terms in the sums $U(f, \mathcal{P}_1)$ and $U(f, \mathcal{P}_2)$ are the same; the only change is that we replace the ith term $\sup(f[x_i, x_{i+1}]) \cdot (x_{i+1} - x_i)$ of $U(f, \mathcal{P}_1)$ with the two terms $\sup(f[x_i, c]) \cdot (c - x_i) + \sup(f[c, x_{i+1}]) \cdot (x_{i+1} - x_c)$. If $A \subseteq B \subset \mathbb{R}$ then $\sup A \leq \sup B$, so we have

$$\sup(f[x_i, c]) \cdot (c - x_i) + \sup(f[c, x_{i+1}]) \cdot (x_{i+1} - x_c)$$

$$\leq \sup(f[x_i, x_{i+1}]) \cdot (c - x_i) + \sup(f[x_i, x_{i+1}]) \cdot (x_{i+1} - c) \leq \sup(f[x_i, x_{i+1}]) \cdot (x_{i+1} - x_i).$$

Thus $U(f, \mathcal{P}_2) \leq U(f, \mathcal{P}_1)$. The same reasoning works for the lower sums: the infimum of f on $[x_i, x_{i+1}]$ is less than or equal to its infimum on $[x_i, c]$ and its infimum on $[c, x_{i+1}]$.

b) Let $\mathcal{R} := \mathcal{P} \cup \mathcal{Q}$; this is a partition of $[a, b]$ that is a *common refinement* of $\mathcal{P}$ and $\mathcal{Q}$. Applying part a) twice, we get

$$L(f, \mathcal{P}) \leq L(f, \mathcal{R}) \leq U(f, \mathcal{R}) \leq U(f, \mathcal{Q}). \qquad \square$$

Now one final definition that we hinted at in the last section: for any function $f : [a, b] \to \mathbb{R}$, we define the **upper Darboux integral**

$$\overline{\int_a^b} f := \inf U(f, \mathcal{P}) \in [-\infty, \infty]$$

and the **lower Darboux integral**

$$\underline{\int_a^b} f := L(f, \mathcal{P}) \in [-\infty, \infty];$$

in each case we are ranging over all partitions $\mathcal{P}$ of $[a, b]$. Notice that $\overline{\int_a^b} f$ is a "minimax": for each partition we maximized f (actually we took suprema, but people don't say "infysup"), collected these values over all partitions and then took the minimum (actually the infimum). Simlarly, $\underline{\int_a^b} f$ is a "maximin." When it makes sense to do, it's often a surprisingly good idea to take minimaxes and maximins and to compare them: see e.g. `https://en.wikipedia.org/wiki/Minimax_theorem`, which is the foundational result in Game Theory. Anyway, there is a clear geometric idea: the upper integral $\overline{\int_a^b} f$ is the best upper bound one can get on the area of

the region S_f using upper rectangles, while the lower integral $\underline{\int}_a^b f$ is the best lower bound one can get on the same area using lower rectangles. So if we want there to be a unique real number lying between all the areas of lower rectangles and all the areas of upper rectangles, then presumably we want the upper and lower integrals to be equal, finite numbers. We are about to show this, but first one "sanity check":

LEMMA 2.5. *Let $f : [a, b] \to \mathbb{R}$ be a function.*

a) *We have*

$$\underline{\int}_a^b f \leq \overline{\int}_a^b f.$$

b) *If f is bounded, then $\overline{\int}_a^b f, \underline{\int}_a^b f \in \mathbb{R}$.*

PROOF. a) The lower integral $\underline{\int}_a^b f$ is the supremum of the set

$$X := \{L(f, \mathcal{P}) \mid \mathcal{P} \text{ is a partition of } [a, b]\},$$

while the upper integral $\overline{\int}_a^b f$ is the infimum of the set

$$Y := \{U(f, \mathcal{P}) \mid \mathcal{P} \text{ is a partition of } [a, b]\}.$$

But Lemma 2.4b) says that for all $x \in X$ and all $y \in Y$ we have $x \leq y$. Thus every $x \in X$ is a lower bound for Y, so $x \leq \inf Y$, and since this holds for all $x \in X$ we have $\sup X \leq \inf Y$.
b) Suppose that f is bounded: there is $M > 0$ such that $|f| \leq M$. Then for any partition $\mathcal{P}$ of $[a, b]$ we have

$$-M(b - a) \leq L(f, \mathcal{P}) \leq U(f, \mathcal{P}) \leq M(b - a),$$

and it follows that $\overline{\int}_a^b f,\ \underline{\int}_a^b f \in [-M(b - a), M(b - a)]$. □

THEOREM 2.6 (Darboux's Integrability Criterion). *For a function $f : [a, b] \to \mathbb{R}$, the following are equivalent:*

(i) *There is a unique real number I such that for all partitions $\mathcal{P}$ of $[a, b]$ we have $L(f, \mathcal{P}) \leq I \leq U(f, \mathcal{P})$.*
(ii) *We have $\underline{\int}_a^b f = \overline{\int}_a^b f \in \mathbb{R}$.*
(iii) *For all $\epsilon > 0$, there is a partition $\mathcal{P}$ such that $U(f, \mathcal{P})$ and $L(f, \mathcal{P})$ are real numbers and $U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon$.*

Henceforth we call a function satisfying these conditions ***Darboux integrable****.*

PROOF. Step 1: Suppose first that f is unbounded. By Proposition 2.3, condition (i) fails. Moreover either f is unbounded above — in which case $U(f, \mathcal{P}) = \infty$ for all $\mathcal{P}$ hence $\overline{\int}_a^b f = \infty$, so (ii) and (iii) fail — or f is unbounded below – in which case $L(f, \mathcal{P}) = -\infty$ for all $\mathcal{P}$ hence $\underline{\int}_a^b f = -\infty$, so again (ii) and (iii) fail.

So it suffices to consider the case in which f is bounded. In this case, by Exercise 2.4 we know that $\overline{\int}_a^b f$ and $\underline{\int}_a^b f$ are both finite.

Step 2: We show that (i) $\iff$ (ii). For a real number S, we have $S \leq \overline{\int}_a^b f$ if and only if I is less than or equal to every upper sum of f, and we have $S \geq \underline{\int}_a^b f$ if and only if S is greater than or equal to every lower sum of f, so a real number S

lies in the interval $[\underline{\int}_a^b f, \overline{\int}_a^b f]$ if and only if it lies in between every lower sum of f and every upper sum of f. So the upper and lower integrals are equal if and only if there is a unique S in between every lower sum of f and every upper sum of f.
Step 3: We show that (ii) $\iff$ (iii): If (ii) holds, let $\epsilon > 0$. Then there is a partition $\mathcal{P}$ of $[a, b]$ such that

$$L(f, \mathcal{P}) > \underline{\int}_a^b f - \frac{\epsilon}{2}$$

and another partition $\mathcal{Q}$ of $[a, b]$ such that

$$U(f, \mathcal{Q}) < \overline{\int}_a^b f + \frac{\epsilon}{2},$$

so

$$U(f, \mathcal{Q}) - L(f, \mathcal{P}) < \overline{\int}_a^b f - \underline{\int}_a^b f + \epsilon = \epsilon.$$

Now let $\mathcal{R} := \mathcal{P} \cup \mathcal{Q}$. Since $\mathcal{R}$ refines both $\mathcal{P}$ and $\mathcal{Q}$, we have

$$U(f, \mathcal{R}) \leq U(f, \mathcal{P}) \text{ and } L(f, \mathcal{R}) \geq L(f, \mathcal{P}),$$

so

$$U(f, \mathcal{R}) - L(f, \mathcal{R}) \leq U(f, \mathcal{Q}) - L(f, \mathcal{P}) < \epsilon.$$

If (iii) holds, then let $\epsilon > 0$, and choose a partition $\mathcal{P}$ such that $U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon$. Then, since $\overline{\int}_a^b f \leq U(f, \mathcal{P})$ and $\underline{\int}_a^b f \geq L(f, \mathcal{P})$, we have

$$\overline{\int}_a^b f - \underline{\int}_a^b f \leq U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon.$$

Since this holds for all $\epsilon > 0$, we have $\overline{\int}_a^b f = \underline{\int}_a^b f$. □

2.2. Verification of the Axioms. Our next order of business is to check that the Darboux integral that we have defined satisfies Axioms (I0), (I1), (I2) and (I3) from §3.1. "Checking axioms" doesn't sound so exciting, but we get quite a payoff: the Fundamental Theorem of Calculus, which includes the fact that every continuous function $f : [a, b] \to \mathbb{R}$ is a derivative.

The verification of (I1) is left as Exercise 2.5 and the verification of (I2) is left as Exercise 2.6. Checking the third axiom is less straightforward:

PROPOSITION 2.7. *Let $f : [a, b] \to \mathbb{R}$ be a Darboux integrable function.*

a) *For any $a \leq c \leq d \leq b$, the function $f|_{[c,d]} : [c, d] \to \mathbb{R}$ is Darboux integrable.*
b) *For any $a \leq c \leq b$, we have $\int_a^b f = \int_a^c f + \int_c^b f$.*

PROOF. a) Let $\epsilon > 0$. Since f is Darboux integrable, there is a partition $\mathcal{P}$ of $[a, b]$ such that $U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon$. Let $\mathcal{P}' := \mathcal{P} \cup \{c, d\}$, and write

$$\mathcal{P}' = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}.$$

Since $\mathcal{P}'$ refines $\mathcal{P}$, we have $U(f, \mathcal{P}') \leq U(f, \mathcal{P})$ and $L(f, \mathcal{P}') \geq L(f, \mathcal{P})$, so

$$U(f, \mathcal{P}') - L(f, \mathcal{P}') < \epsilon.$$

Now $\mathfrak{p} := \mathcal{P}' \cap [c,d]$ is a partition of $[c,d]$: it contains c and d and is a suset of a finite set, hence finite. To be specific, suppose that $c = x_I$ and $d = x_J$. Then

$$U(f,\mathfrak{p}) - L(f,\mathfrak{p}) = \sum_{i=I}^{J-1} \left(\sup(f[x_i,x_{i+1}]) - \inf(f[x_i,x_{i+1}])\right)(x_{i+1}-x_i),$$

whereas

$$U(f,P') - L(f,\mathcal{P}') = \sum_{i=0}^{n-1} \left(\sup(f[x_i,x_{i+1}]) - \inf(f[x_i,x_{i+1}])\right)(x_{i+1}-x_i).$$

The only difference between the former sum and the latter is that in the former sum we are summing from I to $J-1$ and in the latter we are summing from 0 to $n-1$, so the latter sum is the former sum together with some additional terms. But every term in either sum is non-negative, because the supremum of f on any subinterval is at least as large as its infimum on that subinterval. Thus:

$$U(f,\mathfrak{p}) - L(f,\mathfrak{p}) \leq U(f,\mathcal{P}') - L(f,\mathcal{P}') < \epsilon.$$

By Theorem 2.6, $f_{[c,d}$ is Darboux integrable.
b) Let $\mathcal{P}$ be a partition of $[a,b]$, and let $\mathcal{P}' := \mathcal{P} \cup \{c\}$. We also put

$$\mathcal{P}_L := \mathcal{P} \cap [a,c] \text{ and } \mathcal{P}_R := \mathcal{P} \cap [c,b],$$

so $\mathcal{P}_L$ is a partition of $[a,c]$ and $\mathcal{P}_R$ is a partition of $[c,b]$. Similarly to part a), upon writing out the partial sums we find immediately that

$$U(f,\mathcal{P}') = U(f|_{[a,c]},\mathcal{P}_L) + U(f|_{[c,b]},\mathcal{P}_R) \text{ and } L(f,\mathcal{P}') = L(f|_{[a,c]},\mathcal{P}_L) + L(f|_{[c,b]},\mathcal{P}_R).$$

Moreover, since $\mathcal{P}'$ is a refinement of $\mathcal{P}$ we have

$$L(f,\mathcal{P}) \leq L(f,\mathcal{P}') \text{ and } U(f,\mathcal{P}') \leq U(f,\mathcal{P}).$$

By part a), $f|_{[a,c]} : [a,c] \to \mathbb{R}$ and $f|_{[c,b]} : [c,b] \to \mathbb{R}$ are Darboux integrable, so

$$\begin{aligned} L(f,\mathcal{P}) \leq L(f,\mathcal{P}') &= L(f|_{[a,c]},\mathcal{P}_L) + L(f|_{[c,b]},\mathcal{P}_R) \\ &\leq \int_a^c f + \int_c^b f \\ &\leq U(f|_{[a,c]},\mathcal{P}_L) + U(f|_{[c,b]},\mathcal{P}_R) = U(f,\mathcal{P}') \leq U(f,\mathcal{P}). \end{aligned}$$

Thus $\int_a^c f + \int_c^b f$ lies between every lower sum and every upper sum. Since f is Darboux integrable, the unique such real number is $\int_a^b f$, and we conclude:

$$\int_a^b f = \int_a^c + \int_c^b f. \qquad \square$$

We have already shown that every Darboux integrable function is bounded: Proposition 2.3. The last, and most important, thing we have to show is this:

THEOREM 2.8. *Let $f : [a,b] \to \mathbb{R}$ be continuous. Then f is Darboux integrable.*

PROOF. The key is that by Theorem 1.31 we know that f is *uniformly* continuous. So let $\epsilon > 0$; we may choose $\delta > 0$ such that for all $x, y \in [a,b]$, if $|x-y| < \delta$ then $|f(x) - f(y)| < \frac{\epsilon}{b-a}$. Now choose $N \in \mathbb{Z}^+$ such that $\frac{b-a}{N} < \delta$ and let $\mathcal{P}_N$ be the partition that divides $[a,b]$ into N subintervals of equal length. Then

$$U(f,\mathcal{P}_N) - L(f,\mathcal{P}_N) = \left(\frac{b-a}{N}\right) \sum_{i=0}^{N-1} \left(\sup(f[x_i,x_{i+1}]) - \inf(f[x_i,x_{i+1}])\right). \tag{7}$$

Because $x_{i+1} - x_i = \frac{b-a}{N} < \delta$, on the subinterval $[x_i, x_{i+1}]$ any two values of f differ from each other by less than $\frac{\epsilon}{b-a}$, so

$$\forall 0 \leq i \leq N-1,\ \sup(f[x_i, x_{i+1}]) - \inf(f[x_i, x_{i+1}]) \leq \frac{\epsilon}{b-a}.$$

If we apply this inequality to each term of (7), we now get $\frac{b-a}{N}$ times a sum of N terms, each one of which is at most $\frac{\epsilon}{b-a}$, so we get

$$U(f, \mathcal{P}_N) - L(f, \mathcal{P}_N) \leq \frac{b-a}{N} \cdot N \cdot \frac{\epsilon}{b-a} \leq \epsilon.$$

So f is Darboux integrable by Theorem 2.6. □

Finally the circle has been completed: we shown that the Darboux integral satisfies all of our axioms (I0) through (I3), so we do have a gadget $\int : \mathcal{R}_D[a,b] \to \mathbb{R}$ to plug into the hypothesis of the Fundamental Theorem of Calculus. Thus the Fundamental Theorem of Calculus becomes an unconditional result, and in particular we have shown that every continuous function has an antiderivative.

We were fortunate enough to know the Uniform Continuity Theorem (Theorem 1.31), so we used it to get a very agreeable proof of Theorem 2.8. In contrast to the situation of showing that a continuous function has an antiderivative – which I do not know how to show without somehow constructing a definite integral – there are alternate approaches to Theorem 2.8 that avoid the use of uniform continuity. See for instance [**HC**, Thm. 8.9] or [**No52**].

2.3. Linearity of the Darboux Integral. Before proceeding further, it will be helpful to introduce some further notation regarding the quantity $U(f, \mathcal{P}) - L(f, \mathcal{P})$, which appears in condition (iii) in Darboux's Integrability Criterion (Theorem 2.6) and therefore shows up often in our arguments. If $\mathcal{P} = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ then

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) = \sum_{i=0}^{n-1} \left(\sup(f[x_i, x_{i+1}]) - \inf(f[x_i, x_{i+1}])\right)(x_{i+1} - x_i).$$

For a function $f : I \to \mathbb{R}$ defined on an interval I, let us define the **oscillation of f on I** to be

$$\omega(f, I) := \sup(f(I)) - \inf(f(I)) \in [-\infty, \infty].$$

This is an extended real number which lies in $\mathbb{R}$ if and only if f is bounded on I, which will almost always be the case for us. Then we have

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) = \sum_{i=0}^{n-1} \omega(f, [x_i, x_{i+1}])(x_{i+1} - x_i).$$

And let us also put

$$\Delta(f, \mathcal{P}) := U(f, \mathcal{P}) - L(f, \mathcal{P}).$$

Thus f is Darboux integrable if and only if for all $\epsilon > 0$ there is a partition $\mathcal{P}$ of $[a, b]$ with $\Delta(f, \mathcal{P}) < \epsilon$, and moreover if $\mathcal{P}'$ is a partition refining $\mathcal{P}$ then

$$\Delta(f, \mathcal{P}') \leq \Delta(f, \mathcal{P}).$$

THEOREM 2.9. *Let $\mathcal{R}_D[a,b]$ be the set of Darboux integrable functions $f : [a,b] \to \mathbb{R}$. Then the Darboux integral*

$$\int : \mathcal{R}_D[a,b] \to \mathbb{R}$$

is a ***linear functional*** *– that is:*

a) *The set $\mathcal{R}_D[a,b]$ is a subspace of the vector space of all functions $f : [a,b] \to \mathbb{R}$.*
b) *The function $\int : \mathcal{R}_D[a,b] \to \mathbb{R}$ is a linear map: for all $f, g \in \mathcal{R}_D[a,b]$ and all $\alpha, \beta \in \mathbb{R}$, we have*

$$\int_a^b (\alpha f + \beta g) = \alpha \int_a^b f + \beta \int_a^b g.$$

PROOF. Equivalently, and perhaps more plainly, we must prove that if $f, g : [a,b] \to \mathbb{R}$ are Darboux integrable, then:
(i) For all $\alpha \in \mathbb{R}$, αf is also Darboux integrable, and moreover $\int_a^b (\alpha f) = \alpha \int_a^b f$;
(ii) $f + g$ is also Darboux integrable, and moreover $\int_a^b (f+g) = \int_a^b f + \int_a^b g$.
Assertion (i) is mostly a matter of pulling constants through upper and lower sums, so we leave this as Exercise 2.8.

Now let us show assertion (ii). Let $\epsilon > 0$; because f and g are Darboux integrable, there is a partition $\mathcal{P}_1$ of $[a,b]$ such that $\Delta(f, \mathcal{P}_1) < \frac{\epsilon}{2}$ and a partition $\mathcal{P}_2$ of $[a,b]$ such that $\Delta(f, \mathcal{P}_2) < \frac{\epsilon}{2}$. Let $\mathcal{P}_3$ be a common refinement of $\mathcal{P}_1$ and $\mathcal{P}_2$ (e.g. take $\mathcal{P}_3 = \mathcal{P}_1 \cup \mathcal{P}_2$); then $\Delta(f, \mathcal{P}_3)$ and $\Delta(g, \mathcal{P}_3)$ are each less than $\frac{\epsilon}{2}$.

We observe that for an interval I and for functions $f, g : I \to \mathbb{R}$, we have

$$\sup((f+g)(I)) \leq \sup(f(I)) + \sup(g(I)) \text{ and } \inf((f+g)(I)) \geq \inf(f(I)) + \inf(g(I)).$$

You are asked to show this in Exercise 2.9. Using these inequalities we get

$$L(f, \mathcal{P}_3) + L(g, \mathcal{P}_3) \leq L(f+g, \mathcal{P}_3) \leq U(f+g, \mathcal{P}_3) \leq U(f, \mathcal{P}_3) + U(g, \mathcal{P}_3).$$

This shows that

$$\Delta(f+g, \mathcal{P}_3) \leq \Delta(f, \mathcal{P}_3) + \Delta(g, \mathcal{P}_3) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

and thus $f + g$ is Darboux integrable. Moreover, whenever we have a Darboux integrable function $h : [a,b] \to \mathbb{R}$ and a partition $\mathcal{P}$ of $[a,b]$ such that $\Delta(h, \mathcal{P}) \leq \epsilon$, we know $\int_a^b h$ lies in the interval $[L(h, \mathcal{P}), U(h, \mathcal{P})]$ of length $\Delta(h, P) \leq \epsilon$, so we know that $\int_a^b h$ has distance at most ϵ from each of $U(h, \mathcal{P})$ and $L(h, \mathcal{P})$. So:

$$\int_a^b (f+g) \leq U(f+g, \mathcal{P}_3) \leq U(f, \mathcal{P}_3) + U(g, \mathcal{P}_3) \leq \int_a^b f + \int_a^b g + \epsilon$$

and similarly

$$\int_a^b f + \int_a^b g - \epsilon < L(f, \mathcal{P}_3) + L(g, \mathcal{P}_3) \leq L(f+g, \mathcal{P}_3) \leq \int_a^b (f+g).$$

This shows that $\left| \int_a^b (f+g) - (\int_a^b f + \int_a^b g) \right| \leq \epsilon$. Since this holds for all $\epsilon > 0$, we have $\int_a^b (f+g) = \int_a^b f + \int_a^b g$. $\square$

As one simple application of Theorem 2.9, we can reduce the study of $\int_a^b f$ to the case in which f is non-negative and thus officially not worry about signed areas. Indeed, if $f : [a, b] \to \mathbb{R}$ is Darboux integrable, then it is bounded, so there is a $C \geq 0$ such that $f + C \geq 0$ on $[a, b]$. So $\int_a^b (f + C)$ really does represent the area of the region S_{f+C}, and we can recover $\int_a^b f$ from this as

$$\int_a^b f = \int_a^b (f + C) - \int_a^b C = \int_a^b (f + C) - C(b - a).$$

Exercises.

EXERCISE 2.3. *Let X be a subset of $\mathbb{R}^N$, and let $Y_1, \ldots, Y_n$ be finitely many subsets of X such that $X = \bigcup_{i=1}^n Y_i$. Let $f : X \to \mathbb{R}^M$.*

a) *Show: f is bounded if and only if, for each $1 \leq i \leq n$, $f|_{Y_i} : Y_i \to \mathbb{R}^M$ is bounded.*
b) *Suppose $M = 1$. Show: f is bounded above if and only if, for each $1 \leq i \leq n$, $f|_{Y_i} : Y_i \to \mathbb{R}$ is bounded above. Then show the same with "bounded above" replaced everywhere by "bounded below."*

EXERCISE 2.4. *Let $f : [a, b] \to \mathbb{R}$.*

a) *Suppose that f is bounded above by $M \in \mathbb{R}$: we have $f(x) \leq M$ for all $x \in [a, b]$. Show: for every partition $\mathcal{P}$ of $[a, b]$, we have*
$$\overline{\int_a^b} f \leq U(f, \mathcal{P}) \leq M(b - a).$$
b) *Suppose that f is bounded below by $m \in \mathbb{R}$: we have $f(x) \geq m$ for all $x \in [a, b]$. Show: for every partition $\mathcal{P}$ of $[a, b]$, we have*
$$m(b - a) \leq L(f, \mathcal{P}) \leq \underline{\int_a^b} f.$$

EXERCISE 2.5. *Let $f : [a, b] \to \mathbb{R}$ be defined by $f(x) = C$ for all $x \in [a, b]$.*

a) *Show: for every partition $\mathcal{P}$ of $[a, b]$ we have $U(f, \mathcal{P}) = L(f, \mathcal{P}) = C(b-a)$.*
b) *Deduce: f is Darboux integrable and $\int_a^b C = C(b - a)$. Thus Axiom (I1) holds for the Darboux integral.*

EXERCISE 2.6. *Let $f, g : [a, b] \to \mathbb{R}$ be two Darboux integrable functions with $f \leq g$: that is, for all $x \in [a, b]$, we have $f(x) \leq g(x)$.*

a) *Show: for every partition $\mathcal{P}$ of $[a, b]$ we have $U(f, \mathcal{P}) \leq U(g, \mathcal{P})$ and $L(f, \mathcal{P}) \leq L(g, \mathcal{P})$.*
b) *Deduce: $\int_a^b f \leq \int_a^b g$. Thus Axiom (I2) holds for the Darboux integral.*

EXERCISE 2.7. *Let $f : [a, b] \to \mathbb{R}$ and let $c \in (a, b)$. Suppose that each of $f|_{[a,c]} : [a, c] \to R$ and $f|_{[c,b]} : [c, b] \to \mathbb{R}$ are Darboux integrable. Show: $f : [a, b] \to \mathbb{R}$ is Darboux integrable and $\int_a^b f = \int_a^c f + \int_c^b f$.*
(This is similar to Proposition 2.7 and — hint — can be proved in much the same way. Once we establish this, the proof of Proposition 2.7b) applies verbatim to give $\int_a^b f = \int_a^c f + \int_c^b f$. You can just say so: no need to repeat the argument.)

EXERCISE 2.8. *Suppose $f : [a, b] \to \mathbb{R}$ is Darboux integrable. Show: for all $\alpha \in \mathbb{R}$, the function $\alpha f : [a, b] \to \mathbb{R}$ is also Darboux integrable, and moreover*

$$\int_a^b (\alpha f) = \alpha \int_a^b f.$$

EXERCISE 2.9. *Let $X \subset \mathbb{R}^N$ and let $f, g : X \to \mathbb{R}$.*

a) *Show:*
$$\sup((f + g)(X) \leq \sup(f(X)) + \sup(g(X)).$$
b) *Show:*
$$\inf((f + g)(X) \geq \inf(f(X)) + \inf(g(X)).$$

(Comment: in part a), each of the terms is either a real number or ∞. In part b), each of the terms is either a real number or $-\infty$. Standard conventions on the arithmetic of extended real numbers apply, e.g. $\infty + \infty = \infty$ and for all $x \in \mathbb{R}$, $x + \infty = \infty$.)

EXERCISE 2.10 (Mean Value Theorem for Integrals). *Let $f : [a, b] \to \mathbb{R}$ be continuous. Show: there is $c \in [a, b]$ such that*

$$\int_a^b f = f(c) \cdot (b - a).$$

(Hint: let $m = \min f([a, b])$ and $M = \max f([a, b])$. Show that $\frac{\int_a^b f}{b-a} \in [m, M]$. Thus $\frac{\int_a^b f}{b-a}$ is intermediate *between two* values *of f....)*

EXERCISE 2.11. *Let $f : [0, 1] \to \mathbb{R}$ by $f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$.*

a) *Show: $\underline{\int}_0^1 f = 0$ and $\overline{\int}_0^1 f = 1$.*
b) *Deduce: f is bounded function that is not Darboux integrable.*

Problems.

PROBLEM 2.2. *Let $T : [0, 1] \to \mathbb{R}$ be* ***Thomae's function*** *of Problem 1.13, restricted to the unit interval. Recall from that exercise that f is continuous at every rational number and discontinuous at every irrational number.*[1] *Show: T is Darboux integrable and $\int_0^1 T = 0$.*

PROBLEM 2.3. *Let $f : [a, b] \to \mathbb{R}$ be bounded. Suppose that for all $c \in (a, b]$, the restricted function $f|_{[c,b]} : [c, b] \to \mathbb{R}$ is Darboux integrable. Show that f is Darboux integrable and $\lim_{c \to a^+} \int_c^b f = \int_a^b$.*
(Suggestion: use the fact that if $|f| \leq M$, then on any subinterval $[c, d]$, every upper sum of f is at most $M(d - c)$ and every lower sum of f is at least $-M(d - c)$, and note that these quantities approach 0 with the length of $[c, d]$.)

PROBLEM 2.4. *Suppose $f : [a, b] \to \mathbb{R}$ is bounded and has finitely many discontinuities. Show that f is Darboux integrable.*
(You may, or may not, wish to use Problem 2.3.)

[1] Strictly speaking, one should look again at the endpoints 0 and 1 of the interval: there are some functions $f : \mathbb{R} \to \mathbb{R}$ that are discontinuous at 0 and 1 such that after restricting to $[0, 1]$ become continuous at one or both endpoints. But in fact Thomae's function does not have even one-sided limits at any rational point.

PROBLEM 2.5. *Let a, b be positive real numbers, and define $f : \mathbb{R} \to \mathbb{R}$ by*

$$f(x) = \begin{cases} x^a \sin(\frac{1}{x^b}) & x \neq 0 \\ 0 & x = 0 \end{cases}.$$

a) *Show: for all values of a and b, $f_{a,b}$ is continuous.*
b) *Show: $f_{a,b}$ is differentiable if and only if $a > 1$.*
(Here and hereafter, the only issues are at $x = 0$; $f_{a,b}$ is certainly infinitely differentiable on $\mathbb{R} \setminus \{0\}$.)
c) *Show: $f'_{a,b}$ is continuous if and only if $a > b + 1$.*
d) *Show: if $a \in (1, b + 1)$, then $f'_{a,b}$ is unbounded on any open interval containing 0. Deduce: if $c < 0 < d$, $f'_{a,b}$ is not Darboux integrable on $[c, d]$.*
e) *Show: if $1 < a = b + 1$, then $f'_{a,b}$ exists, is discontinuous precisely at 0, and is bounded on any closed, bounded interval. Using Exercise 2.3, deduce that if $c < 0 < d$, then $f'_{a,b}|_{[c,d]} : [c, d] \to \mathbb{R}$ is Darboux integrable.*

PROBLEM 2.6 (Integrability of Monotone Functions).

a) *Let $f : [a, b] \to \mathbb{R}$ be an increasing function, and let $\mathcal{P}_n$ be the partition that divides $[a, b]$ into n equally spaced subintervals. Show:*

$$\Delta(f, \mathcal{P}) = U(f, \mathcal{P}) - L(f, \mathcal{P}) = (f(b) - f(a)) \cdot \left(\frac{b - a}{n}\right).$$

Use this to show that f is Darboux integrable.
b) *Show that if $f : [a, b] \to \mathbb{R}$ is decreasing, then it is Darboux integrable.*
c) *Show: if f is monotone, then $\lim_{n\to\infty} \frac{b-a}{n} \sum_{i=0}^{n-1} f(a + i(\frac{b-a}{n})) = \int_a^b f$.*

PROBLEM 2.7 (Monotone Functions Can Be Pretty Discontinuous). *Let $\{x_n\}_{n=1}^{\infty}$ be an injective sequence of real numbers: i.e., for all $m \neq n$ we have $x_m \neq x_n$. For $x \in \mathbb{R}$, let $S_x := \{n \in \mathbb{Z}^+ \mid x_n \leq x\}$. We define a function $f : \mathbb{R} \to \mathbb{R}$ as follows: for $x \in \mathbb{R}$,*

$$f(x) := \sum_{n \in S_x} 2^{-n}.$$

In other words, $f(x)$ is the sum of an infinite series whose nth term is 2^{-n} if $x_n \leq x$ and is 0 if $x_n > x$.

a) *Show that for all $x \in \mathbb{R}$, the infinite series defining $f(x)$ converges and we have $0 \leq f(x) \leq 1$.*
(Suggestion: compare to the geometric series $\sum_{n=1}^{\infty} 2^{-n} = 1$.)
b) *Show: f is increasing.*
c) *Show: for $n \in \mathbb{Z}^+$, $\lim_{x\to x_n^+} f(x) - \lim_{x\to x_n^-} f(x) = 2^{-n}$. Thus f is discontinuous at x_n.*
d) *Show: if $x \in \mathbb{R} \setminus \{x_n \mid n \in \mathbb{Z}^+\}$, then f is continuous at x.*
e) *Deduce: there is an increasing $f : [0, 1] \to \mathbb{R}$ that is continuous at every irrational point of $[0, 1]$ and discontinuous at every rational point of $[0, 1]$.*

3. Riemann's Riemann Integral

In this section we touch upon Riemann's construction of the Riemann integral, which was earlier than Darboux's. Riemann's construction is a bit more technically elaborate than Darboux's – hence our decision to start with, and mostly focus on, Darboux's – but it has its merits, and it is to our advantage to at least be

familiar with both.

In order to motivate Riemann's construction, imagine you have a function $f : [a, b] \to \mathbb{R}$ that you know is Darboux integrable: to fix ideas, let us suppose that it is continuous. Can we actually compute $\int_a^b f$?

Perhaps your first idea is to find an antiderivative F of f and use the Fundamental Theorem of Calculus: $\int_a^b f = F(b) - F(a)$. If you think this is how most integrals are actually computed, then you have been misled. Despite the time spent in freshman calculus on integration (i.e., antidifferentiation!) techniques, for any function more complicated than a rational function or a polynomial expression in trigonometric functions, it is quite rare to be able to write down an antiderivative "as an elementary function" of the sort you studied in precalculus. If your function is given by a power series expansion such that $[a, b]$ lies inside the open interval of convergence of the series, then you are in business: as you learned in Math 3100, power series can be integrated term by term, and moreover it is easy to estimate the value of a power series at a non-boundary point of convergence: you can cut off after finitely many terms and use geometric series to get an upper bound for the error. But such functions are a lot more than continuous: they are infinitely differentiable (and in fact, most infinitely differentiable functions are still not given by convergent power series expansions).

Returning to the Darboux integral, we point out that things work out very nicely if $f : [a, b] \to \mathbb{R}$ is monotone, as is explored in Exercise 2.6. To fix ideas, let us suppose that f is increasing. First of all, in this case, on any subinterval $[x_i, x_{i+1}]$ the supremum is just $f(x_{i+1})$, the value at the right endpoint, while the infimum is just $f(x_i)$, the value at the left endpoint. So we can actually compute $U(f, \mathcal{P})$ and $L(f, \mathcal{P})$ for any partition $\mathcal{P}$. Moreover, if you just take the partition $\mathcal{P}_n$ that subdivides $[a, b]$ into n equally spaced subintevals, then in the expression for $\Delta(f, \mathcal{P}_n) = U(f, \mathcal{P}_n) - L(f, \mathcal{P}_n)$ almost everything cancels out, and you are left with $(f(b) - f(a)) \cdot (\frac{b-a}{n})$. Evidently this approaches 0 as n approaches ∞, which already shows that f is Darboux integrable. But moreover, it follows easily from this that

$$\lim_{n\to\infty} U(f, \mathcal{P}_n) = \lim_{n\to\infty} L(f, \mathcal{P}_n) = \int_a^n f,$$

so choosing for instance the lower sum, we get concretely that

$$\int_a^n f = \lim_{n\to\infty} \left(\frac{b-a}{n}\right) \sum_{i=0}^{n-1} f(a + i(\frac{b-a}{n})).$$

Even if this limit is too hard to evaluate exactly (which it usually is), we can still compute, for any n, a lower bound $L(f, \mathcal{P}_n)$ for the integral and an upper bound $U(f, \mathcal{P}_n)$ for the integral, and as n approaches ∞, since each sequence approaches $\int_a^b f$, the gap between them $\Delta(f, \mathcal{P}_n)$ approaches 0. Therefore we can compute $\int_a^b f$ degree to accuracy ϵ by choosing a large n and computing $U(f, \mathcal{P}_n)$, $L(f, \mathcal{P}_n)$ and $\Delta(f, \mathcal{P}_n)$: if $\Delta(f, P_n) \leq \epsilon$; great. If not, try again with a larger value of n.

This works so well that we might try to bootstrap it to other functions, e..g. by breaking up $f : [a, b] \to \mathbb{R}$ into finitely many subintervals such that it is monotone

on each one. Unfortunately not every function is piecewise monotone, and for those which are we may have to do quite a lot of work in order to successfully break it up in this way. Or we might try to write $f = g - h$ where each of g and h is monotone, taking advantage of the fact that $\int_a^b f = \int_a^b g - \int_a^b h$. In theory, a large class of functions can be written as the difference of two increasing functions – in particular every function with a continuous derivative can be expressed this way – but in practice finding the g and h is usually not easy.

Why are we avoiding trying to compute the Darboux integral of a non-monotone function $f : [a, b] \to \mathbb{R}$? Because if f is not monotone, then for any partition $\mathcal{P} = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$, then order to compute $U(f, \mathcal{P}_N)$ or $L(f, \mathcal{P}_n)$ we have to solve n optimization problems: we have to maximize (resp. minimize) f on each subinterval $[x_i, x_{i+1}]$. That doesn't sound fun. But we have a more basic issue: which partitions $\mathcal{P}$ should we be using? Darboux integrability means that for each $\epsilon > 0$ there is *some* partition $\mathcal{P}_\epsilon$ of $[a, b]$ for which $\Delta(f, \mathcal{P}_\epsilon) < \epsilon$. It doesn't tell us how to find $\mathcal{P}_\epsilon$. Geometric intuition (recall we have been assuming that f is continuous) suggests we should as in the monotone case be able to use the uniform partitions $\mathcal{P}_n$ for sufficiently large n, or in other words that we should again have

$$\lim_{n\to\infty} \Delta(f, \mathcal{P}_n) = 0, \text{ hence } \lim_{n\to\infty} U(f, \mathcal{P}_n) = \lim_{n\to\infty} L(f, \mathcal{P}_n) = \int_a^b f.$$

We still have the darned upper and lower sums, but....actually, it is clear that the left endpoint sum $\frac{b-a}{n} \sum_{i=0}^{n-1} f(a + i(\frac{b-a}{n}))$ lies in between $L(f, \mathcal{P}_n)$ and $U(f, \mathcal{P}_n)$, so by the Squeeze Theorem for sequences it would indeed then follow that

$$\lim_{n\to\infty} \frac{b-a}{n} \sum_{i=0}^{n-1} f(a + i(\frac{b-a}{n})) = \int_a^b f.$$

Riemann's work shows that all of these things are true and more. There are two key ideas that distinguish Riemann's integral from Darboux's. First, instead of upper and lower sums we work with sums obtained by taking the height of the rectangle to be *any* point in the subinterval $[x_i, x_{i+1}]$. The second idea is that his notion of convergence is *a priori* more demanding than Darboux's in a way that works against you if you are trying to show that a given function is integrable but works for you if you know that it is.

We begin with a function $f : [a, b] \to \mathbb{R}$ and a partition $\mathcal{P} = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ of $[a, b]$, but now we introduce one more piece of data, a **tagging** of $\mathcal{P}$. A tagging is a function $\tau : \{0, 1, \ldots, n\} \to [a, b]$ such that for all i, the point $\tau(i)$ lies in the ith subinterval $[x_i, x_{i+1}]$ determined by the partition $\mathcal{P}$. Instead of using functional notation we may just put $x_i^* = \tau(i)$, and then a tagging is a finite sequence $\{x_0^* \leq x_1^* \leq \ldots \leq x_{n-1}^* \leq x_n^\}$. Notice that this sequence need not be *quite* injective: we could have $x_i^* = x_{i+1}^*$; this holds if and only if both are equal to x_{i+1}, which is both the right endpoint of $[x_i, x_{i+1}]$ and the left endpoint of $[x_{i+1}, x_{i+2}]$. The pair $(\mathcal{P}, \tau)$ is called a **tagged partition**.

To any tagged partition $(\mathcal{P}, \tau)$ and, of course, a function $f : [a, b] \to \mathbb{R}$ we associate a **Riemann sum**

$$R(f, \mathcal{P}, \tau) := \sum_{i=0}^{n-1} f(x_i^*)(x_{i+1} - x_i).$$

It is easy to compare with the upper and lower sums: of course we have

$$\sup(f([x_i, x_{i+1}]) \geq f(x_i^*) \text{ and } \inf(f[x_i, x_{i+1}]) \leq f(x_i^*),$$

so

$$L(f, \mathcal{P}) \leq R(f, \mathcal{P}, \tau) \leq U(f, \mathcal{P}).$$

If f assumes its maximum and minimum value on each subinterval $[x_i, x_{i+1}]$ – so for instance if f is continuous – then $L(f, \mathcal{P})$ and $U(f, \mathcal{P})$ are themselves Riemann sums. In general this is not quite true because the suprema and infima need not be attained, but almost: we will have

$$U(f, \mathcal{P}) = \sup_\tau R(f, \mathcal{P}, \tau) \text{ and } L(f, \mathcal{P}) = \inf_\tau R(f, \mathcal{P}, \tau).$$

Thus for each partition $\mathcal{P}$, the upper sum is the least upper bound of all possible Riemann sums for $\mathcal{P}$ and the lower sum is the greatest lower bound of all possible Riemann sums for $\mathcal{P}$. This is quite clear if f is bounded; it is still true if f is unbounded, but it requires a bit more work:

PROPOSITION 2.10. *Let $f : [a, b] \to \mathbb{R}$ be a function and $\mathcal{P}$ a partition of $[a, b]$.*

a) *If f is unbounded above, then as we range over all possible taggings τ of $[a, b]$, we have*

$$\sup_\tau R(f, \mathcal{P}, \tau) = \infty.$$

b) *If f is unbounded below, then as we range over all possible taggings τ of $[a, b]$, we have*

$$\inf_\tau R(f, \mathcal{P}, \tau) = -\infty.$$

We leave the proof of Proposition 2.10 as Exercise 2.12.

So far this is all pretty similar to Darboux's treatment. The second main idea is that the sense in which the Riemann sums are required to converge to $\int_a^b f$ is quite stringent. To give it, we need just one more definition: for a partition $\mathcal{P} = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ of $[a, b]$, its **mesh** is

$$|\mathcal{P}| := \max_{0 \leq i \leq n-1} x_{i+1} - x_i;$$

that is, the mesh of $\mathcal{P}$ is the largest length of a subinterval $[x_i, x_{i+1}]$. For instance, in the uniform partition $\mathcal{P}_n$ all subintervals have length $\frac{b-a}{n}$, so its mesh is $|\mathcal{P}_n| = \frac{b-a}{n}$.

A function $f : [a, b] \to \mathbb{R}$ is **Riemann integrable** if there is $S \in \mathbb{R}$ such that: for all $\epsilon > 0$, there is $\delta > 0$ such that for every partition $\mathcal{P}$ of $[a, b]$ with mesh $|\mathcal{P}| \leq \delta$ and every tagging τ of $\mathcal{P}$, we have

$$|R(f, \mathcal{P}, \tau) - S| \leq \epsilon.$$

We then put $\int_a^b f := S$.

Let us check that Riemann integrability implies Darboux integrability: let $\epsilon > 0$.

Then there is $\delta > 0$ such that for any partition $\mathcal{P}$ of mesh less than δ we have $|R(f, \mathcal{P}, \tau) - S| \leq \frac{\epsilon}{2}$, which of course means that $R(f, \mathcal{P}, \tau) \in [S - \frac{\epsilon}{2}, S + \frac{\epsilon}{2}]$. Because the upper sum $U(f, \mathcal{P})$ is the supremum of the $R(f, \mathcal{P}, \tau)$'s as we range over τ and $R(f, \mathcal{P}, \tau) \leq S + \frac{\epsilon}{2}$ for all τ, we get $U(f, \mathcal{P}) \leq S + \frac{\epsilon}{2}$; similarly, we get $L(f, \mathcal{P}) \geq S - \frac{\epsilon}{2}$, so $\Delta(f, \mathcal{P}) \leq \epsilon$, and thus f is Darboux integrable and moreover S is the Darboux integral $\int_a^b f$. (In particular, there is at most one $S \in \mathbb{R}$ satisfying the conditions in the definition of Riemann integrability.)

It is **much less obvious** whether every Darboux integrable function is Riemann integrable. But happily it is true:

THEOREM 2.11.

a) *For a function $f : [a, b] \to \mathbb{R}$, the following are equivalent:*
   (i) *The function f is Darboux integrable.*
   (ii) *The function f is Riemann integrable.*
   (iii) *For every sequence $\{(\mathcal{P}_n, \tau_n)\}_{n=1}^{\infty}$ of tagged partitions of $[a, b]$ with $|\mathcal{P}_n| \to 0$, the sequence $\{R(f, \mathcal{P}_n, \tau_n)\}_{n=1}^{\infty}$ of Riemann sums is convergent.*

b) *If the equivalent conditions of part a) hold, then for any sequence $\{(\mathcal{P}_n, \tau_n)\}$ of tagged partitions of $[a, b]$ with $|\mathcal{P}_n| \to 0$, we have*

$$\lim_{n \to \infty} R(f, \mathcal{P}_n, \tau_n) = \int_a^b f.$$

We are not going to prove Theorem 2.11 in our course, but you can find the proof in [**HC**, §8.4]. So that you don't feel short-changed, let me mention that most undergraduate analysis texts do not prove this theorem; many of them just develop Darboux's integral and forget to make the connection with Riemann sums.

Let us sum up the state of affairs: because of Theorem 2.11, the Darboux integral and the Riemann integral, although they were defined differently, turn out to be completely equivalent: a function is integrable in sense if and only if it is in the other sense, and if so they return the same real number $\int_a^b f$. So we no longer need to distinguish between them: henceforth we will *only* speak of Riemann integrable functions and the Riemann integral. This is what is most commonly done, even by people who have much less right to conflate the two than we do.

We end this section with one more result that helps to make the Riemann integral more computable.

THEOREM 2.12. *Let $f : [a, b] \to \mathbb{R}$ be differentiable with bounded derivative: let $M > 0$ be such that $|f'| \leq M$. For $n \in \mathbb{Z}^+$, let*

$$L_n(f) = \sum_{i=0}^{n-1} f(a + i(\frac{b-a}{n}))(\frac{b-a}{n})$$

*be the **left endpoint Riemann sum** of f. Then*

$$\left| \int_a^b f - L_n(f) \right| \leq \left(\frac{(b-a)^2 M}{2} \right) \frac{1}{n}.$$

PROOF. Step 1: We establish the result for $n = 1$. For $x \in [a, b]$, we apply the Mean Value Theorem to f on the interval $[a, x]$: there is $c \in (a, x)$ with

$$f(x) - f(a) = f'(c)(x - a).$$

Since $|f'(c)| \leq M$, we get

$$-M(x - a) + f(a) \leq f(x) \leq M(x - a) + f(a)$$

and thus

$$\int_a^b (-M(x - a) + f(a)) \leq \int_a^b f \leq \int_a^b (M(x - a) + f(a)).$$

Of course we can evaluate the first and last integrals with the Fundamental Theorem of Calculus, and we get

$$\frac{-M}{2}(b - a)^2 + (b - a)f(a) \leq \int_a^b f \leq \frac{M}{2}(b - a)^2 + (b - a)f(a),$$

which is equivalent to

$$\left| \int_a^b f - L_1(f) \right| \leq \frac{M}{2}(b - a)^2.$$

Step 2: Let $n \in \mathbb{Z}^+$. For $0 \leq i \leq n - 1$, put $x_i^* = a + i\frac{b-a}{n}$. Then:

$$\begin{aligned} \left| \int_a^b f - L_n(f) \right| &= \left| \sum_{i=0}^{n-1} \left(\int_{x_i^*}^{x_{i+1}^*} f - f(x_i^*) \left(\frac{b - a}{n} \right) \right) \right| \\ &\leq \sum_{i=0}^{n-1} \left| \int_{x_i^*}^{x_{i+1}^*} f - f(x_i^*) \left(\frac{b - a}{n} \right) \right|. \end{aligned}$$

Step 1 applies to each term in the last sum to give

$$\left| \int_a^b f - L_n(f) \right| \leq \sum_{n=0}^{n-1} \frac{M}{2} \left(\frac{b - a}{n} \right)^2 = \left(\frac{(b - a)^2 M}{2} \right) \frac{1}{n}. \qquad \square$$

Whereas Theorem 2.11 guarantees us that for any Riemann integral $f : [a, b] \to \mathbb{R}$, we can compute $\int_a^b f$ as the limit $\lim_{n\to\infty} L_n(f)$ of the left endpoint Riemann sums, Theorem 2.12 gives us, for functions with a bounded derivative, a precise **error estimate**: it tells us how large n needs to be in order to for $L_n(f)$ to compute $\int_a^b f$ to any prescribed accuracy, where the bound depends on the size of the derivative. To get a bound on f' essentially amounts to solving one optimization problem, which is great progress over the arbitrarily many optimization problems we had to solve to compute a single upper or lower sum. More basically, this result is telling us that the faster f is changing from point to point in the local sense, the more sample points we will need in order to get a handle on $\int_a^b f$: this makes a lot of sense. On the other hand, of course if we don't know anything about f other than that it is, say, differentiable, then we don't know how many sample points we will need to use to usefully approximate $\int_a^b f$ because for any sample points we choose, for all we know f could be oscillating wildly in between them.

Moreover Theorem 2.12 is the first of an infinite sequence of theorems: the rough form of the kth theorem in the sequence is that if we assume that the kth derivative $f^{(k)}$ of f exists and is bounded, then using the values of f at the points of

the partition $\mathcal{P}_n$ of $[a,b]$ into n equally spaced subintervals, one can write down a certain finite sum $S_{k,n}(f)$ that is a certain weighted average of several different Riemann sums, and has the property that

$$\left| S_{k,n}(f) - \int_a^b f \right| \leq C \frac{1}{n^k}.$$

Here C is a certain explicit expression depending only on $(b-a)$, the number n of sample points and an upper bound M for $|f^{(k)}|$. Thus, the more smoothness we assume on f, the more rapidly we can compute $\int_a^b f$. The $k=2$ case is the Trapezoidal Rule [**HC**, Thm. 9.5], while the $k=3$ case is Simpson's rule [**HC**, Thm. 9.8]. The branch of mathematics in which you will learn how this works for all $k \in \mathbb{Z}^+$ and many other similar results is **numerical analysis**.

Exercises.

Exercise 2.12. *Prove Proposition 2.10.*

Problems.

Problem 2.8. *We will show that* $\lim_{n\to\infty} \sum_{k=1}^n \frac{n}{k^2+n^2} = \frac{\pi}{4}$.

a) *Let* $\mathcal{P}_n$ *be the partition of* $[0,1]$ *into* n *equally spaced subintervals. Let* $f : [0,1] \to \mathbb{R}$ *by* $f(x) = \frac{1}{x^2+1}$. *Show:* $\sum_{k=1}^n \frac{n}{k^2+n^2} = R(f, \mathcal{P}_n, \tau_n)$, *where* τ_n *is the right endpoint tagging: for all* $0 \leq i \leq n-1$, $x_i^* = \frac{i}{n}$.

b) *Use Theorem 2.11 and the Fundamental Theorem of Calculus to evaluate* $\lim_{n\to\infty} R(f, \mathcal{P}_n, \tau_n)$.

4. The Class of Riemann Integrable Functions

4.1. More Riemann Integrable Functions. We had a big fish to catch: the existence of an antiderivative for any continuous function. So we built a big net — the Darboux integral — and with that big net we caught our fish. (Then we discussed the construction of a second net — the Riemann integral — that looked rather different from our first net, but we found that in the end the second net catches exactly the same fish as the first. So we stopped distinguishing between the two nets.) It is now time to ask: what other fish have we caught? That is, what can we say about the class $\mathcal{R}[a,b]$ of Riemann integrable $f : [a,b] \to \mathbb{R}$? Again, we know that this class contains all continuous functions, and we also know that every function in the class is bounded.

Problems 2.3 and 2.4 give some instances of functions that are discontinuous but Riemann integrable. Let us concentrate on the latter: according to 2.4, if $f : [a,b] \to \mathbb{R}$ is bounded and has only finitely many discontinuities, then f is Riemann integrable. Let us sketch a proof: let $M > 0$ be such that $|f| \leq M$, fix $\delta > 0$, and choose a partition $\mathcal{P}$ of $[a,b]$ that contains, for each point c of disconinuity of f — to fix ideas, let us assume that the discontinuities occur at interior points of $[a,b]$ – there are consecutive elements $x_i, x_{i+1} \in \mathcal{P}$ with $x_{i+1} - x_i < \delta$. If we remove the open intervals (x_i, x_{i+1}) from $[a,b]$, we get a finite union of closed subintervals — suppose that there are N of them — such that f is continuous on each one, hence Riemann integrable. This means that for any $\epsilon > 0$ we can refine $\mathcal{P}$ to a partition $\mathcal{P}_\epsilon$ such that on the Nth subinterval, the difference between the upper sum and the

lower sum is at most $\frac{\epsilon}{2N}$, so therefore the sum of the differences of the lower sums is at most $\frac{\epsilon}{2}$. Finally, on each subinterval $[x_i, x_{i+1}]$ we have

$$\Delta(f|_{[x_i,x_{i+1}]}, \mathcal{P}_\epsilon) < 2M\delta.$$

This is because since $|f| \leq M$, its oscillation – i.e., its supremum minus its infimum – is at most $2M$, so we multiply this by the length of the subinterval. Thus overall for this partition $\mathcal{P}_\epsilon$ we find that

$$\Delta(F, \mathcal{P}_\epsilon) < \frac{\epsilon}{2} + 2NM\delta.$$

Since M and N are fixed, we can choose δ sufficiently small so that $2NM\delta < \frac{\epsilon}{2}$, and we win: f is integrable by Darboux's Criterion.

So now we are interested in bounded function $f : [a, b] \to \mathbb{R}$ with *infinitely many* discontinuities. At first glance, such a function looks unlikely to be Riemann integrable, at least to me: by Exercise 1.41, the set of discontinuities of f must have an accumulation point in $[a, b]$, and that seems like it could screw things up – arguments like the one we made for finitely many discontinuities are not going to succeed. (Anyway, in our argument above the *number* N of discontinuities appeared in our bound; if there are infinitely many discontinuities, we certainly cannot do this.) However, again some previous exercises show that we've caught profoundly more fish than we thought: Problem 2.6 shows that every monotone function $f : [a, b] \to \mathbb{R}$ is Riemann integrable. That is not so surprising, but Problem 2.7 is: there is a strictly increasing function $f : [a, b] \to \mathbb{R}$ that is discontinuous at every rational point of $[a, b]$! Thus a bounded function can be Riemann integrable even when its set of discontinuities is *dense* in $[a, b]$.

The following result further exhibits the largeness of the class of Riemann integrable functions.

THEOREM 2.13. *Let $f : [a, b] \to [c, d]$ be Riemann integrable, and let $g : [c, d] \to \mathbb{R}$ be continuous. Then the composite function $g \circ f : [a, b] \to \mathbb{R}$ is Riemann integrable.*

We are going to omit the proof of this result because of time constraints and because it is a bit technical: see [**HC**, Thm. 8.17]. It becomes easier in an important special case. For a subset X of $\mathbb{R}^N$, a function $f : X \to \mathbb{R}^M$ is **Lipschitz** if there is a constant $C \in (0, \infty)$ such that

$$\forall x_1, x_2 \in X, \ ||f(x_1) - f(x_2)|| \leq C||x_1 - x_2||.$$

A C that works here is called a **Lipschitz constant** for f. You should think of Lipschitz as a kind of "super-continuity": such functions are uniformly continuous with $\delta = \frac{\epsilon}{C}$. The following result — the second part of which appeared on the 2022 midterm! — showed that this property, although strong, arises in nature.

PROPOSITION 2.14. *Let I be an interval, and let $f : I \to \mathbb{R}$ be differentiable.*

a) *If f' is bounded, then f is Lipschitz.*
b) *If $I = [a, b]$ and f' is continuous, then f is Lipschitz.*

PROOF. a) Let $M > 0$ such that $|f'| \leq M$. Let $x_1 < x_2$ be elements of I. By the Mean Value Theorem there is $c \in (x_1, x_2)$ such that

$$|f(x_1) - f(x_2)| = |f'(c)||x_1 - x_2| \leq M|x_1 - x_2|.$$

Thus M is a Lipschitz contant for f.
b) If $f' : [a, b] \to \mathbb{R}$ is continuous, then by the Extreme Value Theorem, f' is bounded, so part a) applies to show that f is Lipschitz. $\square$

It turns out to be *much* easier to show Theorem 2.13 if we assume that g is not only continuous but Lipschitz: this is Exercise 2.14. Here is a nice consequence:

THEOREM 2.15. *If $f, g : [a, b] \to \mathbb{R}$ are both Riemann integrable, then so is $f \cdot g$.*

PROOF. If $h : [a, b] \to \mathbb{R}$ is any Riemann integrable function, then it follows from Theorem 2.13 that h^2 is also Riemann integrable. In fact, since h is bounded – say $|h| \leq M$ – by Proposition 2.14 we have that $x^2 : [-M, M] \to \mathbb{R}$ is Lipschitz, so this lies in the part of Theorem 2.13 that we (the student who solves the right exercises and I) have proved. Now here is a dirty trick;

$$fg = \frac{(f+g)^2 - f^2 - g^2}{2},$$

which shows that fg is Riemann integrable, since we know that linear combinations of Riemann integrable functions are Riemann integrable and that squares of Riemann integrable functions are Riemann integrable. $\square$

4.2. The Riemann-Lebesgue Criterion. In fact there is a precise characterization of which bounded functions $f : [a, b] \to \mathbb{R}$ are Riemann integrable. This is usually called **Lebesgue's Criterion**, after the leading mathematician Henri Lebesgue who constructed a superior version of the integral to Riemann's. (Lebesgue is also the founder of the subject of measure theory referred to above. Most of Math 8100 concerns measure theory and the Lebesgue integral.) However my former colleague Roy Smith showed me exactly where this criterion occurs in a work of Riemann, so I will call it the Riemann-Lebesgue criterion.

We actually need a tiny piece of measure theory even to state this criterion, namely we need the notion of a subset X of $\mathbb{R}$ having **measure zero**. For this, let $\{[a_n, b_n]\}_{n=1}^{\infty}$ be a sequence of closed bounded intervals. We say that this sequence **covers** X if

$$X \subseteq \bigcup_{n=1}^{\infty} [a_n, b_n],$$

or in words, if every element of X lies in at least one of the subintervals $[a_n, b_n]$. To this sequence of intervals we attach a **total length**

$$\mathbb{L}(\{[a_n, b_n]\} := \sum_{n=1}^{\infty} (b_n - a_n) \in [0, \infty].$$

In other words, we really do just add up the lengths of the subintervals. This is an infinite series with non-negative terms, so it either converges or diverges to ∞. The idea is that, if the total length is finite, it should give an upper bound on the "length" of X. This intriguing idea is the beginning of measure theory, but we only need this one thing: we say that X has **measure zero** if for all $\epsilon > 0$, there is a covering $\{[a_n, b_n]\}$ of X of total length at most ϵ.

This concept is addressed in the exercises. It is pretty clear that every finite subset of $\mathbb{R}$ has measure zero: if we are allowed to use degenerate closed intervals $[a, a]$ this is truly obvious, but actually in the definition of measure zero it doesn't matter

whether we are allowed this or not, so it might be more educational to always use intervals of positive length. More generally, we say that a subset $X \subseteq \mathbb{R}$ is **countable** if there is a surjective sequence in X, i.e., a surjective function $x_\bullet : \mathbb{Z}^+ \to X$. This includes finite subsets, certainly. Moreover, each of $\mathbb{Z}^+$, $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ are countable (these are developed in the exercises). Then any countable subset X has measure zero – again, this is clear if we can use intervals $[a, a]$ but is not much harder to show even if we can't. There are also uncountable subsets of measure zero: one very famous one, the **Cantor set**, is developed in the exercises.

And here is the result:

THEOREM 2.16 (Riemann-Lebesgue Criterion). *For a function $f : [a, b] \to \mathbb{R}$, the following are equivalent:*

(i) *f is Riemann integrable.*
(ii) *f is bounded, and its set of discontinuities has measure zero.*

We will not give a proof of Theorem 2.16. Most proofs use somewhat more advanced material, but this is not necessary: see [**HC**, §8.5] for a proof that you have all the prerequisites to read.

Nevertheless we can stop to appreciate Theorem 2.16: it tells us *exactly* what fish we've caught with our integral! Moreover, if you know this result than many of our other results on Riemann integrability follow immediately. It is easy to see from the definition of measure zero that a finite union of sets, each of measure zero, also has measure zero. (It is not much harder to see that moreover if $\{X_n\}_{n=1}^\infty$ is an infinite sequence of sets of measure zero then $\bigcup_{n=1}^\infty X_n$ also has measure zero...but we don't need this here.)

So: let $f, g : [a, b] \to \mathbb{R}$ be Riemann integrable, so each is bounded and is discontinuous only a set of measure zero. Then:

• For $\alpha \in \mathbb{R}$, αf is bounded and has the same discontinuities as f, so is Riemann integrable.
• $f + g$ is bounded (if $|f| \leq M_1$ and $|g| \leq M_2$ then $|f + g| \leq M_1 + M_2$). If the set of discontinuties of f is X_f and the set of discontinuities of g is X_g, then the set of discontinuities of $f + g$ is contained in $X_f \cup X_g$, so has measure zero. So $f + g$ is Riemann integrable.
• Almost the identical argument works to show that $f \cdot g$ is Riemann integrable (only modification: if $|f| \leq M_1$ and $|g| \leq M_2$, then $|fg| \leq M_1 M_2$).
• If f is monotone, then it is bounded – $f([a, b])$ lies in the interval in between $f(a)$ and $f(b)$ – and it can be shown that the set of discontinuities of f is *countable*. So f is Riemann integrable.
• Suppose $f : [a, b] \to [c, d]$ and $g : [c, d] \to \mathbb{R}$ is continuous. Then g is bounded, hence so is $g \circ f$. Moreover, since g is continuous, the set of discontinuities of $g \circ f$ is contained in the set of discontinuities of f, and a subset of a set of measure zero has measure zero. So $g \circ f$ is Riemann integrable.

Exercises.

EXERCISE 2.13. *Let $f : I \to [c, d]$ be a bounded function, and let $g : [c, d] \to \mathbb{R}$ be a Lipschitz function with Lipschitz constant C. Show:*

$$\omega(g \circ f, I) \leq C\omega(f, I).$$

EXERCISE 2.14. *Let $f : [a, b] \to [c, d]$ be Riemann integrable, and let $g : [a, b] \to \mathbb{R}$ be Lipschitz with Lipschitz constant C. Show that $g \circ f : [a, b] \to \mathbb{R}$ is Riemann integrable as follows: let $\epsilon > 0$, and choose a partition $\mathcal{P}_\epsilon$ for which $\Delta(f, \mathcal{P}_\epsilon) < \frac{\epsilon}{C}$. Use Exercise 2.13 to show that $\Delta(g \circ f, \mathcal{P}_\epsilon) < \epsilon$.*

A nonempty set X is **countable** if there is a surjective function $f : \mathbb{Z}^+ \to X$. By definition, the empty set is also countable.

EXERCISE 2.15. *Let X be a set.*

a) *Show: if X is finite, then X is countable.*
b) *Show: if X is infinite and countable, then there is a bijection $f : \mathbb{Z}^+ \to X$.*
c) *Show: if X is countable and $f : X \to Y$ is a surjection, then also Y is countable.*

Thus every countable set is in bijection with exactly one of the following: (i) the empty set; (ii) the set $\{1, \ldots, n\}$ for some $n \in \mathbb{Z}^+$; or (iii) $\mathbb{Z}^+$.

EXERCISE 2.16. a) *Show: every subset of a countable set is countable.*
b) *Let $\iota : Y \to X$ be an injective function. Show: if X is countable, then so is Y.*

EXERCISE 2.17. *Show: if $n \in \mathbb{Z}^+$ and $X_1, \ldots, X_n$ are countable sets, then their union $\bigcup_{n=1}^{\infty}$ is countable.*

EXERCISE 2.18.

a) *Let $X \subseteq \mathbb{R}$ be a countable subset. Show: X has measure zero.*
b) *Deduce from the Riemann-Lebesgue Criterion that $\mathbb{R}$ is uncountable. (Hint: if $\mathbb{R}$ were countable, then every bounded function would be Riemann integrable.)*

Problems.

PROBLEM 2.9. *Let $f : [a, b] \to \mathbb{R}$ be Riemann integrable.*

a) *Show that $|f| : [a, b] \to \mathbb{R}$ is Riemann integrable. (Suggestions: the absolute value function is Lipschitz, so you can apply Exercise 2.14. Or you can show that for any subinterval I of $[a, b]$ we have $\omega(|f|, I) \leq \omega(f, I)$.)*
b) *Show the **Integral Triangle Inequality**:*

$$\left| \int_a^b f \right| \leq \int_a^b |f|.$$

PROBLEM 2.10. a) *Let X and Y be sets. Show: if X is countable and there is a surjection $f : X \to Y$, then Y is countable.*
b) *Show: $\mathbb{Z}^+ \times \mathbb{Z}^+$ is countable.*
c) *Show: the set $\mathbb{Q}$ of rational numbers is countable. (Suggestion: since $\mathbb{Q} = \mathbb{Q}^{>0} \cup \{0\} \cup \mathbb{Q}^{<0}$ and multiplication by -1 gives a bijection from $\mathbb{Q}^{>0}$ to $\mathbb{Q}^{<0}$, by Exercise 2.15 it is enough to show that $\mathbb{Q}^{>0}$ is countable. Do this by finding a surjective function $f : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{Q}$.)*

Exercise 2.18 gives a proof of the uncountability of $\mathbb{R}$ that is striking, but is also overkill: the Riemann-Lebesgue Criterion is a difficult result whose proof we have omitted. The next exercise outlines a classic proof (due to G. Cantor) of the uncountability of $\mathbb{R}$.

PROBLEM 2.11. *In this exercise we refer to decimal expansions of real numbers. Some real numbers have a unique decimal expansion, but others have (exactly) two different decimal expansions: a real number admitting a decimal expansion ending with all* 0*'s also has a decimal expansion ending with all* 9*'s. For the sake of definiteness, when we refer to "the decimal expansion" of* $x \in \mathbb{R}$ *we will exclude an expansion ending with all* 9*'s.*

Let $f : \mathbb{Z}^+ \to \mathbb{R}$ *be any function, and put* $x_n := f(n)$*. Build a real number* $x = 0.d_1 d_2 \cdots d_n \cdots \in [0,1]$ *as follows: for all* $n \in \mathbb{Z}^+$*, the* n*th decimal digit* d_n *of* x *is different from the* n*th decimal digit of* x_n *and also different from* 0 *and* 9*. (This still leaves us at least* 7 *choices.) Show: for no* $n \in \mathbb{Z}^+$ *do we have* $x = x_n$*, and deduce that* f *is not surjective.*

PROBLEM 2.12. *Let* I *be an interval, and let* $f : I \to \mathbb{R}$ *be a monotone function. Let* X *be the set of* $c \in I$ *such that* f *is discontinuous at* c*. Show:* X *is countable. (Suggestion: we may assume* f *is increasing. An increasing function* f *can only be discontinuous at* c *if* $\lim_{x \to c^-} f(x) < \lim_{x \to c^+} f(x)$*. If so, there is a rational number lying strictly in between the left hand limit and the right hand limit. Use this to build an injective function* $\iota : X \to \mathbb{Q}$ *and then apply Exercise 2.16b).)*

Compare Problem 2.7 with Problem 2.12. Things are getting subtle: monotone functions can have infinitely many discontinuities on a bounded interval, but still their set of discontinuities is "small" in a strong sense.

5. Some Further Exercises and Problems

A subset $X \subseteq \mathbb{R}^N$ is called a $\mathbf{G}\delta$ set if it is a countable intersection of open sets: that is, there is a sequence $\{U_n\}_{n=1}^\infty$ of open subsets of $\mathbb{R}^N$ such that $X = \bigcap_{n=1}^\infty U_n$.[2] A subset $X \subseteq \mathbb{R}^N$ is called a $\mathbf{F}\sigma$ set if it is a countable union of closed sets: that is, there is a sequence $\{C_n\}_{n=1}^\infty$ of closed subsets of $\mathbb{R}^N$ such that $X = \bigcup_{n=1}^\infty C_n$.[3]

Exercises.

EXERCISE 2.19. *Let* $X \subseteq \mathbb{R}^N$*.*

a) *Show that* X *is a* $G\delta$ *set if and only if its complement* $\mathbb{R}^N \setminus X$ *is a* $G\delta$ *set and that* X *is an* $F\sigma$ *set if and only if its complement* $\mathbb{R}^N \setminus X$ *is a* $G\delta$*-set.*
b) *Show: if* X *is open, then* X *is a* $G\delta$ *set.*
c) *Show: if* X *is closed, then* X *is an* $F\sigma$ *set.*
d) *Show:* $\mathbb{Q}$ *is an* $F\sigma$ *set in* $\mathbb{R}$*. Is it a* $G\delta$ *set?*[4]

EXERCISE 2.20. *We would like to define an* $F\sigma$ *subset of* $[a,b]$*. There are two plausible ways to do this:*

[2]This strange terminology is over a a hundred years old. Wikipedia tells us that it comes from the German words *gebiet* (open set) and *durchschnitt* (intersection).

[3]Wikipedia tells us that *this* term comes fom the French words *fermé* (closed) and *somme* (sum, union).

[4]This is asked as a question because unless you know a certain theorem that is not mentioned in these notes, it is *very* hard to answer.

(i) *As an $F\sigma$ subset X of $\mathbb{R}$ that moreover is a subset of $[a,b]$.*
(ii) *As a subset of the form $X \cap [a,b]$, for $X \subseteq \mathbb{R}$ an $F\sigma$ set.*

Show that these define the same class of subsets of $[a,b]$, which we will henceforth call $F\sigma$ subsets of $[a,b]$.
(Remark/hint: you only need to use that $[a,b]$ is a closed subset of $\mathbb{R}$.)

Problems.

PROBLEM 2.13. *Let $X \subseteq \mathbb{R}^N$.*

a) *Show: if X is closed, then X is a $G\delta$ set.*
(Suggestion: for $\epsilon > 0$, let
$$U_\epsilon(X) := \{a \in \mathbb{R}^N \mid d(a,x) < \epsilon \text{ for some } x \in X\}.$$
Show: $U_\epsilon(X)$ is open and $\bigcap_{n=1}^\infty U_{\frac{1}{n}}(X) = X$.)
b) *Deduce: if X is open, then X is an $F\sigma$ set.*

PROBLEM 2.14. *Let $f : \mathbb{R}^N \to \mathbb{R}$.*

a) *Define the **oscillation of f at x** as*
$$\omega_x(f) := \inf_\delta \sup_{y \in B^\circ(x,\delta)} |f(y) - f(x)| \in [0,\infty].$$
Show: f is continuous at x if and only if $\omega_x(f) = 0$.
b) *For $\epsilon > 0$, show: $\{x \in \mathbb{R}^N \mid \omega_x(f) < \epsilon\}$ is open.*
c) *Define the **locus of continuity** of f as*
$$L(f) := \{x \in \mathbb{R}^N \mid f \text{ is continuous at } x\}.$$
Show: $L(f)$ is a $G\delta$ set.
d) *Deduce: the set of points at which f is **dis**continuous is an $F\sigma$ set.*

PROBLEM 2.15. *Prove the following theorem of Young (1903) and Lebesgue (1905): let $X \subseteq \mathbb{R}^N$ be any $G\delta$-set. Then there is a function $f : \mathbb{R}^N \to \mathbb{R}$ with locus of continuity $L(f) = X$. Deduce: the set of discontinuities of a function $f : \mathbb{R}^N \to \mathbb{R}$ can be any $F\sigma$ set in $\mathbb{R}^N$.*

PROBLEM 2.16.

a) *Show that for a subset $X \subseteq [a,b]$, the following are equivalent:*
(i) *X is an $F\sigma$ subset of $[a,b]$ (cf. Exercise 2.20).*
(ii) *There is a function $f : [a,b] \to \mathbb{R}$ such that*
$$\{x \in [a,b] \mid f \text{ is discontinuous at } x\} = X.$$
(iii) *There is a* bounded *function $f : [a,b] \to \mathbb{R}$ such that*
$$\{x \in [a,b] \mid f \text{ is discontinuous at } x\} = X.$$
b) *Deduce: there is a Riemann integrable function $f : [a,b] \to \mathbb{R}$ that is discontinuous at uncountably many points of $[a,b]$.*

CHAPTER 3

Sequences and Series of Functions

1. Pointwise Convergence

1.1. Cautionary tales.

Let X be a set. (The case of most interest to us will be that in which X is a subset of $\mathbb{R}^N$ or better still, of $\mathbb{R}$..) A **sequence of real functions** is a sequence $\{f_n\}_{n=0}^{\infty}$ with each $f_n : X \to \mathbb{R}$ a real-valued function on X.

For us the following example is all-important: let

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

be a power series with radius of convergence $R > 0$. So f may be viewed as a function $f : (-R, R) \to \mathbb{R}$. Put $f_n = \sum_{k=0}^{n} a_k x^k$, so each f_n is a polynomial of degree at most n; therefore f_n makes sense as a function from $\mathbb{R}$ to $\mathbb{R}$, but let us restrict its domain to $(-R, R)$. Then we get a sequence of functions

$$\{f_n : (-R, R) \to \mathbb{R}\}_{n=0}^{\infty}.$$

Our goal is to show that the function f has many desirable properties: it is continuous and indeed infinitely differentiable, and its derivatives and antiderivatives can be computed term-by-term. Since the functions f_n have all these properties (and more – each f_n is a polynomial), it seems like a reasonable strategy to define some sense in which the sequence $\{f_n\}$ **converges** to the function f, in such a way that this converges process *preserves* the favorable properties of the f_n's.

The previous description perhaps sounds overly complicated and mysterious, since in fact there is an evident sense in which the sequence of functions f_n converges to f. Indeed, to say that x lies in the open interval $(-R, R)$ of convergence is to say that the sequence $f_n(x) = \sum_{k=0}^{n} a_k x^k$ converges to $f(x)$.

This leads to the following definition: if $\{f_n\}_{n=1}^{\infty}$ is a sequence of real functions defined on a set X and $f : X \to \mathbb{R}$ is another function, we say f_n **converges to f pointwise** if we have $f_n(x) \to f(X)$ for all $x \in I$. In this situation we also say f is the **pointwise limit** of the sequence $\{f_n\}$. In particular the sequence of partial sums of a power series converges pointwise to the power series on the interval I of convergence.

We have the closely related notion of an infinite series of functions $\sum_{n=0}^{\infty} f_n$ and of pointwise convergence of this series to some limit function f. Indeed, as in the case of just one series, we just define $S_n = f_0 + \ldots + f_n$ and say that $\sum_n f_n$

converges pointwise to f if the sequence of partial sums S_n converges pointwise to f.

The great mathematicians of the 17th, 18th and early 19th centuries encountered many sequences and series of functions (again, especially power series and Taylor series) and often did not hesitate to assert that the pointwise limit of a sequence of functions having a certain nice property itself had that nice property.[1] The problem is that statements like this unfortunately need not be true!

EXAMPLE 3.1. *Define $f_n = x^n : [0,1] \to \mathbb{R}$. Clearly $f_n(0) = 0^n = 0$, so $f_n(0) \to 0$. For any $0 < x \leq 1$, the sequence $f_n(x) = x^n$ is a geometric sequence with geometric ratio x, so that $f_n(x) \to 0$ for $0 < x < 1$ and $f_n(1) \to 1$. It follows that the sequence of functions $\{f_n\}$ has a pointwise limit $f : [0,1] \to \mathbb{R}$, the function which is 0 for $0 \leq x < 1$ and 1 at $x = 1$. Unfortunately the limit function is discontinuous at $x = 1$, despite the fact that each of the functions f_n are continuous. Thus:* ***The pointwise limit of a sequence of continuous functions need not be continuous****.*

Example 3.1 was chosen for its simplicity, not to exhibit maximum pathology. It is possible to construct a sequence $\{f_n\}_{n=1}^{\infty}$ of polynomial functions converging pointwise to a function $f : [0,1] \to \mathbb{R}$ that has infinitely many discontinuities![2]

One can also find assertions in the math papers of old that if f_n converges to f pointwise on an interval $[a,b]$, then $\int_a^b f_n dx \to \int_a^b f dx$. To a modern eye, there are in fact two things to establish here: first that if each f_n is Riemann integrable, then the pointwise limit f must be Riemann integrable. And second, that *if* f is Riemann integrable, its integral is the limit of the sequence of integrals of the f_n's. In fact *both* of these are false!

EXAMPLE 3.2. *Define a sequence $\{f_n\}_{n=0}^{\infty}$ with common domain $[0,1]$ as follows. Let f_0 be the constant function 1. Let f_1 be the function which is constantly 1 except $f(0) = f(1) = 0$. Let f_2 be the function which is equal to f_1 except $f(1/2) = 0$. Let f_3 be the function which is equal to f_2 except $f(1/3) = f(2/3) = 0$. And so forth. To get from f_n to f_{n+1} we change the value of f_n at the rational numbers $\frac{a}{n}$ in $[0,1]$ from 1 to 0. Each f_n is equal to 1 except at a finite set of points, hence bounded with only finitely many discontinuities, hence Riemann integrable.*

The functions f_n converges pointwise to a function f which is 1 on every irrational point of $[0,1]$ and 0 on every rational point of $[0,1]$. Since every open interval (a,b) contains both rational and irrational numbers, the function f is not Riemann integrable: for any partition of $[0,1]$ its upper sum is 1 and its lower sum is 0. Thus a pointwise limit of Riemann integrable functions need not be Riemann integrable.

EXAMPLE 3.3. *We define a sequence of functions $f_n : [0,1] \to \mathbb{R}$ as follows: $f_n(0) = 0$, and $f_n(x) = 0$ for $x \geq \frac{1}{n}$. On the interval $[0, \frac{1}{n}]$ the function forms a*

[1]This is an exaggeration. The precise definition of convergence of real sequences did not come until the work of Weierstrass in the latter half of the 19th century. Thus mathematicians spoke of functions f_n "approaching" or "getting infinitely close to" a fixed function f. Exactly what they meant by this – and indeed, whether even they knew exactly what they meant (presumably some did better than others) is a matter of serious debate among historians of mathematics.

[2]On the other hand, it turns out that it is not possible for a pointwise limit of continuous functions to be discontinuous at *every* point. This is a theorem of R. Baire that belongs in a more advanced course [**GT**, Thm. 4.10].

"spike": $f(\frac{1}{2n}) = 2n$ and the graph of f from $(0,0)$ to $(\frac{1}{2n}, 2n)$ is a straight line, as is the graph of f from $(\frac{1}{2n}, 2n)$ to $(\frac{1}{n}, 0)$. In particular f_n is piecewise linear hence continuous, hence Riemann integable, and its integral is the area of a triangle with base $\frac{1}{n}$ and height $2n$: $\int_0^1 f_n dx = 1$. On the other hand this sequence converges pointwise to the zero function f. So

$$\lim_{n\to\infty} \int_0^1 f_n = 1 \neq 0 = \int_0^1 \lim_{n\to\infty} f_n.$$

EXAMPLE 3.4. *Let $g : \mathbb{R} \to \mathbb{R}$ be a bounded differentiable function such that $\lim_{n\to\infty} g(n)$ does not exist. (For instance, we may take $g(x) = \sin(\frac{\pi x}{2})$.) For $n \in \mathbb{Z}^+$, define $f_n(x) = \frac{g(nx)}{n}$. Let M be such that $|g(x)| \leq M$ for all $x \in \mathbb{R}$. Then for all $x \in \mathbb{R}$, $|f_n(x)| \leq \frac{M}{n}$, so f_n converges pointwise to the function $f(x) \equiv 0$ and thus $f'(x) \equiv 0$. In particular $f'(1) = 0$. On the other hand, for any fixed nonzero x, $f_n'(x) = \frac{ng'(nx)}{n} = g'(nx)$, so*

$$\lim_{n\to\infty} f_n'(1) = \lim_{n\to\infty} g'(n) \text{ does not exist.}$$

Thus

$$\lim_{n\to\infty} f_n'(1) \neq (\lim_{n\to\infty} f_n)'(1).$$

A common theme in all these examples is the **interchange of limit operations**: that is, we have some other limiting process corresponding to the condition of continuity, integrability, differentiability, integration or differentiation, and we are wondering whether it changes things to perform the limiting process on each f_n individually and then take the limit versus taking the limit first and then perform the limiting process on f. As we can see: in general it does matter! This is not to say that the interchange of limit operations is something to be systematically avoided. On the contrary, it is an essential part of the subject, and in "natural circumstances" the interchange of limit operations is probably valid. But we need to develop theorems to this effect: i.e., under *some specific additional hypotheses*, interchange of limit operations is justified.

Exercises.

EXERCISE 3.1. *Let X be a set, and let $\{f_n : X \to \mathbb{R}\}_{n=0}^\infty$ and $\{g_n : X \to \mathbb{R}\}_{n=0}^\infty$ be two sequences of real-valued functions defined on X. Suppose that $\{f_n\}$ converges pointwise on X to f and $\{g_n\}$ converges pointwise on X to g. Let $\alpha, \beta \in \mathbb{R}$. Show: $\{\alpha f_n + \beta g_n\}_{n=0}^\infty$ converges pointwise on X to $\alpha f + \beta g$.*

EXERCISE 3.2. *Let $X \subseteq \mathbb{R}$, and let $\{f_n : X \to \mathbb{R}\}_{n=0}^\infty$ be a sequence of functions convering pointwise on X to f. Let $\{n_k\}_{k=0}^\infty$ be a strictly increasing sequence in $\mathbb{N}$. Show: the sequence $\{f_{n_k} : X \to \mathbb{R}\}_{k=0}^\infty$ converges pointwise on X to f.*

EXERCISE 3.3. *We say that $f : \mathbb{R}^N \to \mathbb{R}$ is **even** if $f(-x) = f(x)$ for all $x \in \mathbb{R}$ and is **odd** if $f(-x) = -f(x)$ for all $x \in \mathbb{R}$.*

a) *Let $\{f_n : \mathbb{R}^N \to \mathbb{R}\}_{n=0}^\infty$ be a sequence of even functions converging pointwise on $\mathbb{R}^N$ to $f : \mathbb{R} \to \mathbb{R}$. Show: f is even.*
b) *Let $\{f_n : \mathbb{R}^N \to \mathbb{R}\}_{n=0}^\infty$ be a sequence of odd functions converging pointwise on $\mathbb{R}^N$ to $f : \mathbb{R} \to \mathbb{R}$. Show: f is odd.*

EXERCISE 3.4. *Let $X \subseteq \mathbb{R}$, and let $\{f_n : X \to \mathbb{R}\}_{n=0}^\infty$ be a sequence of functions converging pointwise on X to $f : \mathbb{R} \to \mathbb{R}$.*

a) *Suppose that each f_n is increasing: that is, for all $x_1, x_2 \in X$, $x_1 \leq x_2 \implies f_n(x_1) \leq f_n(x_2)$. Show: f is increasing.*
b) *Now suppose only that $\{n \in \mathbb{Z}^{\geq 0} \mid f_n : X \to \mathbb{R}$ is increasing$\}$ is infinite. Show: f is increasing.*
c) *For $n \in \mathbb{N}$, put $f_n(x) := \frac{1}{n+1} \arctan x$. Show: $\{f_n : \mathbb{R} \to \mathbb{R}\}_{n=0}^{\infty}$ is a sequence of strictly increasing functions that converges pointwise on $\mathbb{R}$ to a function that is (increasing but) not strictly increasing.*

Problems.

PROBLEM 3.1. [**Ab**, Example 6.222.(iii)] *Consider the sequence of functions $\{h_n : (-1,1) \to \mathbb{R}\}_{n=1}^{\infty}$ given by*

$$h_n(x) = x^{1+\frac{1}{2n-1}}.$$

a) *Show: each h_n is differentiable.*
b) *Show: h_n converges pointwise on $(-1,1)$ to $h(x) := |x|$.*
c) *On the same axes, sketch a graph of $h_1(x)$, $h_2(x)$, $h_3(x)$ and $h(x)$.*

2. Uniform Convergence

All we have to do now is take these lies and make them true somehow. – G. Michael[3]

Most of the above pathologies vanish if we use a stronger notion of convergence.

2.1. Introducing Uniform Convergence. Let $\{f_n : X \to \mathbb{R}\}_{n=0}^{\infty}$ be a sequence of real-valued functions defined on a set X. We say f_n **converges uniformly on** X to f and write $f_n \stackrel{u}{\to} f$ if for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$ and all $x \in I$, we have $|f_n(x) - f(x)| < \epsilon$.

How does this definition differ from that of pointwise convergence? Let's compare: $f_n \to f$ pointwise if for all $x \in X$ and all $\epsilon > 0$, there exists $n \in \mathbb{N}$ such that for all $n \geq N$, we have $|f_n(x) - f(x)| < \epsilon$. The only difference is in the order of the quantifiers: in pointwise convergence we are first given ϵ and x and then must find an $N \in \mathbb{Z}^+$: that is, the N is allowed to depend both on ϵ and the point $x \in X$. In the definition of uniform convergence, we are given $\epsilon > 0$ and must find an $N \in \mathbb{Z}^+$ which works simultaneously (or "uniformly") for all $x \in I$. Thus uniform convergence is a stronger condition than pointwise convergence, and in particular if f_n converges to f uniformly, then certainly f_n converges to f pointwise.

PROPOSITION 3.1 (Cauchy Criterion For Uniform Convergence). *Let X be a set and let $\{f_n : X \to \mathbb{R}\}_{n=0}^{\infty}$ be a sequence of functions. The following are equivalent:*

(i) *We have $f_n \stackrel{u}{\to} f$.*
(ii) *For all $\epsilon > 0$, there is $N \in \mathbb{N}$ such that for all $m, n \geq N$ and all $x \in X$, we have $|f_m(x) - f_n(x)| < \epsilon$.*

[3]George Michael, 1963–2016

PROOF. (i) $\implies$ (ii): Let $\epsilon > 0$, and choose $N \in \mathbb{Z}^+$ such that for all $n \geq N$ and all $x \in X$ we have $|f_n(x) - f(x)| < \frac{\epsilon}{2}$. For all $m, n \geq N$ and all $x \in X$ we have

$$|f_m(x) - f_n(x)| \leq |f_m(x) - f(x)| + |f_n(x) - f(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

(ii) $\implies$ (i): Let $\epsilon > 0$, and choose $N \in \mathbb{Z}^+$ such that for all $m, n \geq N$ and all $x \in X$ we have

$$|f_m(x) - f_n(x)| < \frac{\epsilon}{2}.$$

Let $x \in X$. Since $f_n(x) \to f(x)$, there is $M_x \geq N$ such that

$$f_{M_x}(x) - f(x)| < \frac{\epsilon}{2}.$$

Then: for all $n \geq N$ we have

$$|f_n(x) - f(x)| \leq |f_n(x) - f_{M_x}(x)| + |f_{M_x}(x) - f(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

so $f_n \stackrel{u}{\to} f$. □

2.2. Uniform Convergence and Inherited Properties.

The following result is the most basic one fitting under the general heading "uniform convergence justifies the exchange of limiting operations."

THEOREM 3.2. *Let A be a subset of $\mathbb{R}^N$, and let c be an accumulation point of A. Let $\{f_n : A \to \mathbb{R}\}_{n=0}^{\infty}$ be a sequence of functions. We suppose:*

(i) *For all $n \in \mathbb{N}$, we have $\lim_{x \to c} f_n(x) = L_n \in \mathbb{R}$, and*

(ii) *We have $f_n \stackrel{u}{\to} f$.*

Then the sequence $\{L_n\}$ is convergent, $\lim_{x \to c} f(x)$ exists and we have equality:

$$\lim_{n \to \infty} L_n = \lim_{n \to \infty} \lim_{x \to c} f_n(x) = \lim_{x \to c} f(x) = \lim_{x \to c} \lim_{n \to \infty} f_n(x).$$

PROOF. Step 1: We show that the sequence $\{L_n\}$ is convergent. Since we don't yet have a real number to show that it converges *to*, it is natural to try to use the Cauchy criterion, hence to try to bound $|L_m - L_n|$. Now comes the trick: for all $x \in A$ we have

$$|L_m - L_n| \leq |L_m - f_m(x)| + |f_m(x) - f_n(x)| + |f_n(x) - L_n|.$$

By the Cauchy criterion for uniform convergence, for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $m, n \geq N$ and all $x \in A$ we have $|f_m(x) - f_n(x)| < \frac{\epsilon}{3}$. Moreover, the fact that $f_m(x) \to L_m$ and $f_n(x) \to L_n$ give us bounds on the first and last terms: there exists $\delta > 0$ such that if $0 < |x - c| < \delta$ then $|L_n - f_n(x)| < \frac{\epsilon}{3}$ and $|L_m - f_m(x)| < \frac{\epsilon}{3}$. Combining these three estimates, we find that by taking $x \in B^\circ(c, \delta) \setminus \{c\}$, we have

$$|L_m - L_n| \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

So the sequence $\{L_n\}$ is Cauchy and hence convergent, say to the real number L.
Step 2: We show that $\lim_{x \to c} f(x) = L$ (so in particular the limit exists!). Actually the argument for this is very similar to that of Step 1:

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

Since $L_n \to L$ and $f_n(x) \to f(x)$, the first and last term will each be less than $\frac{\epsilon}{3}$ for sufficiently large n. Since $f_n(x) \to L_n$, the middle term will be less than $\frac{\epsilon}{3}$ for

x sufficiently close to c. Overall we find that by taking x sufficiently close to (but not equal to) c, we get $|f(x) - L| < \epsilon$ and thus $\lim_{x \to c} f(x) = L$. $\square$

COROLLARY 3.3. *Let $A \subseteq \mathbb{R}^N$ and let $\{f_n : A \to \mathbb{R}\}_{n=0}^{\infty}$ be a sequence of continuous functions such that $f_n \stackrel{u}{\to} f$ on A. Then $f : A \to \mathbb{R}$ is continuous.*

Since Corollary 3.3 is easier than Theorem 3.2, we include a separate proof.

PROOF. Let $x \in A$. We need to show that $\lim_{x \to c} f(x) = f(c)$, thus we need to show that for any $\epsilon > 0$ there exists $\delta > 0$ such that for all $x \in B^\circ(c, \delta)$ we have $|f(x) - f(c)| < \epsilon$. The idea – again! – is to trade this one quantity for three quantities that we have an immediate handle on by writing

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

By uniform convergence, there exists $n \in \mathbb{N}$ such that $|f(x) - f_n(x)| < \frac{\epsilon}{3}$ for all $x \in A$: in particular $|f_n(c) - f(c)| = |f(c) - f_n(c)| < \frac{\epsilon}{3}$. Further, since f_n is continuous, there exists $\delta > 0$ such that for all x with $|x - c| < \delta$ we have $|f_n(x) - f_n(c)| < \frac{\epsilon}{3}$. Consolidating these estimates, we get

$$|f(x) - f(c)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \qquad \square$$

THEOREM 3.4. *Let $\{f_n\}$ be a sequence of Riemann integrable functions with common domain $[a, b]$. Suppose that $f_n \stackrel{u}{\to} f$. Then f is Riemann integrable and*

$$\lim_{n \to \infty} \int_a^b f_n = \int_a^b \lim_{n \to \infty} f_n = \int_a^b f.$$

PROOF. Step 1: We prove the integrability of f. Fix $\epsilon > 0$; since $f \stackrel{u}{\to} f$, there is $N \in \mathbb{Z}^+$ such that for all $n \geq N$ and all $x \in [a, b]$, $|f_n(x) - f(x)| < \epsilon$; it follows that for any subinterval $[c, d] \subset [a, b]$,

$$|\sup(f_n, [c, d]) - \sup(f, [c, d])| \leq \epsilon, |\inf(f_n, [c, d]) - \inf(f, [c, d])| \leq \epsilon.$$

So for any partition $\mathcal{P} = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ of $[a, b]$ and $n \geq N$,

$$\begin{aligned} |U(f_n, \mathcal{P}) - U(f, \mathcal{P})| &\leq \sum_{i=0}^{n-1} |\sup(f_n, [x_n, x_{n+1}]) - \sup(f, [x_n, x_{n+1}]|(x_{i+1} - x_i)| \\ &\leq \sum_{i=0}^{n-1} \epsilon(x_{i+1} - x_i) = (b - a)\epsilon, \end{aligned}$$

and similarly,

$$|L(f_n, \mathcal{P}) - L(f, \mathcal{P})| \leq (b - a)\epsilon.$$

Since f_N is integrable, by Darboux's Criterion there is a partition $\mathcal{P}$ of $[a, b]$ such that $U(f_N, \mathcal{P}) - L(f_N, \mathcal{P}) < \epsilon$. Thus

$$\begin{aligned} |U(f, \mathcal{P}) - L(f, \mathcal{P})| &\leq |U(f, \mathcal{P}) - U(f_n, \mathcal{P})| + |U(f_n, \mathcal{P}) - L(f_n, \mathcal{P})| + |L(f_n, \mathcal{P}) - L(f, \mathcal{P})| \\ &\leq (b - a)\epsilon + \epsilon + (b - a)\epsilon = (2(b - a) + 1)\epsilon. \end{aligned}$$

Since $\epsilon > 0$ was arbitrary, Darboux's Criterion shows f is integrable on $[a, b]$.
Step 2: If $f, g : [a, b] \to \mathbb{R}$ are integrable and $|f(x) - g(x)| \leq \epsilon$ for all $x \in [a, b]$, then

$$|\int_a^b f - \int_a^b g| = |\int_a^b f - g| \leq \int_a^b |f - g| \leq (b - a)\epsilon.$$

From this simple observation and Step 1 the fact that $f_n \xrightarrow{u} f$ implies $\int_a^b f_n \to \int_a^b f$ is almost immediate. The details are left to you. $\square$

COROLLARY 3.5. *Let $\{f_n\}$ be a sequence of continuous functions defined on the interval $[a,b]$ such that $\sum_{n=0}^{\infty} f_n \xrightarrow{u} f$. For each n, let $F_n : [a,b] \to \mathbb{R}$ be the unique function with $F_n' = f_n$ and $F_n(a) = 0$, and similarly let $F : [a,b] \to \mathbb{R}$ be the unique function with $F' = f$ and $F(a) = 0$. Then $\sum_{n=0}^{\infty} F_n \xrightarrow{u} F$.*

You are asked to prove Corollary 3.5 in Exercise 3.8.

Our next order of business is to discuss differentiation of sequences of functions. For this we should reconsider Example 4: let $g : \mathbb{R} \to \mathbb{R}$ be a bounded differentiable function such that $\lim_{n\to\infty} g(n)$ does not exist, and let $f_n(x) = \frac{g(nx)}{n}$. Let M be such that $|g(x)| \leq M$ for all $\mathbb{R}$. Then for all $x \in \mathbb{R}$, $|f_n(x)| \leq \frac{M}{n}$, so $f_n \xrightarrow{u} 0$. But as we saw above, $\lim_{n\to\infty} f_n'(1)$ does not exist.

Thus we have shown the following somewhat distressing fact: *uniform convergence of f_n to f does not imply that f_n' converges.*

Well, don't panic. What we want is true *in practice*; we just need suitable hypotheses. We will give a relatively simple result sufficient for our coming applications.

THEOREM 3.6. *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of functions on $[a,b]$. We suppose:*

(i) *Each f_n is continuously differentiable on $[a,b]$;*
(ii) *The functions f_n converge pointwise on $[a,b]$ to some function f; and*
(iii) *The sequence $\{f_n'\}_{n=1}^{\infty}$ converges uniformly on $[a,b]$ to some function g.*

Then f is differentiable and $f' = g$, or in other words

$$(\lim_{n\to\infty} f_n)' = \lim_{n\to\infty} f_n'.$$

PROOF. Let $x \in [a,b]$. Since $f_n' \xrightarrow{u} g$ on $[a,b]$, certainly $f_n' \xrightarrow{u} g$ on $[a,x]$. Since each f_n' is continuous, by Corollary 3.3 g is continuous. Now applying Theorem 3.4 and the Fundamental Theorem of Calculus we have

$$\int_a^x g = \int_a^x \lim_{n\to\infty} f_n' = \lim_{n\to\infty} \int_a^x f_n' = \lim_{n\to\infty} f_n(x) - f_n(a) = f(x) - f(a).$$

Differentiating and applying the Fundamental Theorem of Calculus, we get

$$g = (f(x) - f(a))' = f'. \qquad \square$$

COROLLARY 3.7. *Let $\sum_{n=0}^{\infty} f_n(x)$ be a series of functions converging pointwise to $f(x)$. Suppose that each f_n' is continuously differentiable and $\sum_{n=0}^{\infty} f_n'(x) \xrightarrow{u} g$. Then f is differentiable and $f' = g$:*

$$(\sum_{n=0}^{\infty} f_n)' = \sum_{n=0}^{\infty} f_n'. \tag{8}$$

You are asked to prove Corollary 3.7 in Exercise 3.9.

When for a series $\sum_n f_n$ it holds that $(\sum_n f_n)' = \sum_n f_n'$, we say that the series can be differentiated **termwise** or **term-by-term**. Thus Corollary 3.7 gives a condition under which a series of functions can be differentiated termwise.

Although Theorem 3.6 (or more precisely, Corollary 3.7) will be sufficient for our needs, we cannot help but record the following stronger version.

THEOREM 3.8. *Let $\{f_n\}$ be differentiable functions on the interval $[a,b]$ such that $\{f_n(x_0)\}$ is convergent for some $x_0 \in [a,b]$. If there is $g : [a,b] \to \mathbb{R}$ such that $f_n' \overset{u}{\to} g$ on $[a,b]$, then there is $f : [a,b] \to \mathbb{R}$ such that $f_n \overset{u}{\to} f$ on $[a,b]$ and $f' = g$.*

PROOF. [**R**, pp.152-153]
Step 1: Fix $\epsilon > 0$, and choose $N \in \mathbb{Z}^+$ such that $m, n \geq N$ implies $|f_m(x_0) - f_n(x_0)| \frac{\epsilon}{2}$ and $|f_m'(t) - f_n'(t)| < \frac{\epsilon}{2(b-a)}$ for all $t \in [a,b]$. The latter inequality is telling us that the derivative of $g := f_m - f_n$ is small on the entire interval $[a,b]$. Applying the Mean Value Theorem to g, we get a $c \in (a,b)$ such that for all $x, t \in [a,b]$ and all $m, n \geq N$,

$$|g(x) - g(t)| = |x - t||g'(c)| \leq |x - t| \left(\frac{\epsilon}{2(b-a)} \right) \leq \frac{\epsilon}{2}. \tag{9}$$

It follows that for all $x \in [a,b]$,

$$|f_m(x) - f_n(x)| = |g(x)| \leq |g(x) - g(x_0)| + |g(x_0)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

By the Cauchy criterion, f_n is uniformly convergent on $[a,b]$ to some function f.
Step 2: Now **fix** $x \in [a,b]$ and define

$$\varphi_n(t) = \frac{f_n(t) - f_n(x)}{t - x}$$

and

$$\varphi(t) = \frac{f(t) - f(x)}{t - x},$$

so that for all $n \in \mathbb{Z}^+$, $\lim_{x \to t} \varphi_n(t) = f_n'(x)$. Now by (9) we have

$$|\varphi_m(t) - \varphi_n(t)| \leq \frac{\epsilon}{2(b-a)}$$

for all $m, n \geq N$, so once again by the Cauchy criterion φ_n converges uniformly for all $t \neq x$. Since $f_n \overset{u}{\to} f$, we get $\varphi_n \overset{u}{\to} \varphi$ for all $t \neq x$. Finally we apply Theorem 3.2 on the interchange of limit operations:

$$f'(x) = \lim_{t \to x} \varphi(t) = \lim_{t \to x} \lim_{n \to \infty} \varphi_n(t) = \lim_{n \to \infty} \lim_{t \to x} \varphi_n(t) = \lim_{n \to \infty} f_n'(x). \qquad \square$$

2.3. The Weierstrass M-test.

We have just seen that uniform convergence of a sequence of functions (and possibly, of its derivatives) has many pleasant consequences. The next order of business is to give a useful general criterion for a sequence of functions to be uniformly convergent.

Let X be a nonempty set. For a function $f : X \to \mathbb{R}$, we define

$$||f|| = \sup_{x \in X} |f(x)|.$$

In more words, $||f||$ is the least $M \in [0, \infty]$ such that $|f(x)| \leq M$ for all $x \in X$.

THEOREM 3.9 (Weierstrass M-Test). *Let X be a nonempty set, and let $\{f_n : X \to \mathbb{R}\}_{n=0}^{\infty}$ be a sequence of real-valued functions defined on X. Let $\{M_n\}_{n=0}^{\infty}$ be a non-negative sequence such that:*

(i) *We have $||f_n|| \leq M_n$ for all $n \in \mathbb{N}$, and*

(ii) *We have* $\sum_{n=0}^{\infty} M_n < \infty$.

Then the series $\sum_{n=0}^{\infty} f_n$ *converges uniformly on* X.

PROOF. Let $S_N(x) := \sum_{n=0}^{N} f_n(x)$. Since $\sum_n M_n < \infty$, for each $\epsilon > 0$ there is $N_0 \in \mathbb{Z}^+$ such that for all $N \geq N_0$, we have $\sum_{n>N} M_n < \epsilon$. Then for $x \in X$, $N \geq N_0$ and $k \in \mathbb{N}$, we have

$$|S_{N+k}(x) - S_N(x)| = \left| \sum_{n=N+1}^{N+k} f_n(x) \right| \leq \sum_{n>N} |f_n(x)| \leq \sum_{n>N} M_n < \epsilon.$$

Therefore the series is uniformly convergent by the Cauchy criterion. □

Exercises.

EXERCISE 3.5. *Consider again* $f_n(x) = x^n$ *on the interval* $[0,1]$. *We saw in Example 3.1 that* f_n *converges pointwise to the discontinuous function* f *which is* 0 *on* $[0,1)$ *and* 1 *at* $x = 1$.

a) *Show directly from the definition that the convergence of* f_n *to* f is not *uniform.*
b) *Try to pinpoint exactly where the proof of Theorem 3.2 breaks down when applied to this non-uniformly convergent sequence.*

EXERCISE 3.6. *Let* X *be a set, and let* Y *be a subset of* X *such that* $X \setminus Y$ *is finite. Let* $\{f_n : X \to \mathbb{R}\}_{n=0}^{\infty}$ *be a sequence of functions defined on* X. *Let* $f : X \to \mathbb{R}$ *be a function. Show that the following are equivalent:*

(i) *The sequence* $\{f_n\}$ *converges uniformly on* Y *to* f, *and for all* $x \in X \setminus Y$ *we have* $f_n(x) \to f(x)$.
(ii) *The sequence* $\{f_n\}$ *converges uniformly on* X *to* f.

EXERCISE 3.7. *It follows from Theorem 3.4 that the sequences in Examples 2 and 3 above are not uniformly convergent. Verify this directly.*

EXERCISE 3.8. *Prove Corollary 3.5.*

EXERCISE 3.9. *Prove Corollary 3.7.*

Problems.

PROBLEM 3.2. *Let* X *be a set. Show that the following are equivalent:*

(i) *There is a sequence* $\{f_n : X \to \mathbb{R}\}_{n=0}^{\infty}$ *that converges pointwise on* X *but not uniformly on* X.
(ii) *The set* X *is infinite.*

3. Power Series

3.1. Convergence of power series. Recall that to a real sequence $\{a_n\}_{n=0}^{\infty}$ we can associate the **power series**

$$\sum_{n=0}^{\infty} a_n x^n.$$

We would like to view a power series as a function of x, and in fact one of our main applications of uniform convergence will be to show that it is a very pleasant function of x. But in order to have a function we first have to know what its domain is. This brings us to the following definition: the **domain** of the power series

$\sum_{n=0}^{\infty} a_n x^n$ is the set of $x \in \mathbb{R}$ for which the series converges.

So the natural question is: what is the domain of a power series?

There are two ways to construe this question. The first is: given a *particular* power series — e.g. $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ — how do we determine the domain? This question is addressed in calculus: if a certain *ratio test limit* exists, then that can be used to determine the domain, at least up to determining whether two endpoints are in the domain, for which one can (try to) use other convergence tests. But there is another way to construe this question: what can we say about the domain of convergence of a power series *in general*: e.g. which subsets of $\mathbb{R}$ arise as the domain of convergence of power series? This latter question is not addressed in calculus. It used to be addressed in Math 3100, but with the recent change in the syllabus I'm not sure whether it still is. Anyway, we will give a more vigorous answer than it is standard to give in Math 3100.

I claim that to any power series $\sum_{n=0}^{\infty} a_n x^n$ we can associate an extended real number $R \in [0, \infty]$ called the **radius of convergence** that has the following properties:

- If $R = 0$, then the domain is $D = \{0\}$.
- If $R = \infty$, then the domain is $D = \mathbb{R}$.
- If $R \in (0, \infty)$ then the domain D satisfies $(-R, R) \subseteq D \subseteq [-R, R]$.

In the last case we are very nearly saying what the domain is: the only ambiguity is that $-R$ might or might not be in the domain and also R might or might not be in the domain. In Problem 3.10 you will be asked to give, for every $R > 0$, four different power series with domains $(-R, R)$, $[-R, R)$, $(-R, R]$ and $[-R, R]$.

In particular if this is true then the domain of covnergence of any power series is an interval on the real line.

We will actually give a formula for R in the general case, but before we do that let's revisit the Ratio and Root Tests from a somewhat more sophisticated perspective than we had in fresman calculus.

THEOREM 3.10 (Ratio Test). *Let $\sum_n a_n$ be a real series with $a_n \neq 0$ for all n.*[4]

a) *Suppose there is $N \in \mathbb{Z}^+$ and $r < 1$ such that $|\frac{a_{n+1}}{a_n}| \leq r$ for all $n \geq N$. Then the series is absolutely convergent.*
b) *Suppose there is $N \in \mathbb{Z}^+$ and $R \geq 1$ such that $|\frac{a_{n+1}}{a_n}| \geq R$ for all $n \geq N$. Then $|a_n| \to \infty$, so the series diverges.*
c) *Consider the quantity*

$$\overline{\rho} := \overline{\lim} \left| \frac{a_{n+1}}{a_n} \right| \in [0, \infty].$$

If $\overline{\rho} < 1$, then the hypothesis of part a) holds, so the series is absolutely convergent.

[4]We write $\sum_n a_n$ because we are only concerned with the convergence or divergence of the series, so it doesn't matter whether it starts at 0 or 1...or anywhere else.

d) *Consider the quantity*

$$\underline{\rho} := \underline{\lim} \left| \frac{a_{n+1}}{a_n} \right| \in [0, \infty].$$

If $\underline{\rho} > 1$, then the hypothesis of part b) holds, so the series is divergent.

PROOF. Since we are only concerned about convergence, we may remove finitely terms of a series and thereby pass from a condition holding for all sufficiently large terms to a condition that holds for all terms. We will do so in parts a) and b).
a) We have $|a_0| \leq r|a_1|$, $|a_2| \leq r|a_1| \leq r^2|a_0|$, $a_3 \leq r|a_2| \leq r^3|a_0|$, and so forth. An immedaite inductive argument establishes:

$$\forall n \in \mathbb{N}, \ |a_n| \leq r^n |a_0|,$$

so

$$\sum_{n=0}^{\infty} |a_n| \leq \sum_{n=0}^{\infty} r^n |a_0| = \frac{|a_0|}{1-r}$$

and our series is absolutely convergent by comparsion to a geometric series.
b) Our assumption implies that $|\frac{a_{n+1}}{a_n}| \geq 1$ for all n, which means that $|a_{n+1}| \geq |a_n|$ for all n, which means

$$|a_0| \leq |a_1| \leq |a_2| \leq \ldots \leq |a_n| \leq .$$

Therefore every term has absolute value at least $|a_0| \neq 0$, so a_n does not converge to 0. This implies that the series diverges (e.g. [**SS**, Proposition 3.1.9] or [**HC**, Theorem 11.4].
c) If $\overline{\rho} < 1$, choose r with $\overline{\rho} < r < 1$. By our "creeping" interpretation of the limsup — Proposition 1.14 — there is $N \in \mathbb{Z}^+$ such that $|\frac{a_{n+1}}{a_n}| \leq r$ for all $n \geq N$, so the hypothesis of part a) holds.
d) We use the analogous "creeping" interpretation of the liminf (see Exercise 1.14): if $\underline{\rho} > 1$, choose R with $1 < R < \underline{\rho}$. Then there is $N \in \mathbb{Z}^+$ such that $|\frac{a_{n+1}}{a_n}| \geq R$ for all $n \geq N$, so the hypothesis of part b) holds. □

In freshman calculus, one usually assumes that the Ratio Test limit

$$\rho = \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right|$$

exists, which is equivalent to the equality $\underline{\rho} = \overline{\rho}$. So the calculus version of the Ratio Test follows from Theorem 3.10: if ρ exists and is less than 1, then $\overline{\rho} = \rho < 1$, so we have absolute convergence by part a), while if ρ exists and is greater than 1, then $\underline{\rho} = \rho > 1$, so we have divergence by part b). If $\rho = 1$, then parts c) and d) clearly don't apply, and in fact part a) does not either, since the hypothesis implies $\overline{\rho} \leq r < 1$. However part b) may still apply: e.g. it applies to a series with $a_n = C \neq 0$ for all n.

Another way for Theorem 3.10 to fail is if $\overline{\rho} > 1$ and $\underline{\rho} < 1$. In Exercise 3.20 you are asked to show that this holds for the series

$$\sum_{n=0}^{\infty} 2^{-n+(-1)^n}.$$

In this example the presence of $(-1)^n$ is enough to trip up the Ratio Test because it messes with the relatve sizes of the terms. Nevertheless, whether the nth term is

2^{-n-1} or 2^{-n+1} it is still exponentially small, so we certainly expect the series to converge and indeed this can be shown by observing that

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

and comparing to a convergent geometric series. But there is a close relative of the Ratio Test that is not tripped up by this example.

THEOREM 3.11 (Root Test). *Let $\sum_n a_n$ be a real series.*

a) *Suppose there is $N \in \mathbb{Z}^+$ and $r < 1$ such that $|a_n|^{\frac{1}{n}} \leq r$ for all $n \geq N$. Then the series is absolutely convergent.*
b) *Suppose we have $|a_n|^{\frac{1}{n}} \geq 1$ for infinitely many n. Then $a_n \nrightarrow 0$, so the series diverges.*
c) *Consider the quantity*

$$\overline{\theta} := \overline{\lim}\ |a_n|^{\frac{1}{n}}.$$

Then:
(i) *If $\overline{\theta} < 1$, then the hypothesis of part a) holds, so the series is absolutely convergent.*
(ii) *If $\overline{\theta} > 1$, then the hypothesis of part b) holds, so the series is divergent.*

We leave the proof of Theorem 3.11 as Problem 3.11, but we will make a few comments. The proof of part a) has the same strategy as that of the corresponding part of the Ratio Test: we compare to a geometric series, while the proof of part b) is even easier than that of the corrsponding part of the Ratio Test. The proof of c) part (i) is also similar to the proof of part c) of the Ratio Test. But in c) part (ii) it is initially surprising that the limsup appears again, whereas in the Ratio Test the liminf appears in the corresponding part b). To sound this out a bit: knowing that $|a_n|^{\frac{1}{n}} \geq 1$ on any infinite set of n is indeed enough to know that the series diverges, whereas knowing that $|\frac{a_{n+1}}{a_n}| \geq R > 1$ on an infinite set of n is not telling us enough to conclude anything about convergence: it is good to think about why.

In Problem 3.12 you are asked to show that for any series $\sum_n a_n$ with nonzero terms, the upper and lower ratio and root test limits are related as follows:

$$\underline{\rho} \leq \underline{\theta} \leq \overline{\theta} \leq \overline{\rho}. \tag{10}$$

From (10) it follows that if the Ratio Test limit ρ exists, then also the Root Test limit θ exists and they are equal. However we have already seen an example where θ exists but ρ does not. If you have ever heard someone say "The Root Test is stronger than the Ratio Test," then that is what they meant (and they were right!).

THEOREM 3.12 (Cauchy-Hadamard Formula). *Let $\{a_n\}_{n=0}^{\infty}$ be any real sequence. Put*

$$\overline{\theta} := \overline{\lim}\ |a_n|^{\frac{1}{n}} \in [0, \infty]$$

and

$$R := \frac{1}{\overline{\theta}} \in [0, \infty].$$

Then the power series $\sum_{n=0}^{\infty} a_n x^n$ has radius of convergence R. More precisely, the power series converges absolutely for $|x| < R$ and diverges for $|x| > R$.

PROOF. Our version of the Root Test must be pretty good, because we can deduce this important result almost immediately.

First let $x \in \mathbb{R}$ be such that $|x| < R$. (Thus $R > 0$; we allow $R = +\infty$.) Since $\overline{\lim}\ |a_n|^{\frac{1}{n}} = \overline{\theta}$, we have

$$\overline{\lim}\ |x^n a_n|^{\frac{1}{n}} = \overline{\lim}\ |x||a_n|^{\frac{1}{n}} = |x| \cdot \overline{\lim}\ |a_n|^{\frac{1}{n}} = |x|\overline{\theta} = \frac{|x|}{R} < 1.$$

So the series $\sum_n a_n x^n$ is absolutely convergent by part c)(i) of the Root Test.
Now let $x \in \mathbb{R}$ be such that $|x| > R$. (Thus $R \neq +\infty$; we allow $R = 0$.) This time

$$\overline{\lim}\ |x^n a_n|^{\frac{1}{n}} = \frac{|x|}{R} > 1,$$

so the series diverges by part c)(ii) of the Root Test. □

As discussed at the beginning of this section, it follows immediately from Theorem 3.12 that the domain of convergence of any real power series is an interval: more precisely takes one of the forms $\{0\}$, $\mathbb{R}$, $(-R, R)$, $[-R, R)$, $(-R, R]$ or $[-R, R]$ where R is as in Theorem 3.12.

Let me end with a few words about the history of this beautiful result. Jacques Hadamard was a French mathematician who lived from 1865 to 1963. Among his many achievements was giving one of the first two proofs of the celebrated Prime Number Theorem, which is that if $\pi(x) : [1, \infty) \to \mathbb{R}$ is defined as the number of prime numbers $p \leq x$, then

$$\lim_{x \to \infty} \frac{\pi(x)}{\frac{x}{\log x}} = 1.$$

He did this in 1896, at the same time as another mathematician, Charles de la Vallée Poussin. Both mathematicians were following a strategy layed out by Bernhard Riemann (1826-1866), but this strategy was not easy to implement.

Hadamard published a proof of Theorem 3.12 in 1888 [**Ha88**] while he was still a student, and he included the proof in his 1892 PhD thesis. As we saw, this result is a quick consequence of the Root Test. The Root Test is due to one of the true masters, Augustin-Louis Cauchy (1789-1857); in some circles it is called "Cauchy's Root Test" whereas the Ratio Test is attributed to Jean le Rond d'Alembert (1717-1783). How could The Master know the Root Test and not its application to the radius of convergence of a power series?!? The answer is that *he couldn't **not** know it*: this formula appears in an 1821 textbook of Cauchy [**Ca21**]. Cauchy is a household name unto this very day and course, but most of us don't read his original works because they are written in French in the style of 19th century mathematics. But Hadamard and his teachers were...19th century French mathematicians. Cauchy would have been their patron saint. How *in the world* was Cauchy's contribution forgotten in Hadamard's time? The answer is that many good mathematicians are bad academics: almost no research mathematicians spend any significant time reading primary source material. So this kind of independent rediscovery – even decades or centuries later – is in fact rather common.

3.2. Power series as functions.

THEOREM 3.13 (Wonderful Properties of Power Series). *Let $\sum_{n=0}^{\infty} a_n x^n$ be a power series with radius of convergence $R > 0$. Consider $f(x) = \sum_{n=0}^{\infty} a_n x^n$ as a function $f : (-R, R) \to \mathbb{R}$. Then:*

a) *The function f is continuous.*
b) *The function f is differentiable. Morever, its derivative may be computed* termwise*:*
$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}.$$
c) *Since the power series f' has the same radius of convergence $R > 0$ as f, the function f is in fact infinitely differentiable.*
d) *For all $n \in \mathbb{N}$, we have $f^{(n)}(0) = (n!)a_n$.*

Proof.

a) Let $0 < A < R$, so f defines a function from $[-A, A]$ to $\mathbb{R}$. We claim that the series $\sum_n a_n x^n$ converges to f uniformly on $[-A, A]$. Indeed, as a function on $[-A, A]$, we have $||a_n x^n|| = |a_n| A^n$, and thus
$$\sum_n ||a_n x^n|| = \sum_n |a_n| A^n < \infty,$$
because power series converge absolutely on the interior of their interval of convergence. By the Weierstrass M-test, f is the uniform limit of the sequence $S_n(x) = \sum_{k=0}^{n} a_k x^k$. But each S_n is a polynomial function, hence continuous and infinitely differentiable. So by Theorem 3.2, the function f is continuous on $[-A, A]$. Since any $x \in (-R, R)$ lies in $[-A, A]$ for some $0 < A < R$, also the function f is continuous on $(-R, R)$.
b) According to Corollary 3.7, in order to show that $f = \sum_n a_n x^n = \sum_n f_n$ is differentiable and the derivative may be computed termwise, it is enough to check that (i) each f_n is continuously differentiable and (ii) $\sum_n f_n'$ is uniformly convergent. But (i) is trivial, since $f_n = a_n x^n$ — of course monomial functions are continuously differentiable. As for (ii), we compute that
$$\sum_n f_n' = \sum_n (a_n x^n) = \sum_n n a_{n-1} x^{n-1}.$$
Since $\lim_{n\to\infty} |n|^{\frac{1}{n}} = 1$, by Cauchy-Hadamard the power series $\sum_n n a_{n-1} x^{n-1}$ also has radius of convergence R, hence by the result of part a) it is uniformly convergent on $[-A, A]$. Therefore Corollary 3.7 applies to show $f'(x) = \sum_{n=0}^{\infty} n a_n x^{n-1}$.
c) We have just seen that for a power series f convergent on $(-R, R)$, its derivative f' is also given by a power series convergent on $(-R, R)$. So we may continue in this way: by induction, derivatives of all orders exist.
d) The formula $f^{(n)}(0) = (n!)a_n$ is simply what one obtains by repeated termwise differentiation. We leave this as an exercise to the reader. □

The fact that for any power series $f(x) = \sum_n a_n x^n$ with positive radius of convergence we have $a_n = \frac{f^{(n)}(0)}{n!}$ yields the following important result.

Corollary 3.14. *(Uniqueness Theorem) Let $f(x) = \sum_n a_n x^n$ and $g(x) = \sum_n b_n x^n$ be two power series with radii of convergence R_a and R_b with $0 < R_a \leq R_b$, so that both f and g are infinitely differentiable functions on $(-R_a, R_a)$. Suppose that for some δ with $0 < \delta \leq R_a$ we have $f(x) = g(x)$ for all $x \in (-\delta, \delta)$. Then $a_n = b_n$ for all n.*

The upshot of Corollary 3.14 is that the only way that two power series can be equal as functions – even in some very small interval around zero – is if all of their

coefficients are equal. This is not obvious, since in general $\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} b_n$ does not imply $a_n = b_n$ for all n. Another way of saying this is that the only power series a function can be equal to on a small interval around zero is its Taylor series.

Exercises.

EXERCISE 3.10. *Prove Theorem 3.13d).*

EXERCISE 3.11. *Show that if $f(x) = \sum_{n=0}^{\infty} a_n x^n$ has radius of convergence $R > 0$, then $F(x) = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}$ is an anti-derivative of f.*

EXERCISE 3.12. *Suppose $f(x) = \sum_n a_n x^n$ and $g(x) = \sum_n b_n x^n$ are two power series each converging on some open interval $(-A, A)$. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence of elements of $(-A, A) \setminus \{0\}$ such that $\lim_{n\to\infty} x_n = 0$. Suppose that $f(x_n) = g(x_n)$ for all $n \in \mathbb{Z}^+$. Show that $a_n = b_n$ for all n.*

Problems.

PROBLEM 3.3. *Let $\sum_n a_n x^n$ be a power series with infinite radius of convergence, hence defining a function $f : \mathbb{R} \to \mathbb{R}$. Show that the following are equivalent:*

(i) *The series $\sum_n a_n x^n$ is uniformly convergent on $\mathbb{R}$.*
(ii) *We have $a_n = 0$ for all sufficiently large n.*

PROBLEM 3.4. *Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ be a power series with $a_n \geq 0$ for all n. Suppose that the radius of convergence is 1, so that f defines a function on $(-1, 1)$. Show that the following are equivalent:*

(i) *The series $\sum_n a_n$ converges.*
(ii) *The power series converges uniformly on $[0, 1]$.*
(iii) *The function f is bounded on $[0, 1)$.*

4. The Weierstrass Approximation Theorem

4.1. Statement of Weierstrass Approximation.

THEOREM 3.15 (Weierstrass Approximation Theorem). *Let $f : [a, b] \to \mathbb{R}$ be a continuous function and ϵ any positive number. Then there exists a polynomial function P such that for all $x \in [a, b]$, $|f(x) - P(x)| < \epsilon$. In other words, any continuous function defined on a closed, bounded interval is the uniform limit of a sequence of polynomials.*

It is interesting to compare Theorem 3.15 with Taylor's theorem, which gives conditions for a function to be equal to its Taylor series. Note that any such function must be C^∞ (i.e., it must have derivatives of all orders), whereas in the Weierstrass Approximation Theorem we can get *any* continuous function. An important difference is that the Taylor polynomials $T_N(x)$ have the property that $T_{N+1}(x) = T_N(x) + a_{N+1}x^N$, so that in passing from one Taylor polynomial to the next, we are not changing any of the coefficients from 0 to N but only adding a higher order term. In contrast, for the sequence of polynomials $P_n(x)$ uniformly converging to f in Theorem 1, $P_{n+1}(x)$ is not required to have any simple algebraic relationship to $P_n(x)$.

Theorem 3.15 was first established by Weierstrass in 1885. To this day it is one of the most central and celebrated results of mathematical analysis. Many mathematicians have contributed novel proofs and generalizations. We will give a simply

remarkable 1912 proof of S.J. Bernstein [**Be12**], which is strongly motivated by probability theory but can be understood without it.

4.2. The proof. Knowing the Weierstrass Approximation Theorem on any one closed bounded interval $[a, b]$ with $a < b$ easily implies it on all closed bounded intervals. We leave this as Exercise 3.6. Bernstein's proof concerns the case $[a, b] = [0, 1]$.

Recall that for integers n and k with $0 \leq k \leq n$ we have the binomial coefficient

$$\binom{n}{k} := \frac{n!}{k!(n-k)!} \in \mathbb{Z}^+.$$

These appear in the **binomial theorem**:

$$\forall x, y \in \mathbb{R}, \forall n \in \mathbb{N},\ (x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

For integers $0 \leq k \leq n$ we define the **Bernstein polynomial**

$$B_{n,k}(x) := \binom{n}{k} x^k (1-x)^{n-k}.$$

Then $B_{n,k}(x)$ is a polynomial of degree n such that $B_{n,k}(x) \geq 0$ for all $x \in [0, 1]$.

Example 3.5.

a) *For $n = 0$ we have $B_{0,0}(x) = 1$.*
b) *For $n = 1$ we have $B_{1,0} = 1 - x$, $B_{1,1} = x$.*
c) *For $n = 2$ we have $B_{2,0} = (1-x)^2$, $B_{2,1} = 2x(1-x)$, $B_{2,2} = x^2$.*

Lemma 3.16. *Let $n \in \mathbb{N}$. Then:*

a) *We have $\sum_{k=0}^{n} B_{n,k}(x) = 1$.*
b) *We have $\sum_{k=0}^{n} kB_{n,k}(x) = nx$.*
c) *We have $\sum_{k=0}^{n} k(k-1)B_{n,k}(x) = n(n-1)x^2$.*

Proof. a) Taking $y = 1 - x$ in the binomial theorem yields

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

b) For $n = 0$ both sides of the desired identity are 0, so we may assume that $n \geq 1$. Then we have

$$\sum_{k=0}^{n} kB_{n,k}(x) = \sum_{k=1}^{n} kB_{n,k}(x) = \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k}$$

$$= n \sum_{k=1}^{n} x \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} x^{k-1} (1-x)^{(n-1)-(k-1)} = nx \sum_{k=0}^{n-1} B_{n-1,k}(x) = nx,$$

where in the last equality we used part a).
c) When $n = 0$ or $n = 1$, both sides of the desired identity are 0, so we may assume that $n \geq 2$. Then we have

$$\sum_{k=0}^{n} k(k-1)B_{n,k}(x) = \sum_{k=0}^{n} n(n-1)x^2 \frac{(n-2)!}{(k-2)!((n-2)-(k-2))!} x^{k-2} (1-x)^{(n-2)-(k-2)}$$

$$= n(n-1)x^2 \sum_{k=0}^{n-2} B_{n-2,k}(x) = n(n-1)x^2. \qquad \square$$

LEMMA 3.17. *Let $n \in \mathbb{Z}^+$. Then:*

a) *We have $\sum_{k=0}^{n} \frac{k}{n} B_{n,k}(x) = x$.*

b) *We have $\sum_{k=0}^{n} \left(\frac{k}{n} - x\right)^2 B_{n,k}(x) = \frac{x(1-x)}{n}$.*

PROOF. Parts a) and b) follow from parts b) and c) of Lemma 3.16 respectively by simple algebraic manipulations. We leave the details as Exercise 3.13. $\square$

I will admit that my probability theory is rusty, but neverthless I will mention the probabilistic interpretation of Lemma 3.17. For $n \in \mathbb{N}$, $0 \leq k \leq n$ and $x \in [0,1]$ we have that $B_{n,k}(x) \geq 0$ and $\sum_{k=0}^{n} B_{n,k}(x) = 1$. Thus we may interpret $B_{n,k}(x)$ as giving a probability distribution on the finite space $\mathbf{P} := \{0, 1, \ldots, n\}$: the probability of the basic event k is $B_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}$. This is a famous probability distribution, called **Bernoulli trials**: if we have a coin for which each time it is flipped, the probability of getting heads is $x \in [0,1]$, then the probability that we have exactly k heads after n flips is $B_{n,k}(x)$.

Now suppose $n \geq 1$. We may view $k \mapsto \frac{k}{n}$ as a random variable on $\mathbf{P}$, i.e., as a function $X : \mathbf{P} \to \mathbb{R}$. Lemma 3.17a) is then saying that the expected value $E(X)$ of X is x, while Lemma 3.17b) is saying the variance of X — that is, the expected value of $(X - E(X))^2$ — is $\frac{x(1-x)}{n}$, which converges to 0 uniformly for $x \in [0,1]$. Note also that $\max_{x \in [0,1]} x(1-x) = \frac{1}{4}$.

Now we are ready to state and prove:

THEOREM 3.18 (Bernstein's Theorem).
Let $f : [0,1] \to \mathbb{R}$ be continuous, and for $n \in \mathbb{Z}^+$ define

$$B_n(f)(x) := \sum_{k=0}^{n} f\left(\frac{k}{n}\right) B_{n,k}(x).$$

Then the sequence $B_n(f)$ converges uniformly to f on $[0,1]$.

Notice that $B_n(f)(x)$ is a polynomial of degree at most n. Therefore Theorem 3.18 is a stronger version of the Weierstrass Approximation Theorem: not only does ensure that every continuous function $f : [0,1] \to \mathbb{R}$ is a uniform limit of polynomials, it gives one such sequence of polynomials explicitly, namely $B_n(f)$.

PROOF. By Lemma 3.16a) we have

$$|B_n(f)(x) - f(x)| = |\sum_{k=0}^{n} \left(f\left(\frac{k}{n}\right) - f(x)\right) |B_{n,k}(x).$$

Let $\epsilon > 0$. By Theorem 1.31 the function f is uniformly continuous, so there is $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \frac{\epsilon}{2}$. Let $x \in [0,1]$. We have

$$|B_n(f)(x) - f(x)| \leq \sum_{|\frac{k}{n} - x| < \delta} \left| f\left(\frac{k}{n}\right) - f(x) \right| B_{n,k}(x) + \sum_{|\frac{k}{n} - x| \geq \delta} \left| f\left(\frac{k}{n}\right) - f(x) \right| B_{n,k}(x)$$

$$\leq \frac{\epsilon}{2} \sum_{|\frac{k}{n} - x| < \delta} B_{n,k}(x) + 2||f|| \sum_{|\frac{k}{n} - x| \geq \delta} B_{n,k}(x).$$

To deal with the second term, we use Lemma 3.17b) and write

$$\sum_{|\frac{k}{n}-x|\geq\delta} B_{n,k}(x) = \sum_{|\frac{k}{n}-x|\geq\delta} \frac{\left|\frac{k}{n}-x\right|^2}{\left|\frac{k}{n}-x\right|^2} B_{n,k}(x) \leq \frac{1}{\delta^2} \sum_{|\frac{k}{n}-x|\geq\delta} \left|\frac{k}{n}-x\right|^2 B_{n,k}(x) \leq \frac{x(1-x)}{n\delta^2}.$$

So altogether we have – using $x(1-x) \leq \frac{1}{4}$ – that

$$|B_n(f) - f(x)| \leq \frac{\epsilon}{2} + 2||f|| \frac{x(1-x)}{n\delta^2} \leq \frac{\epsilon}{2} + \frac{||f||}{2n\delta^2}.$$

This last quantity is independent of x and is, for sufficiently large n, at most ϵ, completing the proof. □

4.3. Generalizations. There have been many generalizations of the Weierstrass Approximation Theorem. First, the result holds verbatim with $[a, b]$ replaced by any closed and bounded subset A of $\mathbb{R}$: this is Problem 3.7. The next step is to consider a bounded subset $A \subset \mathbb{R}^N$. We still have the notion of a polynomial function $P : \mathbb{R}^N \to \mathbb{R}$: see §1.4.3. Then:

THEOREM 3.19 (Weierstrass Approximation Theorem in $\mathbb{R}^N$).
Let $A \subset \mathbb{R}^N$ be closed and bounded. Then every continuous function $f : A \to \mathbb{R}$ is a uniform limit of polynomials.

If $A \subset \mathbb{R}$ is closed and bounded, then in fact every continuous function $f : A \to \mathbb{R}$ has a continuous extension to all of $\mathbb{R}^N$: as mentioned in §1.6, this is a special case of the Tietze Extension Theorem [**GT**, Theorem 2.89]. In particular f extends to a closed box $B = \prod_{i=1}^N [a_i, b_i]$ containing A. If we can write the extended function $F : B \to \mathbb{R}$ as a uniform limit of polynomials, then just by restricting to A we get f as a uniform limit of polynomials, so we have reduced Theorem 3.19 to the case of a closed box. In this case it is not so hard to push Bernstein's proof through: again we can get explicit polynomials, which are linear combinations of products of univariate Bernstein polynomials. This is done for instance in [**HS33**]. However, it is not really worth going into the details because much stronger results are known, as we will now explain.

In fact it is fruitful to take a more general perspective. For a subset $A \subseteq \mathbb{R}^N$, the set $\mathcal{C}(A)$ of continuous real-valued functions $f : A \to \mathbb{R}$ has, as we know, a natural algebraic structure: first of all for every $\alpha \in \mathbb{R}$, we have the constant function $\alpha \in \mathcal{C}(A)$. Moreover, if $f, g \in \mathcal{C}(A)$ then $f + g$ and $f \cdot g$ also lie in $\mathcal{C}(A)$. We summarize these properties by saying that $\mathcal{C}(A)$ forms an **$\mathbb{R}$-algebra**. Exactly the same holds for the set $\mathcal{P}(A)$ of polynomial functions on A: it contains the constant functions and its closed under addition and multiplication. A restatement of Theorem 3.19 is that, when A is closed and bounded, every element of the larger $\mathbb{R}$-algebra $\mathcal{C}(A)$ is a unfiorm limit of functions lying in the smaller $\mathbb{R}$-algebra $\mathcal{P}(A)$. If S is a subset of $\mathcal{C}(A)$ such that every element is a uniform limit of a sequence in S, we will say that S is **dense** in $\mathcal{C}(A)$, so once again we can restate Theorem 3.19 by saying that $\mathcal{P}(A)$ is a dense $\mathbb{R}$-subalgebra of $\mathcal{C}(A)$. This raises a natural question: let $\mathcal{A}$ be any $\mathbb{R}$-subalgebra of $\mathcal{C}(A)$. Must $\mathcal{A}$ be dense in $\mathcal{C}(A)$? If not, can we find conditions to ensure the density?

The answer to the first question is *no*. For a simple example, let

$$\mathcal{A} := \{f : [0, 1] \to \mathbb{R} \mid f \text{ is continuous and } f(0) = f(1)\}$$

be the set of all continuous real-valued functions on $[0,1]$ taking the same value at 0 and 1. It is immediate that this is an $\mathbb{R}$-subalgebra of $\mathcal{A}$. However, let $\{f_n\}_{n=1}^{\infty}$ be a sequence in $\mathcal{A}$ and suppose that $f_n \stackrel{u}{\to} f$. Then

$$f(1) - f(0) = \lim_{n\to\infty} f_n(1) - \lim_{n\to\infty} f_n(0) = \lim_{n\to\infty} (f_n(1) - f_n(0)) = \lim_{n\to\infty} 0 = 0,$$

so $f(0) = f(1)$. (In fact we only used the pointwise convergence to deduce this.) So any uniform limit of a sequence in $\mathcal{A}$ satisfies $f(0) = f(1)$ and thus e.g. the function $f(x) = x$ is not such a uniform limit: $\mathcal{A}$ is not dense in $\mathcal{C}(A)$.

This example generalizes as follows:

PROPOSITION 3.20. *Let $A \subseteq \mathbb{R}^N$, and let $\mathcal{A}$ be a dense $\mathbb{R}$-subalgebra of $\mathcal{C}(A)$. Then $\mathcal{A}$* ***separates points*** *of A: that is, for all $\mathbf{x} \neq \mathbf{y} \in A$ there is $f \in \mathcal{A}$ such that $f(\mathbf{x}) \neq f(\mathbf{y})$.*

You are asked to prove Proposition 3.20 in Exercise 3.14.

Notice that for *any* subset $A \subseteq \mathbb{R}^N$, the subalgebra $\mathcal{P}(A)$ of polynomial functions on A separates points of A: indeed if we have distinct points $\mathbf{x} \neq \mathbf{y}$ of A, then for at least one i with $1 \leq i \leq N$ we have that the ith coordinates of $\mathbf{x}$ and $\mathbf{y}$ differ. But the ith coordinate function $(x_1, \ldots, x_N) \mapsto x_i$ *is* a polynomial function, so $\mathcal{P}(A)$ separates points of A almost tautologically.[5]

Remarkably, when A is closed and bounded, the converse of Proposition 3.20 holds. This is a special case of a celebrated result of M. Stone [**St37**], [**St48**]:

THEOREM 3.21 (Stone-Weierstrass). *Let $X \subset \mathbb{R}^N$ be closed and bounded, and let $\mathcal{A}$ be an $\mathbb{R}$-subalgebra of $\mathcal{C}(A)$ that separates points. Then $\mathcal{A}$ is dense in $\mathcal{C}(A)$.*

We will not prove Theorem 3.21 here, but see e.g. [**R**, Theorem 7.2].

Since the $\mathbb{R}$-algebra $\mathcal{P}(A)$ of polynomial functions on A separates points, the Stone-Weierstrass Theorem in $\mathbb{R}^N$ immediately implies the Weierstrass Theorem in $\mathbb{R}^N$. However, it is considerably more general. We give two examples.

PROPOSITION 3.22. *Let $f : [a,b] \to \mathbb{R}$ be continuous and injective. Let*

$$\mathcal{A}(f) := \{a_n f^n + \ldots + a_1 f + a_0 : [a,b] \to \mathbb{R} \mid a_0, \ldots, a_n\}$$

be the set of polynomial functions in f. Then $\mathcal{A}(f)$ is dense in $\mathcal{C}([a,b])$.

PROOF. It is immediate that $\mathcal{A}(f)$ is an $\mathbb{R}$-subalgebra of $\mathcal{C}(A)$. Moreover, to say that f is injective is to say that f itself separates points of $[a,b]$: for all $x \neq y$ in $[a,b]$ we have $f(x) \neq f(y)$. So certainly $\mathcal{A}(f)$ separates points of $[a,b]$. By the Stone-Weierstrass Theorem we have that $\mathcal{A}(f)$ is dense in $\mathcal{C}([a,b])$. $\square$

Notice that if in Proposition 3.22 we take $f(x) = x$, then we recover the Weierstrass Approximation Theorem. Taking e.g. $f(x) = e^x$, we get: every continuous function is a uniform limit of polynomials in e^x. For any positive odd $d \in \mathbb{Z}^+$, taking $f(x) = x^d$ we see that every continous function is a limit of polynomials in x^d, i.e.,

[5]In fact, if $\mathcal{A}$ is an algebra of functions on a subset A of $\mathbb{R}^N$ and B is a subset of A, then it makes sense to restrict each element of $\mathcal{A}$ to B to get $\mathcal{A}|_B$, an algebra of functions on B. It is immediate that if $\mathcal{A}$ separates points of A, then $\mathcal{A}|_B$ separates points of B. So in the above discussion the key case is that of $A = \mathbb{R}^N$.

polynomials in which each power of x is divisible by d. If d is even, then x^d is not injective, so this doesn't work. The following result addresses the $d = 2$ case:

THEOREM 3.23. *Let $R > 0$. A continuous function $f : [-R, R] \to \mathbb{R}$ is* ***even*** *if $f(-x) = f(x)$ for all $x \in [-R, R]$. Then every even continuous function is a uniform limit of even polynomial functions.*

PROOF. First we observe that every continuous function $f : [0, R] \to \mathbb{R}$ extends uniquely to an even continuous function $F : [-R, R] \to \mathbb{R}$: for $x \in [-R, 0)$ we may — and must — take $f(x) = f(-x)$. Let $\mathcal{P}_e$ be the set of all even polynomial functions on $[-R, R]$. Then $\mathcal{P}_e$ consists of polynomials $\sum_{k=0}^n a_k x^{2k}$ in which only even degree powers of x appear: see Exercise 3.16. Thus in the notation of Proposition 3.22 we have $\mathcal{P}_e = \mathcal{A}(x^2)$.

Although the injectivity hypothesis of Proposition 3.22 does not apply to $f(x) = x^2$ on $[-R, R]$, it does apply on $[0, R]$, so that result implies that $\mathcal{P}_e|_{[0,R]}$ is dense in $\mathcal{C}([0, R])$. Finally, if $f : [-R, R] \to \mathbb{R}$ is continuous and even and $\{P_n\}$ is a sequence of even polynomials converging uniformly to f on $[0, R]$, then the evenness implies that also $P_n \stackrel{u}{\to} f$ on $[-R, R]$. $\square$

One may wonder why Proposition 3.22 was stated only in the one variable case. The argument will hold verbatim with $[a, b]$ replaced by any closed, bounded subset in $\mathbb{R}^N$. However, when $N \geq 2$ injective continuous functions on subsets of $\mathbb{R}^N$ are much harder to come by: it follows from Exericse 3.17 that for any nondegenerate closed box $\mathcal{B} = \prod_{i=1}^N [a_i, b_i]$ (nondegenerate means $a_i < b_i$ for all i), there is no continuous injection $f : \mathcal{B} \to \mathbb{R}$.

Exercises.

EXERCISE 3.13. *Prove Lemma 3.17.*

EXERCISE 3.14. *Let $f : [a, b] \to \mathbb{R}$ be continuous, and let $c \in [a, b]$. Show: f is the uniform limit of a sequence of polynomials $\{P_n : [a, b] \to \mathbb{R}\}$ such that $P_n(c) = f(c)$ for all $c \in [a, b]$.*

EXERCISE 3.15. *Prove Proposition 3.20.*

EXERCISE 3.16. *Let $P : \mathbb{R} \to \mathbb{R}$ be a polynomial function.*

a) *Suppose that P is an even function: for all $x \in \mathbb{R}$, $P(-x) = x$. Show: $P(x) = \sum_{k=0}^n a_k x^{2k}$.*
b) *Suppose that P is an odd function: for all $x \in \mathbb{R}$, $P(-x) = -x$. Show: $P(x) = \sum_{k=0}^n a_k x^{2k+1}$.*

EXERCISE 3.17.

a) *Let $C := \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$ be the unit circle in $\mathbb{R}^2$. Show that there is no continuous,* injective *function $f : C \to \mathbb{R}$.*
(Hint: we may assume that f is nonconstant, so by the Extreme Value Theorem f assumes a minimum value m and a maximum value $M > m$. Show: every $L \in (m, M)$ is assumed at least twice on C.)
b) *Let X be a subset of $\mathbb{R}^N$ such that there is a continuous injection $\iota : C \hookrightarrow X$. Show: there is no injective continuous function $f : X \to \mathbb{R}$.*
c) *Let $N \geq 2$, and let X be a subset of $\mathbb{R}^N$ that contains an open ball. Show: there is no continuous injection $\iota : X \to \mathbb{R}$.*

Problems.

PROBLEM 3.5. *Let $\{P_n : \mathbb{R} \to \mathbb{R}\}_{n=1}^{\infty}$ be a sequence of polynomials converging uniformly on $\mathbb{R}$ to a function $f : \mathbb{R} \to \mathbb{R}$.*

a) *Show: there is $N \in \mathbb{Z}^+$ such that $P_n - P_N$ is constant for all $n \geq N$. (Hint: a polynomial is bounded on $\mathbb{R}$ if and only if it is constant.)*
b) *Deduce that there is $C \in \mathbb{R}$ such that $f = P_N + C$.*

In particular, a sequence of polynomials that is uniformly convergent on $\mathbb{R}$ has eventually constant degree and the limit function is also a polynomial.

PROBLEM 3.6. *Let $a < b$ and $c < d$ be real numbers.*

a) *Show: there is a unique linear function $\ell : \mathbb{R} \to \mathbb{R}$ such that $\ell(a) = c$ and $\ell(b) = d$. Show for that for this function we have $\ell([a, b]) = [c, d]$.*
b) *We define a function $\mathcal{L} : \mathcal{C}[c, d] \to \mathcal{C}[c, d]$ by mapping a continuous function $f : [c, d] \to \mathbb{R}$ to $f \circ \ell : [a, b] \to \mathbb{R}$. Show that $\mathcal{L}$ is a bijection. (Hint: the inverse function is the same function with $[a, b]$ and $[c, d]$ reversed.) Show also that $f : [c, d] \to \mathbb{R}$ is a polynomial function if and only if $\mathcal{L}(f)$ is a polynomial function.*
c) *Let $\{f_n : [c, d] \to \mathbb{R}\}_{n=1}^{\infty}$ be a sequence of functions and let $f : [c, d] \to \mathbb{R}$ be a function. Show: $f_n \overset{u}{\to} f$ on $[c, d] \iff \mathcal{L}(f_n) \overset{u}{\to} \mathcal{L}(f)$ on $[a, b]$.*
d) *Deduce: every element of $\mathcal{C}[a, b]$ is a uniform limit of polynomials if and only if every element of $\mathcal{C}[c, d]$ is a uniform limit of polynomials.*

PROBLEM 3.7. *Let $A \subseteq \mathbb{R}$. Show that the following are equivalent:*

(i) *Every continuous function $f : A \to \mathbb{R}$ is a uniform limit of polynomial functions.*
(ii) *The set A is closed and bounded.*

You may use without proof that every continuous function on a closed subset of $\mathbb{R}$ extends continuously to all of $\mathbb{R}$.

5. A Continuous, Nowhere Differentiable Function

We are going construct a function $f : \mathbb{R} \to \mathbb{R}$ with the following striking property: for all $x_0 \in \mathbb{R}$, f is continuous at x_0 but f is *not* differentiable at x_0. In short, we say that f is *continuous but nowhere differentiable.*

The first such construction (accompanied by a complete, correct proof) was given in a seminal 1872 paper of Weierstrass. Weierstrass's example was as follows: let $\alpha \in (0, 1)$, and let b be a positive odd integer such that $\alpha b > 1 + \frac{3\pi}{2}$. Then the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \sum_{n=0}^{\infty} \alpha^n \cos(b^n \pi x) \tag{11}$$

is continuous on $\mathbb{R}$ but not differentiable at any $x \in \mathbb{R}$.

By far the easier part of this is to show that f is continuous: this is Exercise 3.18. The proof that f is nowhere differentiable is not so easy, as indicated by the rather specific conditions given on the parameters α, b. (For less carefully chosen α, b the function f can have a "small" set of points of differentiability.) Thus, as

with most other contemporary treatments, we will switch to a different function for which the nowhere differentiability calculation is more straightforward. More specifically, we will switch from trigonometric functions to piecewise linear functions, so first we interpose the following result nailing down some further (simple) properties of these functions.

LEMMA 3.24.

a) *Let $f : [a, b] \to \mathbb{R}$ be a piecewise linear function with slopes $m_1, \ldots, m_n$. Show that f is Lipschitz, and the smallest possible Lipschitz constant is $C = \max_i |m_i|$.*
b) *Let $f : \mathbb{R} \to \mathbb{R}$ be a function. Suppose that there is $C > 0$ such that for every closed subinterval $[a, b]$ of $\mathbb{R}$, C is a Lipschitz constant for the restriction of f to $[a, b]$. Show that C is a Lipschitz constant for f.*
c) *Let $f : \mathbb{R} \to \mathbb{R}$ be a piecewise linear function with "corners" at the integers – i.e., f is differentiable on $(n, n + 1)$ for all $n \in \mathbb{Z}^+$ and is not differentiable at any integer n. For $n \in \mathbb{Z}$, let m_n be the slope of f on the interval $(n, n + 1)$. Let $C = \sup_{n\in\mathbb{Z}} m_n$. Then f is Lipschitz if and only if $C < \infty$, in which case C is the smallest Lipschitz contant for f.*

You are asked to prove Lemma 3.24 in Exercise 3.8.

Now we begin our construction with the "sawtooth function" $S : \mathbb{R} \to \mathbb{R}$: the unique piecewise linear function with corners at the integers and such that $S(n) = 0$ for every even integer n and $S(n) = 1$ for every odd integer n. The slopes of S are all ± 1, so by the preceding exercise S is Lipschitz (hence continuous):

$$\forall x, y \in \mathbb{R}, |S(x) - S(y)| \leq |x - y|.$$

Also S is 2-periodic: for all $x \in \mathbb{R}$, $S(x + 2) = S(x)$. For $k \in \mathbb{N}$, define

$$f_k : \mathbb{R} \to \mathbb{R},\ f_k(x) = \left(\frac{3}{4}\right)^k S(4^k x).$$

We suggest that the reader sketch the graphs of the functions f_k: roughly speaking they are sawtooth functions which, as k increases, oscillate more and more rapidly but with smaller amplitude: indeed $||f_k(x)|| = \left(\frac{3}{4}\right)^k$. We define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \sum_{k=0}^{\infty} f_k(x) = \sum_{k=0}^{\infty} \left(\frac{3}{4}\right)^k S(4^k x).$$

Since $\sum_{k=0}^{\infty} ||f_k|| = \sum_{k=0}^{\infty} \left(\frac{3}{4}\right)^k < \infty$, the series defining f converges uniformly by the Weierstrass M-Test. This also gives that f is continuous, since f is a uniform limit of a sequence of continuous functions. We claim however that f is nowhere differentiable. To see this, fix $x_0 \in \mathbb{R}$. We will define a sequence $\{\delta_n\}$ of nonzero real numbers such that $\delta_n \to 0$ and the sequence

$$D_n = \frac{f(x_0 + \delta_n) - f(x_0)}{\delta_n}$$

is divergent. This implies that f is not differentiable at x_0.

Let's do it. First suppose that the fractional part of x_0 lies in $[0, \frac{1}{2})$, so that the interval $(x_0, x_0 + \frac{1}{2})$ contains no integers. In this case we put

$$\delta_n = \frac{4^n}{2},$$

and the reason for our choice is that the interval $(4^n x_0, 4^n(x_0+\delta_n))$ contains no integers. Let $k, n \in \mathbb{N}$. We claim the following inequalities:

$$\forall k > n,\ |S(4^k x_0 + 4^k \delta_n) - S(4^k x_0)| = 0. \tag{12}$$

$$\forall k = n,\ |S(4^k x_0 + 4^k \delta_n) - S(4^k x_0)| = \frac{1}{2}. \tag{13}$$

$$\forall k < n,\ |S(4^k x_0 + 4^k \delta_n) - S(4^k x_0)| \leq |4^k \delta_n|. \tag{14}$$

Indeed: (12) holds because if $k > n$ then

$$4^k x_0 + 4^k \delta_n - 4^k x_0 = 4^k \delta_n = \frac{4^{k-n}}{2}$$

is a multiple of 2 and S is a 2-periodic function; (13) holds because if $k = n$ then

$$4^k x_0 + 4^k \delta_n = 4^k x_0 + \frac{1}{2},$$

so by our choice of δ_n, the function S is linear on $[4^k x_0, 4^k x_0 + \frac{1}{2}]$ of slope ± 1, hence the difference between its values at the endpoints is $\frac{\pm 1}{2}$. Finally, (14) holds because 1 is a Lipschitz constant for S. Using these results and the Reverse Triangle Inequality gives

$$\begin{aligned}
\left| \frac{f(x_0+\delta_n) - f(x_0)}{\delta_n} \right| &= \left| \sum_{k=0}^{n} \left(\frac{3}{4}\right)^k \frac{S(4^k x_0 + 4^k \delta_n) - S(4^k x_0)}{\delta_n} \right| \\
&\geq \left(\frac{3}{4}\right)^n 4^n - \sum_{k=0}^{n-1} \left(\frac{3}{4}\right)^k \cdot \left| \frac{S(4^k x_0 + 4^k \delta_n) - S(4^k x_0)}{\delta_n} \right| \\
&\geq 3^n - \sum_{k=0}^{n-1} 3^k = 3^n - \frac{3^n - 1}{2} \geq \frac{3^n}{2}.
\end{aligned}$$

Thus $D_n \to \infty$, so f is not differentiable at x_0.

We're not quite done: recall that we assumed that the fractional part of x_0 lay in $[0, \frac{1}{2})$, with the consequence that S was linear on the interval $[(4^n x_0, 4^n(x_0+\delta_n)]$. What to do if the fractional part of x_0 lies in $[\frac{1}{2}, 1)$? In this case we take $\delta_n = \frac{-4^n}{2}$ so that the interval $(4^n(x_0+\delta_n), 4^n x_0)$ contains no integers so S is linear on the interval $[4^n(x_0+\delta_n), 4^n x_0]$, and the rest of the proof goes through as above.

So, albeit with a different function, we have proved Weierstrass's Theorem.

THEOREM 3.25. *(Weierstrass, 1872) There is a function $f : \mathbb{R} \to \mathbb{R}$ that is continuous at every point of $\mathbb{R}$ but differentiable at no point of $\mathbb{R}$.*

Notice that if we restrict f to some closed interval, say $[0, 2]$, then by the Weierstrass Approximation Theorem f is – like any continuous function on $[0, 2]$ – a uniform limit of polynomials. Thus even a uniform limit of polynomials on a closed, bounded interval need not have any good differentiability properties whatsoever!

Exercises.

EXERCISE 3.18. *Show that the function defined by (11) above is continuous.*

Problems.

PROBLEM 3.8. *Prove Lemma 3.24.*

6. Some Further Exercises and Problems

Exercises.

EXERCISE 3.19. *Let $A \subseteq \mathbb{R}^N$, and let $\{f_n : A \to \mathbb{R}\}_{n=0}^{\infty}$ be a sequence of functions such that $f_n \xrightarrow{u} f$ on A.*

a) *Suppose that the set $\{n \in \mathbb{N} \mid f_n$ is continuous$\}$ is infinite. Show: the function f is continuous.*
b) *Give an example of a sequence $f_n : \mathbb{R} \to \mathbb{R}$ such that each f_n is discontinuous at* every *point of $\mathbb{R}$ that converges uniformly to the continuous function $f = 0$.*

EXERCISE 3.20. *This exercise concerns ratio and root test upper and lower limits for the series $\sum_{n=0}^{\infty} 2^{-n+(-1)^n}$.*

a) *Show that $\underline{\rho} = \frac{1}{8}$ and $\overline{\rho} = 2$. Deduce that the Ratio Test fails to determine the convergence of the series.*
b) *Show that $\underline{\theta} = \overline{\theta} = \frac{1}{2}$. Deduce from the Root Test that the series is absolutely convergent.*

EXERCISE 3.21. *Construct a convergent series $\sum_n a_n$ with positive terms such that the upper Ratio Test limit $\overline{\rho} = \overline{\lim} \frac{a_{n+1}}{a_n}$ is infinite.*

Problems.

PROBLEM 3.9. *Let $A \subseteq \mathbb{R}^N$, and let $\{f_n : A \to \mathbb{R}\}_{n=0}^{\infty}$ be a sequence of functions such that $f_n \xrightarrow{u} f$ on A. Suppose that the set $\{n \in \mathbb{N} \mid f_n$ is uniformly continuous$\}$ is infinite. Show: the function f is uniformly continuous.*

PROBLEM 3.10. *Let $R \in (0, \infty)$.*

a) *Exhibit a power series $\sum_{n=0}^{\infty} a_n x^n$ with domain $(-R, R)$.*
b) *Exhibit a power series $\sum_{n=0}^{\infty} a_n x^n$ with domain $[-R, R)$.*
c) *Exhibit a power series $\sum_{n=0}^{\infty} a_n x^n$ with domain $(-R, R]$.*
d) *Exhibit a power series $\sum_{n=0}^{\infty} a_n x^n$ with domain $[-R, R]$.*

PROBLEM 3.11. *Prove the Root Test (Theorem 3.11).*

PROBLEM 3.12. *Let $\sum_n a_n$ be any series with $a_n \neq 0$ for all n. Prove (10) relating the Ratio and Root Test lower and upper limits.*

CHAPTER 4

Real Induction and Compactness

1. Real Induction

1.1. Statement and first application. A subset $S \subseteq [a, b]$ is **inductive** if:

(RI1) $a \in S$.
(RI2) If $a \leq x < b$, then $x \in S \implies [x, y] \subseteq S$ for some $y > x$.
(RI3) If $a < x \leq b$ and $[a, x) \subseteq S$, then $x \in S$.

THEOREM 4.1. *(Real Induction) For $S \subseteq [a, b]$, the following are equivalent:*

(i) *The set S is inductive.*
(ii) *We have $S = [a, b]$.*

PROOF. (i) $\implies$ (ii): let $S \subseteq [a, b]$ be inductive. Seeking a contradiction, suppose $S' = [a, b] \setminus S$ is nonempty, so $\inf S'$ exists and is finite.
Case 1: $\inf S' = a$. Then by (RI1), $a \in S$, so by (RI2), there exists $y > a$ such that $[a, y] \subseteq S$, and thus y is a greater lower bound for S' then $a = \inf S'$: contradiction.
Case 2: $a < \inf S' \in S$. If $\inf S' = b$, then $S = [a, b]$. Otherwise, by (RI2) there exists $y > \inf S'$ such that $[\inf S', y] \subseteq S$, contradicting the definition of $\inf S'$.
Case 3: $a < \inf S' \in S'$. Then $[a, \inf S') \subset S$, so by (RI3) $\inf S' \in S$: contradiction!
(ii) $\implies$ (i) is immediate. $\square$

Theorem 4.1 is due to D. Hathaway [**Ha11**] and, independently, to me [**Cl19**]. It is really an equivalent formulation of Dedekind completeness (see [**Cl19**, Thm. 14]), but just as Mathematical Induction provides a platform or structure that can be extremely helpful in "getting a clue" as to what kind of argument to make, Real Induction provides a similar structure that, with some practice, makes proofs of some of the major results in undergraduate real analysis and topology become rather routine. Here is a first example.

THEOREM 4.2 (Intermediate Value Theorem). *Let $f : [a, b] \to \mathbb{R}$ be continuous. If M lies in between $f(a)$ and $f(b)$, then there is $c \in [a, b]$ such that $f(c) = M$.*

PROOF. We will show the following: if $f : [a, b] \to \mathbb{R} \setminus \{0\}$ is continuous and $f(a) > 0$, then also $f(b) > 0$. Assuming this, it follows that if $f : [a, b] \to \mathbb{R}$ is continuous and moreover $f(a) > 0$ and $f(b) < 0$, then there must be $c \in (a, b)$ such that $f(c) = 0$ (for if not, the preceding claim implies $f(b) > 0$). In turn, if we apply this second claim to $\pm(f(x) - M)$ then we get the full result.

Let $S := \{x \in [a, b] \mid f(x) > 0\}$. Our strategy of proof is to show that S is inductive; then by Real Induction we will have $S = [a, b]$, and that $b \in S$ means $f(b) > 0$, the desired conclusion.
(RI1) By hypothesis we have $f(a) > 0$, meaning that $a \in S$.
(RI2) Let $x \in [a, b)$ and suppose that $x \in S$: that is, $f(x) > 0$. Because f is

continuous, there is some $\delta > 0$ such that f remains positive on $[x-\delta, x+\delta] \cap [a,b]$. Choosing δ small enough so that $x + \delta \leq b$ we get that f is positive on $[x, x+\delta]$.
(RI3) Let $x \in (a,b]$ and suppose that $f(y)$ is positive for all $y \in [a,b]$ with $y < x$. We want to show that $f(x) > 0$. Because $f(x) \neq 0$ by hypothesis, the only other possibility is that $f(x) < 0$. But as above, this means that there is $\delta < 0$ such $a \leq x - \delta$ and f is negative on $[x-\delta, x)$, which contradicts our assumption. $\square$

We will give a more significant application of Real Induction in the following section.

2. A Mean Value Inequality For All Functions

Once, over the course of a single week I attended five short talks on this linchpin of calculus:

THEOREM 4.3 (Mean Value Theorem, henceforth "MVT"). *Let $f : [a,b] \to \mathbb{R}$ be a function. Suppose that f is continuous and that its restriction to (a,b) is differentiable. Then there is $c \in (a,b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \tag{15}$$

As one of the speakers pointed out, MVT is *not* just the equation (15): rather, (15) is its conclusion, which holds under the continuity and differentiability hypotheses.

Most freshman calculus students will struggle to find meaning in MVT unless it is placed in good context. The speakers did this very well. All of them gave the geometric interpretation involving a tangent line parallel to the secant line. Most also gave the physical interpretation: suppose $f(x)$ gives the position of a particle at time x. Then $f(b) - f(a)$ is the displacement of the particle between time a and time b, while $b - a$ is the elapsed time, so the quantity $\frac{f(b)-f(a)}{b-a}$ is the particle's *average* velocity on the time interval $[a,b]$. In turn $f'(c)$ can be interpreted as the *instantaneous* velocity at time c. Thus MVT asserts that the average velocity is also equal to the instantaneous velocity for at least one point in time.

Several of the speakers gave the following application: **MVT could be used by the highway patrol in order to award speeding tickets even when the speeding is not directly witnessed.** Namely, if the authorities know that at time a your position is $f(a)$ – perhaps you pass a camera that views your license plate – and also that at time $b > a$ your position is $f(b)$, they can apply MVT to conclude that at some point in between your instantaneous velocity must have been $\frac{f(b)-f(a)}{b-a}$. So if for the entire stretch of the highway between $f(a)$ and $f(b)$ the speed limit was at most $M < \frac{f(b)-f(a)}{b-a}$, they can write you a ticket.

So far this is quite familiar. But during the last lecture a new thought poked through: there is a snag that I didn't see for years. Do you see it now?

Here it is: in order to apply MVT we need to know that f is continuous on $[a,b]$ and differentiable on (a,b). Is it clear that one must drive in a differentiably?

I think it is not! For instance, if you are driving at 30 mph and get rear ended, that will instantly bump up your speed. It seems natural to model this position function as having a corner point at the point of collision. But even one point of nondifferentiability can falsify the conclusion of MVT: e.g. the function

$$f : [-1,1] \to \mathbb{R},\ x \mapsto |x|$$

is differentiable at every point other than $x = 0$ with derivative either 1 or -1: so $f'(c)$, when it exists, gets nowhere near the average velocity, which is 0.

If called to testify on the nondifferentiability problem in traffic court, I would say: while it's highly plausible that a driver's position must be a continuous function of time, differentiability feels like a simplifying assumption made so as to mathematically model a real world situation. There are also results suggesting that a "generic" continuous function is not differentiable.[1]

What is your reaction to this objection? Mine is to question our apparent definition of "exceeding the speed limit of M" as $f'(c) > M$ for at least one c in the interval (a, b). Isn't this more an *interpretation* of the derivative than a principled requirement that it exist?

Here is a definition of exceeding a speed limit M that applies to any function $f : [a, b] \to \mathbb{R}$ whatsoever: for $M > 0$, a function $f : [a, b] \to \mathbb{R}$ is an **M-speeder** if for all $\delta > 0$ there are $c, d \in [a, b]$ with $0 < d - c < \delta$ such that $\frac{|f(d)-f(c)|}{d-c} > M$. Thus you are an M-speeder if there are arbitrarily short subintervals of $[a, b]$ on which your average speed exceeds M. This definition seems physically appealing and even in accordance with the type of unsafe driving speed limits are designed to prevent: covering too much ground in too short a time for you and others to respond appropriately to changes in traffic conditions.

And here our main result:

THEOREM 4.4. *If $f : [a, b] \to \mathbb{R}$ is a function, then f is an M-speeder for all $0 < M < \frac{|f(b)-f(a)|}{b-a}$.*

Thus the indirect procedure for awarding (so to speak) speeding tickets seems justified after all.

We give the proof of Theorem 4.4 in §2. In §3 we explore some complements, including a version of Theorem 4.4 for functions taking values in any metric space. In §4 we will discuss more conventional Mean Value Inequalities and see that they are implied by our results.

2.1. The Proof. Our proof will use the analogue of the limit superior and limit inferior of a real sequence for functions $g : (-\delta, \delta) \setminus \{0\} \to \mathbb{R}$. Namely we put

$$\overline{\lim}_{x\to 0} g(x) := \inf_{0<\epsilon\leq\delta} \left(\sup_{0<|x|<\epsilon} g(x) \right) \in [-\infty, \infty],$$

$$\underline{\lim}_{x\to 0} g(x) := \sup_{0<\epsilon\leq\delta} \left(\inf_{0<|x|<\epsilon} g(x) \right) \in [-\infty, \infty].$$

The usual limit $\lim_{x\to 0} g(x)$ exists in $[-\infty, \infty]$ and only if $\overline{\lim}_{x\to 0} g(x) = \underline{\lim}_{x\to 0} g(x)$. If g is defined only on $(0, \delta)$ (resp. only on $(-\delta, 0)$) we can still define $\overline{\lim}_{x\to 0} g(x)$ and $\underline{\lim}_{x\to 0}$ just by restricting to $0 < x < \epsilon$ (resp. to $-\epsilon < x < 0$).

Here is a $\overline{\lim}_{x\to 0}$ cheatsheet: we have $\overline{\lim}_{x\to 0} g(x) = \infty$ if and only if g is unbounded above on every interval $(-\delta, \delta) \setminus \{0\}$. If $\overline{\lim}_{x\to 0} g(x) = L \in \mathbb{R}$, then

[1] A "generic" continuous function $f : [a, b] \to \mathbb{R}$ is differentiable *at no point* in the sense that the somewhere differentiable functions form a meager subset of the Baire space of all continuous functions [**?**, §49].

L is the least real number with the property that for all $M > L$ there is $\epsilon > 0$ such that $g(x) \leq M$ for all $0 < |x| < \epsilon$. We have $\overline{\lim}_{x\to 0} g(x) = -\infty$ if and only if $\lim_{x\to 0} g(x) = -\infty$ in the usual sense. One can get the corresponding $\underline{\lim}_{x\to 0}$ cheatsheet using the fact that

$$\underline{\lim}_{x\to 0} g(x) = -\overline{\lim}_{x\to 0} - g(x).$$

For $f : [a, b] \to \mathbb{R}$ and $x \in [a, b]$, we put

$$\overline{D}_f(x) := \overline{\lim}_{h\to 0} \frac{f(x+h) - f(x)}{h} \in [-\infty, \infty],\ \underline{D}_f(x) := \underline{\lim}_{h\to 0} \frac{f(x+h) - f(x)}{h} \in [-\infty, \infty].$$

Thus f is differentiable at $x \in [a, b]$ – in the one-sided sense at the endpoints – if and only if $\overline{D}_f(x)$ and $\underline{D}_f(x)$ are equal and finite. However both $\overline{D}_f$ and $\underline{D}_f$ are defined for arbitrary f.[2]

All of the content of Theorem 4.4 resides in the following result.

THEOREM 4.5. *Let $f : [a, b] \to \mathbb{R}$, and let $A, B \in \mathbb{R}$.*

a) *If $\overline{D}_f(x) \leq A$ for all $x \in [a, b]$, then $\frac{f(b)-f(a)}{b-a} \leq A$.*

b) *If $\underline{D}_f(x) \geq B$ for all $x \in [a, b]$, then $\frac{f(b)-f(a)}{b-a} \geq B$.*

PROOF. a) For $\epsilon > 0$, put

$$S_\epsilon := \{x \in [a, b] \mid f(x) - f(a) \leq (A + \epsilon)(x - a)\}.$$

It suffices to show:

$$\forall \epsilon > 0,\ S_\epsilon = [a, b]. \tag{16}$$

For if so, then for all $\epsilon > 0$, because $b \in S_\epsilon$ we have

$$f(b) - f(a) \leq (A + \epsilon)(b - a),$$

and it follows that $\frac{f(b)-f(a)}{b-a} \leq A$. We will show that $S_\epsilon = [a, b]$ by Real Induction.
(RI1): It is immediate that $a \in S_\epsilon$.
(RI2): Let $x \in [a, b) \cap S_\epsilon$. Since $\overline{D}_f(x) \leq A$, there is $\delta > 0$ such that for all $y \in [x, x + \delta]$ we have $\frac{f(y)-f(x)}{y-x} \leq A + \epsilon$, and thus we have

$$f(y)-f(a)-(f(y)-f(x))+(f(x)-f(a)) \leq (A+\epsilon)(y-x)+(A+\epsilon)(x-a) = (y-a)(A+\epsilon),$$

so $[x, x + \delta] \subseteq S_\epsilon$.
(RI3): Let $x \in (a, b]$ and suppose that $[a, x) \subseteq S_\epsilon$. Since $\overline{D}_f(x) \leq A$, there is $\delta > 0$ such that for all $y \in [x - \delta, x]$ we have $\frac{f(x)-f(y)}{x-y} \leq A + \epsilon$, and thus we have

$$f(x)-f(a) = (f(x)-f(y))+(f(y)-f(a)) \leq (A+\epsilon)(x-y)+(A+\epsilon)(y-a) = (x-a)(A+\epsilon),$$

so $x \in S_\epsilon$.
b) This is very similar to part a). Or, since $\underline{D}_{-f}(x) = -\overline{D}_f(x)$, we can reduce to part a). □

Proof of Theorem 4.4: if $0 < M < \frac{|f(b)-f(a)|}{b-a}$, then either $\frac{f(b)-f(a)}{b-a} > M$ or $\frac{f(b)-f(a)}{b-a} < -M$.

• If $\frac{f(b)-f(a)}{b-a} > M$, then by Theorem 4.5a) we have $\overline{D}_f(c) > M$ for some $c \in [a, b]$, so for all $\delta > 0$ there is $x \in [a, b]$ with $0 < |c - x| < \delta$ such that

[2]These are two-sided versions of the **Dini derivatives**, though you need not know what that means.

$\frac{f(x)-f(c)}{x-c} > M$.

• If $\frac{f(b)-f(a)}{b-a} < -M$, then by Theorem 4.5b) we have $\underline{D}_f(c) < -M$ for some $c \in [a,b]$, so for all $\delta > 0$ there is $x \in [a,b]$ with $0 < |c-x| < \delta$ such that $\frac{f(x)-f(c)}{x-c} < -M$.

REMARK. *We actually proved something stronger than Theorem 4.4, namely: for $f : [a,b] \to \mathbb{R}$, suppose there is a real number $0 < M < \frac{|f(b)-f(a)|}{b-a}$. Then:*

a) *If $f(b) > f(a)$, there is $c \in [a,b]$ such that for all $\delta > 0$ there is $x \in [a,b]$ such that $0 < |c-x| < \delta$ and $\frac{f(x)-f(c)}{x-c} > M$.*
b) *If $f(b) < f(a)$, there is $c \in [a,b]$ such that for all $\delta > 0$ there is $x \in [a,b]$ such that $0 < |c-x| < \delta$ and $\frac{f(x)-f(c)}{x-c} < -M$.*

2.2. Consequences. If f is continuous, we get a slightly stronger conclusion:

THEOREM 4.6. *Let $f : [a,b] \to \mathbb{R}$ be continuous, and let $0 < M < \frac{|f(b)-f(a)|}{b-a}$. Then there is $c \in (a,b)$ such that either:*

(i) *For all $\delta > 0$, there is $x \in [a,b]$ such that $0 < |c-x| < \delta$ and $\frac{f(x)-f(c)}{x-c} > M$; or*
(ii) *For all $\delta > 0$, there is $x \in [a,b]$ such that $0 < |c-x| < \delta$ and $\frac{f(x)-f(c)}{x-c} < -M$.*

Moreover (i) occurs if $f(b) > f(a)$ and (ii) occurs if $f(b) < f(a)$.

PROOF. After Remark 2.1, all that is left is to show that we can take $c \in (a,b)$. For this: cine f is continuous and $M < \frac{|f(b)-f(a)|}{b-a}$, there are $a < a' < b' < b$ such that $M < \frac{|f(b')-f(a')|}{b'-a'}$. Applying Remark 2.1 to the restriction of f to $[a', b'$ gives $c \in [a'b'] \subset (a,b)$. □

And if we put back the differentiability assumption, we get:

COROLLARY 4.7 (Mean Value Inequality, or MVI). *Let $f : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). If $m \leq f'(x) \leq M$ for all $x \in (a,b)$, then*

$$m(y-x) \leq f(y) - f(x) \leq M(y-x)$$

for all $x, y \in [a,b]$ with $x \leq y$.

PROOF. Suppose that the second inequality does not hold. Then there are $a \leq x < y \leq b$ with $f(y) - f(x) > M(y-x)$. Then there is $\epsilon > 0$ such that $f(y) - f(x) > (M+\epsilon)(y-x)$. By Theorem 4.6 there is $c \in (a,b)$ and a sequence of points $\{x_n\}$ in $[a,b] \setminus \{c\}$ converging to c such that $\frac{f(x_n)-f(c)}{x_n-c} > M + \epsilon$ for all $n \in \mathbb{Z}^+$. By assumption f is differentiable at c, so

$$f'(c) = \lim_{n\to\infty} \frac{f(x_n) - f(c)}{x_n - c} \geq M + \epsilon,$$

a contradiction. The first inequality is established in a very similar way. □

MVI is itself a corollary of the Mean Value Theorem (MVT). MVT is certainly not very difficult to prove: we covered it in the Fall 2022 course. My understanding is that you have seen MVT in the last course, but just to be sure Exercise 4.1 outlines the standard proof.

During my first work on Real Induction in 2010, I saw how to use it to prove

MVI. Our Theorem 4.4 is essentially the recognition (which came in 2022) that this argument is not really using the differentiability of f. I did not see in 2010 and still do not see now how to prove the MVT using Real Induction.[3]

Despite the fact that MVT is stronger than MVI, several leading mathematical expositors have argued that MVI is somehow "more natural" than MVT and that calculus texts should use MVI instead: this goes back at least to 1967 works of Bers [**Be67**] and Cohen [**Co67**] and has also been made by Dieudonné [**D**] and Boas [**Bo81**], among others. I never reeally understood the passion that several leading analysts brought to this; maybe by now the fuss has died down. The one thing I will say in their favor is that it turns out that many important applications of MVT one can use MVI instead. The exercises treat some classic examples.

Exercises.

EXERCISE 4.1. *Let $f : [a, b] \to \mathbb{R}$ be continuous and differentiable on (a, b).*

a) *Prove **Rolle's Theorem**: if $f(a) = f(b) = 0$, then there is $c \in (a, b)$ such that $f'(c) = 0$.*
(Hint: first deal with the case in which f assumes its maximum and minimum values at the endpoints, then treat the case in which some interior point c is either a maximum or minimum for f. You may use without proof that if f has a local maximum or minimum at $c \in (a, b)$, then $f'(c) = 0$.)
b) *Deduce MVT by subtracting a suitable linear function from f.*

EXERCISE 4.2. *Let $f : [a, b] \to \mathbb{R}$ be a function with $f(b) - f(a) > 0$, and let $0 < M < \frac{f(b)-f(a)}{b-a}$. By Remark 2.1, there is $c \in [a, b]$ such that for all $\delta > 0$ there is $x \in [a, b]$ such that $0 < |c - x| < \delta$ and $\frac{f(x)-f(c)}{x-c} < M$. According to Theorem 4.6, if f is moreover continuous, we may choose this c to lie in the open interval (a, b). Give an example of a discontinuous $f : [a, b] \to \mathbb{R}$ such that the only such c is a, and give another example such that the only such c is b.*

For the following exercises we need to introduce some terminology which will be at least roughly familiar but we need to make some precise distinctions: let $X \subseteq \mathbb{R}$ and let $f : X \to \mathbb{R}$. We say f is:

- **increasing** if for all $x_1, x_2 \in X$, $x_1 \leq x_2 \implies f(x_1) \leq f(x_2)$;
- **strictly increasing** if for all $x_1, x_2 \in X$, $x_1 < x_2 \implies f(x_1) < f(x_2)$;
- **decreasing** if for all $x_1, x_2 \in X$, $x_1 \leq x_2 \implies f(x_1) \geq f(x_2)$;
- **strictly decreasing** if for all $x_1, x_2 \in X$, $x_1 < x_2 \implies f(x_1) > f(x_2)$.

EXERCISE 4.3. *Let I be an interval, and let $f : I \to \mathbb{R}$ be differentiable.*

a) *Show: if $f'(x) \geq 0$ for all $x \in I$, then f is increasing.*
b) *Show: if $f'(x) > 0$ for all $x \in I$, then f is strictly increasing.*
c) *Show: if $f'(x) \leq 0$ for all $x \in I$, then f is decreasing.*
d) *Show: if $f'(x) < 0$ for all $x \in I$, then f is strictly decreasing.*

EXERCISE 4.4. *Let $f : I \to \mathbb{R}$ be a differentiable function.*

[3]To be sure, there are no negative consequences to this whatseover: again, pretty much everyone knows how to prove the Mean Value Theorem. But it is interesting to know whether or not something can be done, and so I still wonder...

a) *[Zero Velocity Theorem]*
Show: if $f'(x) = 0$ for all $x \in I$, then f is constant.
b) *[(Almost) Uniqueness of Antiderivatives] Suppose that $f, g : I \to \mathbb{R}$ are differentiable functions and that $f' = g'$. Show: there is $C \in \mathbb{R}$ such that $g = f + C$: i.e., for all $x \in I$, we have $g(x) = f(x) + C$.*

EXERCISE 4.5. *Let $k \in \mathbb{Z}^+$. Suppose that the kth derivative $f^{(k)}$ of f exists and is identically 0: $f^{(k)}(x) = 0$ for all $x \in I$. Show: f is a polynomial function of degree at most k. (Suggestion: use induction on k.)*

Problems.

PROBLEM 4.1. *A function $f : [a, b] \to \mathbb{R}$ is an* ***infinite-speeder*** *if it is an M-speeder for all $M > 0$.*

a) *Let $f : [a, b] \to \mathbb{R}$ be a function that is* not *continuous. Show: f is an infinite-speeder.*
b) *Show: $g : [-1, 1] \to \mathbb{R}$ by $g(x) = x^{\frac{1}{3}}$. Show: g is an infinite-speeder that is continuous and differentiable except at $x = 0$.*
c) *Define $h : [-1, 1] \to \mathbb{R}$ by*
$$h(x) = \begin{cases} x \mapsto x^2 \sin(\frac{1}{x^2}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$
Show: h is a differentiable infinite-speeder.
d) *Show: no Lipschitz function $f : [a, b] \to \mathbb{R}$ is an infinite-speeder. Deduce: if f has a continuous derivative, then it is not an infinite-speeder.*

Here is a version of Theorem 4.4 for speeding in Euclidean N-space:

PROBLEM 4.2. *Let $f : [a, b] \to \mathbb{R}^N$ be a function, and let $0 < M < \frac{||f(b)-f(a)||}{b-a}$.*

a) *Show: there is $c \in [a, b]$ such that for all $\delta > 0$ there is $x \in [a, b]$ such that $0 < |c - x| < \delta$ and $\frac{||f(x)-f(c)||}{x-c} > M$.*
b) *Show: if f is moreover continuous, then the c of part a) may be taken to lie in (a, b).*

PROBLEM 4.3. *Let I be an interval, and let $f : I \to \mathbb{R}$ be differentiable.*

a) *Suppose that $f'(x) \geq 0$ for all $x \in I$, so by Exercise 4.3, f is increasing. If we moreover had that $f'(x) > 0$ for all $x \in I$, then by Exercise 4.3b) f would be strictly increasing. However, f may be strictly increasing even when $f'(x) = 0$ for some $x \in I$. Show that the following are equivalent:*
   (i) *f is* not *strictly increasing.*
   (ii) *There are $a < b$ in I such that $f|_{[a,b]} : [a, b] \to \mathbb{R}$ is constant.*
b) *Use the criterion of part a) to show that for all odd integers $n \geq 1$, the function $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^n$ is strictly increasing.*
c) *State an analogous criterion for a function to be strictly decreasing. (You need not prove it.)*

3. Compactness

3.1. Definition and first properties. Let A be a subset of $\mathbb{R}^N$, and let $\{U_i\}_{i \in I}$ be an indexed family of subsets of $\mathbb{R}^N$. We say that $\{U_i\}_{i \in I}$ **covers** A (or **is a cover of A**) if $A \subseteq \bigcup_{i \in I} U_i$: in other words, every point of A lies in at least

one of the sets A_i. We say that a cover $\{U_i\}_{i \in I}$ of A **admits a finite subcover** if there is a finite subset $J \subseteq I$ such that $\{U_i\}_{i \in J}$ still covers A.

EXAMPLE 4.1. *Let A be a subset of $\mathbb{R}^N$. We claim that every cover of A admits a finite subcover if and only if A is finite.*

Indeed, suppose $A = \{a_1, \ldots, a_n\}$ is finite. Then if $\{U_i\}_{i \in I}$ is a cover of A, for $1 \leq j \leq n$ the element a_j must lie in U_i for at least one $i \in I$. Choose such an i and call it i_j. Then $J := \{i_1, \ldots, i_n\}$ is a finite subset of I such that $A \subseteq \bigcup_{i \in J} U_i$.

Conversely, suppose that A is infinite. For $a \in A$, let $U_a := \{a\}$. Take $I = A$. Clearly $\bigcup_{a \in I} U_a = A$, so we have a covering of A. In this case if we remove even a single element of I then we don't have a cover anymore, so there is no proper subset J of I such that $\bigcup_{i \in J} U_i \supseteq A$. In particular there is no finite subcover.

Thus for a subset A of $\mathbb{R}^N$ (or a subset of any set, for that matter), the condition that every cover of A admits a finite subcover is precisely the condition that A be finite. Now we will restrict to a certain class of covers using the topology of $\mathbb{R}^N$ and get a condition that can (helpfully, I feel, though it is up to you whether you agree) be viewed as a "topological finiteness condition." Here goes:

For a subset A of $\mathbb{R}^N$, an **open cover of A** is a cover $\{U_i\}_{i \in I}$ of A in which each U_i is an open subset of $\mathbb{R}^N$. We say that a subset $A \subseteq \mathbb{R}^N$ is **compact** if every open cover admits a finite subcover.

In Chapter 1 we introduced the concept of a subset A of $\mathbb{R}^N$ being sequentially compact – every sequence in A admits a subsequence that converges to an element of A – and found that A is sequentially compact if and only if it is closed and bounded, a result that is equivalent to the Bolzano-Weierstrass Theorem in $\mathbb{R}^N$. Compactness is a more abstract concept, involving indexed families of sets rather than just sequences and subsequences. Nevertheless, could it be that a subset of $\mathbb{R}^N$ is compact if and only if it is closed and bounded? Half of this is easy to prove:

PROPOSITION 4.8. *Let $A \subseteq \mathbb{R}^N$ be compact. Then A is closed and bounded.*

PROOF. First suppose that A is not closed, so there is a point $x \in \mathbb{R}^N \setminus A$ that is an accumulation point of A. For $n \in \mathbb{Z}^+$, let

$$U_n := \mathbb{R}^N \setminus B^\bullet(x, \frac{1}{n}).$$

Then

$$\bigcup_{n=1}^{\infty} U_n = \mathbb{R}^N \setminus \bigcap_{n=1}^{\infty} B^\bullet(x, \frac{1}{n}) = \mathbb{R}^N \setminus \{x\} \supseteq A,$$

so $\{U_n\}_{n=1}^{\infty}$ is an open cover of A. We also have $U_1 \subseteq U_2 \subseteq \ldots \subseteq U_n \subseteq$ so the union over any finite number of sets $U_{n_1}, \ldots, U_{n_k}$ is $U_{\max(n_1, \ldots, n_k)}$. However A contains points arbitrarily close to x, so A is not contained in any set U_n. Therefore the cover $\{U_n\}_{n=1}^{\infty}$ has no finite subcover, so A is not compact.

Now suppose that A is unbounded. For each $a \in A$, let $U_a := B^\circ(a, 1)$ be the open ball of radius 1 centered at a. Clearly $\{U_a\}_{a \in A}$ is an open cover of A. However each U_a is bounded, so the union of any finite number of U_a's is also bounded. Since A is unbounded we cannot have $A \subseteq \bigcup_{a \in J} U_a$ for any finite subset J of A, so this cover has no finite subcover and thus A is not compact. $\square$

The celebrated Heine-Borel Theorem is Proposition 4.8 together with its converse. We will give two proofs, one for $N = 1$ using real induction and one for all $N \in \mathbb{Z}^+$

using "lion-hunting." But first let us establish some other facts about compact subsets following what we already know about sequential compactness.

PROPOSITION 4.9. *Let A be a compact subset of $\mathbb{R}^N$.*

a) *If B is a closed subset of A, then B is also compact.*
b) *Let $f : A \to \mathbb{R}^M$ be a continuous function. Then $f(A)$ is a compact subset of $\mathbb{R}^M$.*

PROOF. a) Let $\{U_i\}_{i\in I}$ be an open cover of B. Since B is closed, $U_B := \mathbb{R}^N \setminus B$ is open. Consider the family $\tilde{U} := \{U_i \mid i \in I\} \cup \{U_B\}$ of open sets that consists of all of the U_i's together with U_B. Since every element of B is contained in some U_i and every other element of $\mathbb{R}^N$ is contained in U_B, we have $\bigcup_{i\in I} U_i \cup U_B = \mathbb{R}^N$. In particular $\tilde{U}$ is an open covering of the compact subset A, which means that it has a finite subcover: that is, there is a finite subset J of I such that

$$A \subseteq \bigcup_{i\in J} U_i \cup U_B.$$

(We don't know whether U_B lies in the finite subcover or not, but if it doesn't then adding it in we still get a finite subcover, so we may as well assume that it does.) Since B is a subset of A we have

$$B \subseteq \bigcup_{i\in J} U_i \cup U_B.$$

But U_B is disjoint from B, so also

$$B \subseteq \bigcup_{i\in J} U_i.$$

This means that our open cover of B has a finite subcover, so B is compact.
b) Let $\{V_i\}_{i\in I}$ be an open cover of $f(A)$ in $\mathbb{R}^M$. For each $i \in I$, let $U_i := f^{-1}(V_i) = \{x \in \mathbb{R}^N \mid f(x) \in V_i\}$. We claim that because V_i is open and f is continuous, then U_i is an open subset of $\mathbb{R}^N$. Indeed, let $x \in U_i$, so $f(x) \in V_i$. Since V_i is open in $\mathbb{R}^M$ there is $\epsilon > 0$ such that $B^\circ(f(x), \epsilon) \subseteq V_i$. By continuity there is $\delta > 0$ such that if $x' \in B^\circ(x, \delta)$, then $f(x') \in B^\circ(f(x), \epsilon) \subset V$, so $B^\circ(x, \delta) \subseteq f^{-1}(V_i) = U_i$. This shows that U_i is open. Next we claim that $\{U_i\}_{i\in I}$ is a cover of A. Indeed, if $a \in A$, then there is some $i \in I$ such that $f(a) \in V_i$, hence $a \in U_i$. Because A is compact, there is a finite subset $J \subseteq I$ such that $A \subseteq \bigcup_{i\in J} U_i$. Thus if a is in A, then $a \in U_i$ for some $i \in J$, which implies that $f(a) \in V_i$ for some $i \in J$, so every element of $f(A)$ lies in V_i for some $i \in J$, so $\{V_i\}_{i\in J}$ is a finite subcover of $f(A)$. Thus $f(A)$ is compact. □

3.2. Heine-Borel in $\mathbb{R}$.

THEOREM 4.10 (Heine-Borel in $\mathbb{R}$). *For $B \subseteq \mathbb{R}$, the following are equivalent:*

(i) *The set B is compact.*
(ii) *The set B is closed and bounded.*

PROOF. (i) $\implies$ (ii) is Proposition 4.8.
(ii) $\implies$ (i): Let B be closed and bounded. Then for some $a \leq b$ we have $B \subseteq [a, b]$. By Proposition 4.9a), if $[a, b]$ is compact then so is $[a, b]$. So it suffices to show that the closed bounded interval $[a, b]$ is compact, which we will do by Real Induction. Let $\{U_i\}_{i\in I}$ be an open cover of $[a, b]$. We define S to be the set of $x \in [a, b]$ such that some finite number of the U_i's cover $[a, x]$. It will suffice to show that S is

inductive, for then by Real Induction we have $S = [a, b]$, and that $b \in S$ means that $[a, b]$ has a finite subcover, which we want to show.
(RI1) Choose $i \in I$ such that $a \in U_i$. Then U_i is a singleton subcover of $[a, a] = \{a\}$, hence certainly a finite subcover.
(RI2) Let $x \in [a, b)$ and suppose that $x \in S$: thus there is a finite subset J of I such that $[a, x] \subseteq \bigcup_{i \in J} U_i$. In particular x lies in the open set $\bigcup_{i \in J} U_i$, so there is some $\delta > 0$ such that $[x, x + \delta]$ also lies in $\bigcup_{i \in J} U_i$, so overall $[a, x + \delta] \subseteq \bigcup_{i \in J} U_i$, which shows that $x + \delta \in S$. Notice that if S contains an element of $[a, b]$ then it also contains all smaller elements, so certainly $[x, x + \delta] \subseteq S$.
(RI3) Let $x \in (a, b]$ and suppose that S contains every element of $[a, b]$ that is smaller than x. Choose $i_\bullet \in I$ such that $x \in U_{i_\bullet}$. Because $U_{i_\bullet}$ is open there is $\delta > 0$ such that $[x - \delta, x] \subseteq U_{i_\bullet}$. Because $x - \delta < x$ we have $x - \delta \in S$, so there is a finite subset J of I such that $[a, x - \delta] \subseteq \bigcup_{i \in J} U_i$. Thus overall we have

$$[a, x] \subseteq \bigcup_{i \in J} U_i \cup U_{i_\bullet},$$

so the covering of $[a, x]$ admits a finite subcover and thus $x \in S$. □

3.3. Heine-Borel in $\mathbb{R}^N$. Recall from Exercise 1.6 that the diameter $\operatorname{diam}(A)$ of a nonempty subset $A \subseteq \mathbb{R}^N$ is the supremum of the distances between pairs of points of A and that $\operatorname{diam}(A)$ is finite if and only if A is bounded. Also we have $\operatorname{diam}(A) = 0$ if and only if A consists of a single point.

THEOREM 4.11 (Cantor Intersection Theorem). *Let $\{B_n\}_{n=1}^\infty$ be a sequence of nonempty closed subsets of $\mathbb{R}^N$ such that:*

(i) *We have $B_1 \supseteq B_2 \supseteq \ldots \supseteq B_n \supseteq \ldots$.*
(ii) *We have $\operatorname{diam}(B_n) \to 0$.*

Then there is $x \in \mathbb{R}^N$ such that $\bigcap_{n=1}^\infty B_n = \{x\}$.

PROOF. Step 1: For $n \in \mathbb{Z}^+$, choose any $\mathbf{x}_n \in B_n$. We claim that the sequence $\{\mathbf{x}_n\}_{n=1}^\infty$ is a Cauchy sequence. Indeed, let $\epsilon > 0$, and choose $N \in \mathbb{Z}^+$ such that for all $n \geq N$ we have $\operatorname{diam}(B_n) < \epsilon$. Then for all $m, n \geq N$ we have that x_m lies in B_m, which is a subset of B_N and $\mathbf{x}_n$ lies in B_n, which is a subset of B_N. So $\mathbf{x}_m$ and $\mathbf{x}_n$ each lie in B_N, hence $||\mathbf{x}_m - \mathbf{x}_n|| < \epsilon$. Cauchy sequences in $\mathbb{R}^N$ are convergent, so there is $\mathbf{x} \in \mathbb{R}^N$ such that $\mathbf{x}_n \to \mathbf{x}$. Let $n \in \mathbb{Z}^+$. Because B_n contains all but finitely many terms of the sequence $\{\mathbf{x}_n\}$ and is closed, it also contains its limit $\mathbf{x}$. So indeed $\mathbf{x} \in \bigcap_{n=1}^\infty B_n$.
Step 2: Seeking a contradiction we suppose there are distinct points $\mathbf{x} \neq \mathbf{y}$ in $\bigcap_{n=1}^\infty$, let $d := ||\mathbf{x} - \mathbf{y}|| > 0$. For any $n \in \mathbb{Z}^+$, since $\mathbf{x}$ and $\mathbf{y}$ both lie in B_n we have $\operatorname{diam}(B_n) \geq d$. A sequence cannot converge to 0 if each of its terms is at least a positive number: contradiction. □

THEOREM 4.12 (Heine-Borel). *For $B \subseteq \mathbb{R}^N$, the following are equivalent:*

(i) *The set B is compact.*
(ii) *The set B is closed and bounded.*

PROOF. (i) $\implies$ (ii) is Proposition 4.8.
(ii) $\implies$ (i): Since B is closed and bounded, there is a closed box A containing B. If A is compact, then by Proposition 4.9 it follows that B is compact. So it suffices to show that closed boxes in $\mathbb{R}^N$ are compact.

Let $\mathcal{B}$ be a closed box in $\mathbb{R}^N$, and seeking a contradiction, suppose there

is an open cover $\{U_i\}_{i \in I}$ with no finite subcover. We may *bisect* $\mathcal{B}$ into 2^N closed subboxes all of the same size: if $\mathcal{B} = \prod_{i=1}^N [a_i, b_i]$, then each of the boxes is obtained by selecting for each $1 \leq i \leq N$ either the left interval $[a_i, \frac{a_i+b_i}{2}]$ or the right interval $[\frac{a_i+b_i}{2}, b_i]$. The diameter of the closed box $\mathcal{B}$ is the length $\sqrt{(a_1-b_1)^2 + \ldots + (a_N-b_N)^2}$ of its longest diagonal, from which it is easy to see that each of the 2^N subboxes has half the diameter of the original box. It must be the case that for at least one of the subboxes $\mathcal{B}_1$ cannot be covered by any finite number of the sets U_i: indeed there are fintiely many subboxes overall and their union is $\mathcal{B}$, so if each subbox could be so covered, then taking the union of those covers would give a finite subcover of $\mathcal{B}$.

Now we repeat the above argument with $\mathcal{B}_1$ in place of $\mathcal{B}$: there must be some subbox $\mathcal{B}_2$ with $\operatorname{diam}(\mathcal{B}_2) = \frac{\operatorname{diam}(\mathcal{B}_1)}{2}$ that is not covered by any finitely many of the U_i's. And so forth: repeating this argument gives us an infinite nested sequence of closed subboxes

$$\mathcal{B} \supsetneq \mathcal{B}_1 \supsetneq \mathcal{B}_2 \supsetneq \ldots$$

such that for all n we have $\operatorname{diam}(\mathcal{B}_{n+1}) = \frac{\operatorname{diam}(\mathcal{B}_n)}{2}$, from which it follows that $\operatorname{diam}(\mathcal{B}_n) \to 0$. The Cantor Intersection Theorem therefore applies: there is a unique point $x \in \bigcap_{n=1}^\infty \mathcal{B}_n$. Since $x \in B$, there is $i \in I$ be such that $x \in U_i$. Because U_i is open, there is some $\delta > 0$ such that U_i contains the closed ball $B^\bullet(x, \delta)$ (by definition this is true for an open ball, but any open ball contains a closed ball of any smaller radius). For any set X containing x and of diameter at most δ, we have $X \subseteq B^\bullet(x, \delta)$. This applies to the box $\mathcal{B}_n$ for all sufficiently large n, but to get a contradiction we only need one, so let $n \in \mathbb{Z}^+$ be such that

$$\mathcal{B}_n \subseteq B^\bullet(x, \delta) \subseteq U_i.$$

By our construction $\mathcal{B}_n$ is not supposed to be covered by any finite number of U_i's....but it is clearly covered by U_i alone: contradiction! $\square$

The Heine-Borel Theorem actually implies the Bolzano-Weierstrass Theorem, as we will now show. By Exercise 1.40 it suffices to show that if A is a subset of $\mathbb{R}^N$ that is infinite and bounded, then it has an accumulation point in $\mathbb{R}^N$, so seeking a contradiction suppose that A has no accumulation point in $\mathbb{R}^N$. Then A is closed, e.g. by Exercise 1.17, so by Heine-Borel the subset A is compact. Let $a \in A$. Since a is not an accumulation point of A, Proposition 1.34 implies that a is an isolated point of a: there is $\delta_a > 0$ such that $A \cap B^\circ(a, \delta_a) = \{a\}$. Then $\{B^\circ(a, \delta_a)\}_{a \in A}$ is an open cover of A that has no proper subcover: indeed for any $a \in A$, by construction the ball $B^\circ(a, \delta_a)$ is the only element of the cover that contains a, so we cannot remove it. Since A is infinite so is our covering, and thus we have found an open cover of A without a finite subcover, contradicting the compactness of A.

It is not as easy to deduce Heine-Borel from Bolzano-Weierstrass, though in some sense that's what we did to prove Heine-Borel, since our proof of Cantor Intersection used the Cauchy completeness of $\mathbb{R}^N$ which we deduced from Bolzano-Weierstrass. However Exercise 4.7 gives a version of the Cantor Intersection Theorem for nested sequences of closed boxes specifically which can be proved directly from the least upper bound axiom.

In many applications of Bolzano-Weierstrass, we could equally well use Heine-Borel.

In particular, the results of this section equally well imply that the continuous image of a closed and bounded subset of Euclidean space is a closed and bounded subset of Euclidean space, which is what is needed to prove e.g. Corollary 1.28. As one goes on in mathematics, the primary of compactness over sequential compactness becomes increasingly pronounced. We give one example in the next section.

3.4. Dini's Theorem. Let X be a set, and let $\{f_n : X \to \mathbb{R}\}_{n=0}^{\infty}$ be a sequence of functions on X. We say that we have an **increasing sequence** if

$$\forall x \in X,\ \forall n \in \mathbb{N}, f_n(x) \leq f_{n+1}(x)$$

and similarly we say that we have a **decreasing sequence** if

$$\forall x \in X, \forall n \in \mathbb{N}, f_n(x) \geq f_{n+1}(x).$$

A good way to think about this is that having an increasing sequence of functions means that for each $x \in X$ we have an increasing sequence $\{f_n(x)\}$ of real numbers, while having a decreasing sequence of functions means that for each $x \in X$ we have a decreasing sequence $\{f_n(x)\}$ of real numbers.

We hasten to add a possible point of confusion: when X is a subset of $\mathbb{R}$ we have the notion of an increasing function $f : X \to \mathbb{R}$: this is a function such that $x_1 \leq x_2 \implies f(x_1) \leq f(x_2)$. We could of course have a sequence of such functions, which would give us a *sequence of increasing functions*. Thus

$$\text{sequence of increasing functions} \neq \text{ increasing sequence of functions},$$

so one must listen carefully. It may help to note that the notion of an increasing sequence of functions makes sense when the domain is any set X, whereas to define an increasing function $f : X \to \mathbb{R}$ we would need an order relation on X.

Now we present a result of Dini that gives a new criterion for uniform convergence.

THEOREM 4.13 (Dini's Theorem). *Let $A \subseteq \mathbb{R}^N$ be compact, and let $\{f_n : A \to \mathbb{R}\}_{n=0}^{\infty}$ be a sequence of real-valued functions on A. Suppose all of the following:*

(i) *Each f_n is continuous.*
(ii) *The sequence $\{f_n\}$ is increasing or decreasing.*
(iii) *$\{f_n\}$ converges pointwise on A to a* ***continuous*** *function A.*

Then $f_n \overset{u}{\to} f$ on A.

PROOF. Step 1: Replacing $\{f_n\}$ by $\{-f_n\}$ if necessary, we may assume that the sequence is decreasing. Siince f is continuous, the sequence $\{f_n\}$ is continuous, decreasing and converges pointwise to f on A if and only if the sequence $\{f_n - f\}$ is continuous, decreasing and converges pointwise to 0 on A, so we may assume that $f = 0$. The hypotheses then imply that $f_n(x) \geq 0$ for all $n \in \mathbb{N}$ and all $x \in A$.
Step 2: Let $\epsilon > 0$ and let $x \in A$. Since $f_n(x) \to 0$, there is $N_x \in \mathbb{N}$ such that $f_{N_x}(x) \in [0, \frac{\epsilon}{2})$. Since f_{N_x} is continuous at x, there is $\delta_x > 0$ such that

$$\forall y \in B^\circ(x, \delta_x) \cap A,\ |f_{N_x}(y) - f_{N_x}(x)| < \frac{\epsilon}{2},$$

from which it follows that

$$\forall y \in B^\circ(x, \delta_x) \cap A,\ f_{N_x}(y) = |f_{N_x}(y)| \leq |f_{N_x}(y) - f_{N_x}(x)| + |f_{N_x}(x)| < \epsilon.$$

Since the sequence is decreasing, we get that $|f_n(y)| < \epsilon$ for all $n \geq N_x$ and all $y \in B^\circ(x, \delta_x) \cap A$.

Step 3: The open balls $\{B^\circ(x, \delta_x)\}_{x \in A}$ form an open cover of A. Since A is compact, there is a finite subcover: thus there are $x_1, \ldots, x_n \in A$ such that $A \subseteq \bigcup_{i=1}^n B^\circ(x, \delta_x)$. Put

$$N := \max(N_{x_1}, \ldots, N_{x_n}).$$

Then for all $x \in A$ and all $n \geq N$, there is some $1 \leq i \leq n$ such that $x \in B^\circ(x_i, \delta_{x_i})$ so $|f_n(x)| < \epsilon$. This shows that the convergence is uniform on A. □

The proof of Dini's Theorem is a good illustration of the power of compactness: it allowed us to pass from a condition that held "locally at every point of A" to a global condition on A almost immediately. So far as I know, it is not possible to use Bolzano-Weierstrass in place of Heine-Borel in the proof of Dini's Theorem.

In the setup of Dini's Theorem, suppose that we assume conditions (i) and (ii) only. Then condition (iii) is actually equivalent to the convergence to f being uniform. Indeed, one direction is Dini's Theorem, while the other direction is Corollary 3.3. We state this again in the special case of a sequence of partial sums:

Corollary 4.14. *Let $A \subseteq \mathbb{R}^N$ be compact, and let $\{f_n : A \to [0, \infty)\}_{n=0}^\infty$ be a sequence of non-negative continuous real functions defined on A. Suppose that the series $\sum_{n=0}^\infty f_n$ converges pointwise on A to a function $S : A \to \mathbb{R}$. Then the convergence is uniform if and only if the sum S is a continuous function.*

Proof. For $N \in \mathbb{N}$, put $S_N := \sum_{n=0}^N f_n$. Since each f_n is non-negative, the sequence $\{S_N\}$ of partial sums is increasing, and thus hypotheses (i) and (ii) of Dini's Theorem apply to $\{S_N\}$. So: if S is continuous, then Dini's Theorem applies and the convergence is uniform. Conversely, if the convergence is uniform then S is continuous by Corollary 3.3. □

Exercises.

Exercise 4.6. *For $n \in \mathbb{N}$, let $f_n : [0, 1] \to [0, \infty)$ by $f_n(x) = x^n(1 - x)$. Use Corollary 4.14 to show: $\sum_{n=0}^\infty f_n$ converges pointwise but not uniformly on $[0, 1]$.*

Problems.

Exercise 4.7. *To prove Heine-Borel we applied the Cantor Intersection in the case where each B_n was a closed box in $\mathbb{R}^N$. In this case one can give a more elementary proof and indeed see directly that the hypothesis that the diameters converge to 0 is not needed.*

a) *Let $\{B_n\}_{n=1}^\infty$ be a nested sequence of nonempty closed bounded intervals in $\mathbb{R}$, so $B_n = [a_n, b_n]$ for some real numbers $a_n \leq b_n$. Put*

$$A := \sup_n a_n \text{ and } B := \inf_n b_n.$$

Using only the Dedekind completeness of $\mathbb{R}$, show:

$$A \leq B \text{ and } \bigcap_{n=1}^\infty B_n = [A, B].$$

b) *Let $\{B_n\}_{n=1}^\infty$ be a nested sequence of closed boxes in $\mathbb{R}^N$. Show: $\bigcap_{n=1}^\infty B_n$ is itself a closed box. (Note that a point counts as a closed box.)*

CHAPTER 5

Metric Spaces

1. A look ahead

In the last two chapters we sketched out some of the terrain of more advanced analysis, in which function theory and set theory interact in more subtle ways. In truth, graduate level real analysis is a very challenging course that relatively few students will take. So I want to end by previewing a different course: undergraduate general topology. This course gives a generalization and abstraction of most of the material from Chapter 1, to the context of metric spaces.

Let X be a set. A **metric function** is a function $d : X \times X \to \mathbb{R}$ such that all of the following hold:

(D1) (Positive Definiteness) For all $x, y \in X$, we have $d(x, y) \geq 0$, with equality if and only if $x = y$.
(D2) (Symmetry) For all $x, y \in X$, we have $d(x, y) = d(y, x)$.
(D3) (Triangle Inequality) For all $x, y, z \in X$, we have $d(x, z) \leq d(x, y) + d(y, z)$.

A **metric space** is a pair (X, d), where X is a set and $d : X \times X \to \mathbb{R}$ is a metric function. In our course, the shining example was to take $X = \mathbb{R}^N$ and d to be the Euclidean distance function: $d(\mathbf{x}, \mathbf{y}) := ||\mathbf{x} - \mathbf{y}||$. Many other examples come from this, since if (X, d) is a metric space and Y is any subset of X, then if $d_Y : Y \times Y \to \mathbb{R}$ is just the metric function restricted to $Y \times Y$, then d_Y is a metric function on Y, so (Y, d_Y) is again a metric space.

In Chapters 1 and 4 we studied:

- Convergence of sequences in $\mathbb{R}^N$.
- Continuity of functions $f : X \to \mathbb{R}^M$ where X is a subset of $\mathbb{R}^N$.
- Bounded, open and closed sets in $\mathbb{R}^N$.
- Sequential compactness of subsets of $\mathbb{R}^N$.
- Compactness of subsets of $\mathbb{R}^N$.

These concepts translate essentially verbatim to the context of a general metric space (X, d), and there is a useful general theory that parallels *much* of what we did in Euclidean spaces. However, in several ways, $\mathbb{R}^N$ and various subsets of it (especially, closed and bounded subsets) behave more nicely than an arbitrary metric space. When this occurs, it is important to think deeply about why: usually one can isolate a certain specific feature of $\mathbb{R}^N$ and use it to define classes of metric spaces in which these good things continue to happen.

Let us now give just a flavor of this.

We said that a sequence $\{x_n\}$ in $\mathbb{R}^N$ converges to a point p in $\mathbb{R}^N$ if the real sequence $d(x_n, p)$ converges to 0. This definition makes sense in any metric space, and the geometric intuitition is the same: all sufficiently large terms of the sequence should lie arbitrarily close to the limit. (It does *not* make sense to talk about divergence to ∞ without some extra structure.) In any metric space, a sequence can have at most one limit, and if a sequence converges to p than all of its subsequences converge to p. And again, it is interesting to explore to what extent we can get a divergent sequence to converge by passing to subsequences.

We can define open and closed balls in any metric space (X, d) in exactly the same way, for $x \in X$ and $\epsilon > 0$, we put

$$B^\circ(x, \epsilon) := \{y \in X \mid d(x, y) < \epsilon\} \text{ and } B^\bullet(x, \epsilon) := \{y \in X \mid d(x, y) \leq \epsilon\}.$$

They may not *look* like balls anymore – we will see an interesting example of this later – but if you think about it, the finer geometry of balls was never really used.[1]

Again we can define a subset U of a metric space (X, d) to be **open** if for every $x \in U$ there is $\epsilon > 0$ such that $B^\circ(x, \epsilon) \subseteq U$. Moreover we can define limit points of a subset Y in the same way: these are the limits of convergent sequences whose terms lie in Y. (We can also define accumulation points.) Then we can say that a subset Y is **closed** if it contains all of its limit points. Again it turns out that $Y \subseteq X$ is closed if and only if its complement $X \setminus Y$ is open...and the proof is really the same. We can also define boundedness: a subset $Y \subseteq X$ is bounded if it lies in some closed ball $B^\bullet(x, R)$. Equivalently, for a nonempty subset Y of a metric space X we can define its **diameter**

$$\operatorname{diam}(Y) := \sup\{\ d(y_1, y_2) \mid y_1, y_2 \in Y\} \in [0, \infty]$$

and put $\operatorname{diam} \varnothing = 0$; then a subset Y is bounded if and only if it has finite diameter.

If (X, d_X) and (Y, d_Y) are two metric spaces and $f : X \to Y$ is a function between them, then all of the following definitions go through using the metric functions instead of Euclidean norms: continuous, uniformly continuous, Lipschitz. Just to spell out the first one: we say that $f : X \to Y$ is **continuous** at $c \in X$ if for all $\epsilon > 0$, there is $\delta > 0$ such that for all $x \in X$, if $d(x, c) < \delta$ then $d(f(x), f(c)) < \epsilon$. Once again continuous functions are characterized by preservation of limits of convergent sequences, we have a sequential characterization of uniform continuity, and so forth.

For a subset Y of a metric space (X, d), we say that Y is **sequentially compact** if every sequence $\{x_n\}$ in Y admits a subsequence converging to an element of Y. This is the same definition as before. A subset Y of a metric space (X, d) is **compact** if for every family $\{U_i\}_{i \in I}$ of open subsets of X that covers Y in the sense that $Y \subseteq \bigcup_{i \in I} U_i$, there is a finite subset $J \subseteq I$ such that $\bigcup_{i \in J} U_i$ also covers Y. In brief: "every open cover of Y has a finite subcover."

The proofs that we gave in $\mathbb{R}^N$ readily adapt to show: sequentially compact subsetse must be closed and bounded, and the same holds for compact subsets. But

[1]Perhaps the closest we came to this was showing that balls are convex. Convexity does *not* make sense in an arbitrary metric space. One needs the structure of a real vector space for this.

now a surprise occurs: for a subset Y of a general metric space X, being closed and bounded is not sufficient for either sequential compactness or compactness.

EXAMPLE 5.1. *For a set X, the* ***discrete metric*** d_d *on X is*

$$d(x,y) := \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases}.$$

For any $c \in X$ and any $0 < \epsilon < 1$, we have $B^\circ(c,\epsilon) = \{x\}$. Then (X, d_d) is bounded: if $X \neq \varnothing$, then $\operatorname{diam}(X) = 1$. *Also X is closed as a subset of itself.*

But: if a sequence $\{x_n\}$ in a metric space converges to a point c, then for all $\epsilon > 0$, we must have $x_n \in B^\circ(c,\epsilon)$ for all sufficiently large n. So a sequence in the discrete metric space (X, d_X) converges to c if and only if all sufficiently large terms are equal to c, i.e., is eventually constant. If X is moreover infinite, then there is an injective sequence $x_\bullet : \mathbb{Z}^+ \to X$. Every subsequence of $\{x_n\}$ remains injective and therefore divergent. Therefore X itself is closed, bounded but not sequentially compact. Neither is X compact, because the singleton cover $\{\{x\}\}_{x \in X}$ is an open cover and does not have any proper subcover, hence – since X is infinite – does not have any finite subcover.

If one looks back at the proof of the Bolzano-Weierstrass Theorem in $\mathbb{R}^N$, we get referred back to Bolzano-Weierstrass in $\mathbb{R}$ which was proved in a previous course using the completeness properties of $\mathbb{R}$. In a metric space X we do *not* have a notion of ordering of the points, so upper bounds and Dedekind completeness doesn't make sense. However, Cauchy sequences do: a sequence $\{x_n\}$ in a metric space (X, d) is **Cauchy** if for all $\epsilon > 0$ there is $N \in \mathbb{Z}^+$ such that for all $m, n \geq N$ we have $d(x_m, x_n) < \epsilon$. Again it is easy to see that convergent sequences are Cauchy but the converse does *not* generally hold: we say that a metric space is **complete** if every Cauchy sequence in that space is convergent.

For instance, there are Cauchy sequences in $\mathbb{Q}$ (which becomes a metric space by restricting the metric function on $\mathbb{R}$) that converge only to elements of $\mathbb{R}$ – e.g. a sequence of rational approximations to $\sqrt{2}$ – so $\mathbb{Q}$ is not a complete metric space. Actually, a little thought shows that this phenomenon is much more general: if (X, d) is any metric space whatsoever and Y is a subset of X that is *not* closed, then by definition there is a sequence $\{y_n\}$ in Y converging to an element $x \in X \setminus Y$; any subsequence still converges to x and therefore not to any element of Y. So incomplete metric spaces abound.

One more definition: a subset Y of a metric space X is **totally bounded** if for every $\epsilon > 0$, it admits a finite cover by subsets of diameter at most ϵ. Equivalently, for every $\epsilon > 0$, Y admits a finite cover by closed ϵ-balls. Since sets of finite diameter are bounded and finite unions of bounded sets are bounded, certainly totally bounded implies bounded. The terminology is of course suggesting that totally bounded could be stronger. This is an absolutely key example of a difference between $\mathbb{R}^N$ and a general metric space: in $\mathbb{R}^N$ every bounded set is totally bounded (this was a homework problem). However, an infinite set with the discrete metric is bounded but not totally bounded: the only sets of diameter at most $\frac{1}{2}$ are single points, an an infinite set is not a finite union of singleton subsets!

Now we can state what is probably the most important theorem of metric topology:

THEOREM 5.1. *For a metric space (X, d), the following are equivalent:*

(i) *X is compact: every open cover of X has a finite subcover.*
(ii) *X is sequentially compact: every sequence in X has a convergent subsequence.*
(iii) *X is "accumulation point compact": every infinite subset of X has an accumulation point in X.*
(iv) *X is complete (Cauchy sequences converge) and totally bounded (for all $\epsilon > 0$, X can be covered by finitely many subsets of arbitrarily small diameter).*

PROOF. See e.g. [**GT**, Thm. 2.78]. (Warning: in those notes, where we say "limit point" they say "adherent point" and where we say "accumulation point" they say "limit point.") □

Theorem 5.1 is not quite a generalization of either the Bolzano-Weierstrass or Heine-Borel Theorems, but it helps us to understand them more deeply: it shows that the two key facts that go into these results are the Cauchy completeness of $\mathbb{R}$ (from which the Cauchy completeness of $\mathbb{R}^N$ follows almost immediately) and the fact that bounded subsets in Euclidean space are totally bounded, which can be thought of as a consequence of the Archmedean property in the form that if you divide a real number by 2 enough times, it gets arbitrarily small.

If we have two metric spaces (X, d_X) and (Y, d_Y), a map $f : X \to Y$ is an **isometric embedding** if it preserves distances between points:

$$\forall x_1, x_2 \in X,\ d_Y(f(x_1), f(x_2)) = d_X(x_1, x_2).$$

Such maps are in particular Lipschitz with Lipschitz constant 1, so they are uniformly continuous, and so forth. But really this is much stronger: Lipschitz maps are maps that only stretch distances between points by a bounded factor, while isometric embeddings *preserve* distances. You should think of an isometric embedding $f : X \to Y$ as giving you a "perfect copy" $f(X)$ of X as a subset of Y.

Now here is another big theorem:

THEOREM 5.2. *Let (X, d_X) be a metric space. Then there is a metric space $(\tilde{X}, d_{\tilde{X}})$ and an isometric embedding*

$$\iota : X \to \tilde{X}$$

such that:

(i) *$\tilde{X}$ is a complete metric space, and*
(ii) *The image $\iota(X)$ is dense in $\tilde{X}$: that is, for every $x \in \tilde{X}$ there is a sequence $\{x_n\}$ in X such that $\iota(x_n) \to x$.*

The metric space $\tilde{X}$ is called the **completion** of the metric space X, and it can be thought of as "filling in the missing holes" that prevent Cauchy sequences in X from converging. Moreover, the completion $\tilde{X}$ is essentially *unique*, although I don't have the time to explain exactly what that means here. This is such a profound idea: you have a space in which not every Cauchy sequence converges, which robs you of and essential tool to show convergence of sequences. So you faithfully embed your space inside a larger space (in a parsimonious way: every point you

have added is the limit of a sequence in your original space) and in that larger space all Cauchy sequences converge.

One thing that the formalism of metric spaces buys you is the idea to consider two different metric functions *on the same space*. This turns out to be very natural and useful, because indeed there is often more than one sense in which things can get "close together" and you want to compare the two. Let me end by mentioning an example of this: let

$$X = \mathcal{C}[a,b] = \{\text{continuous } f : [a,b] \to \mathbb{R}\}$$

be the set of continuous real-valued functions defined on $[a,b]$. We want to make this into a metric space, i.e., we want to measure the distance between two functions. How might we do this?

One way to do this was given in Math 3100: for $f, g \in \mathcal{C}[a,b]$, put

$$d_\infty(f,g) := \max_{x\in[a,b]} |f(x) - g(x)|.$$

(It is easy to see that this is a metric: the triangle inequality follows from the usual triangle equality in $\mathbb{R}$.) Convergence of sequences in the d_∞-metric is precisely uniform convergence. However, there is another metric that is arguably even more natural: for $f, g \in \mathcal{C}[a,b]$, put

$$d_1(f,g) := \int_a^b |f - g|.$$

I claim that d_1 is a metric function. This time the Triangle Inequality is not the hardest part: for $f, g, h \in \mathcal{C}[a,b]$ we have

$$\begin{aligned} d_1(f,h) &= \int_a^b |f-h| \leq \int_a^b (|f-g| + |g-h|) \\ &= \int_a^b |f-g| + \int_a^b |g-h| = d_1(f,g) + d_1(g,h). \end{aligned}$$

Because $|f-g| = |g-f|$, clearly $d_1(f,g) = d_1(g,f)$. Also clearly $f_1(f,g) \geq 0$, because the integral of a non-negative function is non-negative. Also clearly $d_1(f,f) = 0$. However, it takes some work to show that if if $d_1(f,g) = 0$ then $f = g$: this comes down to showing: if $f : [a,b] \to \mathbb{R}$ is continuous and non-negative, then $\int_a^b f = 0$ implies $f = 0$. (This is a good exercise! I recommend you try it.)

We have moreover that

$$d_1(f,g) = \int_a^b |f-g| \leq \int_a^b d_\infty(f,g) = (b-a)d_\infty(f,g),$$

so that the d_1-metric is, up to the constant $(b-a)$, the smaller of the two, and intuitively it measures the distance in a more refined way: whereas $d_\infty(f,g)$ measures the *maximum* distance between $f(x)$ and $g(x)$, $\frac{1}{b-a}d_1(f,g)$ measures the *average* distance between $f(x)$ and $g(x)$. In more advanced analysis *both* of these metrics are important, and they fit into an infinite family of metrics d_p for $p \in [1,\infty]$.

It turns out that $\mathcal{C}[a,b]$ with the d_∞-metric is complete: this is a variant of the Math 3100 fact that a uniform limit of continuous functions remains continuous. On the other hand, $\mathcal{C}[a,b]$ with the d_1-metric is *not* complete.

EXAMPLE 5.2. *For $n \in \mathbb{Z}^+$, let $f_n : [0,2]$ be the function*

$$f_n(x) = \begin{cases} x^n & \text{if } x \in [0,1) \\ 1 & \text{if } x \in [1,2] \end{cases}.$$

This sequence converges pointwise the to the function:

$$f : [0,2] \to \mathbb{R},\ f(x) = \begin{cases} 0 & \text{if } x \in [0,1) \\ 1 & \text{if } x \in [1,2] \end{cases},$$

which is discontinuous at 1. Since f is bounded with a single discontinuity, it is Riemann integrable, and

$$d_1(f_n, f) = \int_0^2 |f_n - f| = \int_0^1 |x^n| + \int_1^2 0 = \int_0^1 x^n = \frac{1}{n+1} \to 0.$$

This implies that the sequence $\{f_n\}$ is Cauchy in $\mathcal{C}[a,b]$ with the d_1-metric: indeed, for all $m, n \in \mathbb{Z}^+$, we have

$$d_1(f_m, f_n) \leq d_1(f_m, f) + d_1(f_n, f),$$

so if we choose $N \in \mathbb{Z}^+$ such that for all $n \geq N$ we have $d_1(f_n, f) < \frac{\epsilon}{2}$ then for all $m, n \geq N$ we have $d_1(f_m, f_n) < \epsilon$. If there were a continuous *function g such that $f_n \to g$ in the d_1-metric, then for all $n \in \mathbb{Z}^+$ we have*

$$d_1(f, g) \leq d_1(f, f_n) + d_2(f_n, g),$$

so

$$d_1(f, g) \leq \lim_{n \to \infty} d_1(f, f_n) + d_2(f_n, g) = 0.$$

Since f and g are both continuous on $[1,2]$, we must have $g(x) = f(x) = 1$ for all $x \in [1,2]$. Similarly, for any $\delta \in (0,1)$, we have

$$0 = \int_0^2 |f - g| = \int_0^\delta |f - g| + \int_\delta^2 |f - g|;$$

since both terms are non-negative, we conclude $\int_0^\delta |f - g| = 0$. Since f and g are continuous on $[0, \delta]$, we must have $g = f = 0$. Therefore we know that g is continuous on $[0,2]$, is equal to 0 for all $x \in [0,1)$ and is equal to 1 for all $x \in [1,2]$. But there is no such function, so $\{f_n\}$ is Cauchy in $(\mathcal{C}[a,b], d_1)$ but not convergent.

But all is not lost! By Theorem 5.2 one can consider the completion of $\mathcal{C}[a,b]$ with respect to the d_1-metric. This is called the **Lebesgue space** $L^1([a,b])$...and now we are back to real analysis. Indeed, these Lebesgue spaces are discussed in Math 8100 perhaps more than any other topic.

Bibliography

[Ab] S. Abbott, *Understanding analysis.* Second edition. Undergraduate Texts in Mathematics. Springer, New York, 2015.

[AM] P.L. Clark, *Introduction to Advanced Mathematics.* `http://alpha.math.uga.edu/~pete/3200_supplemental.pdf`

[Be12] S.J. Bernstein, *Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités.* Communications of the Kharkov Mathematical Society 13 (1912), 1–2.

[Be14] Á. Besenyei, *Peano's unnoticed proof of Borel's theorem.* Amer. Math. Monthly 121 (2014), 69–72.

[Be67] L. Bers, *Classroom Notes: On Avoiding the Mean Value Theorem.* Amer. Math. Monthly 74 (1967), 583.

[Bo95] É. Borel, *Sur quelque points de la théorie des fonctions.* Ann. Sci. lÉcole Norm. Sup. 12 (1895), 9–55.

[Bo81] R.P. Boas, *Who Needs Those Mean-Value Theorems, Anyway?* The Two-Year College Mathematics Journal 12 (1981), 178–181.

[Ca21] A.L. Cauchy, *Analyse algébrique*, 1821.

[Cl19] P.L. Clark, *The instructor's guide to real induction.* Math. Mag. 92 (2019), 136–150.

[Cl22] P.L. Clark, *A Mean Value Inequality for all functions.* `http://alpha.math.uga.edu/~pete/MVT_Dini.pdf`

[Co67] L.W. Cohen, *Classroom Notes: On Being Mean to the Mean Value Theorem.* Amer. Math. Monthly 74 (1967), 581–582.

[D] J. Dieudonné, *Foundations of Modern Analysis* 2nd edition, Academic Press, New York, 1969.

[dBR76] P. du Bois-Reymond, *Über den Gültigkeitschbereich der Taylor'schen Reihenentwickelung.* Sitzungb. k. Bayer. Akad. Wiss., math.-phys. Klasse (1876), 225–237.

[dBR83] P. du Bois-Reymond, *Über den Gültigkeitschbereich der Taylor'schen Reihenentwickelung.* Math. Ann. 21 (1883), 107–119.

[Go16] R.A. Gordon, *A bounded derivative that is not Riemann integrable.* Math. Mag. 89 (2016), 364–370.

[GT] P.L. Clark, *General Topology.* `http://alpha.math.uga.edu/~pete/pointset.pdf`

[Ha88] J. Hadamard, *Sur le rayon de convergence des śeries ordonnées suivant les puissances d'une variable.* C.R. Acad. Sci. Paris 106 (1888), 259–262.

[Ha11] D. Hathaway, *Using Continuity Induction.* College Math. Journal 42 (2011), 229–231.

[HC] P.L. Clark, *Honors Calculus* `http://alpha.math.uga.edu/~pete/2400full.pdf`

[HS33] T.H. Hildebrant and I.J. Schoenberg, *On linear functional operations and the moment problem for a finite interval in one or several dimensions.* Ann. of Math. (2) 34 (1933), 317–328.

[No52] M.J. Norris, *Integrability of continuous functions.* Amer. Math. Monthly 59 (1952), 244–245.

[R] W. Rudin, *Principles of mathematical analysis.* Third edition. International Series in Pure and Applied Mathematics. McGraw-Hill, New York-Auckland-Düsseldorf, 1976.

[SS] M.R. Adams and P.L. Clark, *Sequences and series.*

[St37] M.H. Stone, *Applications of the theory of Boolean rings to general topology.* Trans. Amer. Math. Soc. 41 (1937), 375–481.

[St48] M.H. Stone, *The generalized Weierstrass approximation theorem.* Math. Mag. 21 (1948), 167–184.