1 The Real Numbers

In this course, we will primarily be studying properties of the real numbers. (In fact, we'll study a generalization called "metric spaces", which we'll define soon. But we can't talk about metric spaces without first understanding the real numbers well).

The real numbers are defined entirely by the fact that they are a "complete ordered field". They are in fact only complete ordered field that exists. But that probably doesn't tell you very much right now, because you probably don't know what any of those three things mean. So our first goal is to understand those three terms, and thus effectively define the real numbers.

1.1 Fields

First we need to define what a field is. The basic idea is that a field is a set where you can do addition, subtraction, multiplication, and division. More formally:

Definition 1.1. Suppose F is a set with two binary operations, $+$ and $\times$. We say F is a *field* if it satisfies the following axioms:

1. (Closure) If $x, y \in F$ then $x + y, xy \in F$.
2. (Commutativity) $x + y = y + x$ and $xy = yx$ for all $x, y \in F$.
3. (Associativity) $(x + y) + z = x + (y + z)$ and $(xy)z = x(yz)$ for all $x, y, z \in F$.
4. (Identities) There is an element $0 \in F$ such that $x + 0 = x$ for all $x \in F$. There is an element $1 \in F$ such that $1x = x$ for all $x \in F$.
5. (Inverses) For every $x \in F$ there is a $-x \in F$ such that $x + (-x) = 0$. For every non-zero $x \in F$ there is an element $x^{-1} \in F$ such that $xx^{-1} = 1$.
6. (Distributivity) $x(y + z) = xy + xz$ for all $x, y, z \in F$.

Example 1.2. The set $\mathbb{Q}$ of rational numbers is a field. The sets $\mathbb{R}$ and $\mathbb{C}$ of real and complex numbers are fields.

The set $\mathbb{Z}$ of integers is not a field, because it does not have multiplicative inverses. (We call this set a *ring*).

The set $\mathbb{N}$ of natural numbers is not a field. It does not have multiplicative or additive inverses.

The set $\mathbb{Z}/n\mathbb{Z}$ of integers modulo n is is a field if n is prime, and is not a field if n is composite. (It is a fact we won't discuss again in this course that every number has a multiplicative inverse modulo n if and only if n is prime).

Proposition 1.3. *If $a, b \in F$ then there is a unique solution to the equation $x + a = b$. That is, there exists a unique $x \in F$ such that $x + a = b$.*

Proof. First we prove uniqueness. That is, we will suppose that x is a solution, and prove that there is only one possible value of x.

By inverses, we know that $-a$ exists. Since $x + a = b$ we have $(x + a) + (-a) = b + (-a)$.

By associativity, $(x + a) + (-a) = x + (a + (-a))$, and by inverses, $a + (-a) = 0$ so $x + (a + (-a)) = x + 0 = x$ by identity. Thus we have $x = b + (-a)$. So the only possible solution is $x = b + (-a)$.

Now we prove existence—which in this case, means proving that $b + (-a)$ is a solution. But we have

$$\begin{aligned}
(b + (-a)) + a &= b + (-a + a) && \text{Associativity} \\
&= b + (a + (-a)) && \text{Commutativity} \\
&= b + 0 && \text{Inverses} \\
&= b && \text{Identity}
\end{aligned}$$

and thus $b + (-a)$ is a solution.

□

Proposition 1.4. *Let F be a field and let $x_1, \dots, x_n \in F$. The meaning of the expression $x_1 + x_2 + \cdots + x_n$ does not depend on the location of the parentheses, and thus we may omit parentheses without ambiguity.*

Proof. We prove this by induction. For a base case, suppose $n = 3$. Then by associativity, we know that $(x_1 + x_2) + x_3 = x_1 + (x_2 + x_3)$, and these are the only possible uses of parentheses, so the claim is true.

Now suppose the claim is true for a sum of n or fewer terms. Then we can rewrite any such sum to have the parentheses all on the left, as $(((x_1 + x_2) + x_3) + \dots) + x_n$. Now consider the sum $x_1 + \cdots + x_{n+1}$. Any parenthesization will split it up as $(x_1 + \cdots + x_k) + (x_{k+1} + \cdots + x_{n+1})$ for some k.

By our inductive hypothesis, since the right-hand term has fewer than n terms, we can reparenthesize it with all the parentheses on the left to get $(x_1 + \cdots + x_k) + (((x_{k+1} + \dots) +$

$x_n)+x_{n+1})$. By associativity, this is the same as $((x_1+\cdots+x_k)+((x_{k+1}+\dots))+x_n)+x_{n+1}$, and now the left-hand term has n terms, so by our inductive hypothesis we can put all the parentheses on the left, to get $(((x_1+x_2)+\dots)+x_n)+x_{n+1}$.

Thus every sum of n_1 terms, however parenthesized, can be rewritten with all parentheses to the left without changing the value. This proves our original claim that the value does not depend on the location of the parentheses. □

Exercise 1.5. *Prove that if F is a field and $x \in F$, then $0x = 0$.*

1.2 Ordered Fields

We see that $\mathbb{R}$ is one of many fields, so we need to be more specific in our description of it. One of the most important features of the real numbers is the real number line—which is just a way of saying we can put all the real numbers in order.

Definition 1.6. An *ordered set* is a set S with a total binary transitive anti-symmetric relation $\leq$. That is:

1. (Trichotomy) For any x, y in S, exactly one of the following is true: $x < y$ or $x = y$ or $y < x$; and
2. (Transitivity) If $x < y$ and $y < z$ then $x < z$.

We write $x \leq y$ if either $x < y$ or $x = y$. We define $\geq$ and $\geq$ in the obvious way: $x > y$ if and only if $y < x$.

It's possible to put an order on many sets just by arbitrarily deciding which things are smaller than which other things. (It's reasonable to say Red < Orange < Yellow < Green < Blue < Violet, for instance). The interesting question is whether you can order a set in a way compatible with its other properties.

Definition 1.7. A field F is an *ordered field* if it is an ordered set such that

1. (Order Additivity) If $x, y, z \in F$ and $x < y$ then $x + z < y + z$.
2. (Order Multiplicativity) If $x, y \in F$ with $x > 0, y > 0$, then $xy > 0$.

Remark 1.8. Rosenlicht gives an equivalent but distinct characterization, where he separates $\mathbb{R}$ into the disjoint sets $\mathbb{R}_+, \{0\}, \mathbb{R}_-$, and asserts that $\mathbb{R}_+$ is closed under addition and multiplication; he then defines $x > y$ to mean that $x - y \in \mathbb{R}_+$.

This is sufficient to imply that $\mathbb{R}$ is an ordered field under our definition; in fact, the properties we have stated are O1 through O4 in Rosenlicht. Rosenlicht's characterization has some benefits for the purpose of universal axiomatic constructions, but I think is a bit less clear about what's actually going on. My presentation thus follows Lebl instead.

We only stated a couple of principles here, but they imply a lot more. Most of the things they imply are things you're already taking for granted, but it's worth spelling them out—both because we need to know them, and because proving them is good practice for the sort of arguments we'll be making for the rest of the course.

Proposition 1.9. *1. $x > 0$ if and only if $-x < 0$.*

2. If $x > 0$ and $y < z$ then $xy < xz$.

3. If $x < 0$ and $y < z$ then $xy > xz$.

4. If $x \neq 0$ then $x^2 > 0$.

5. $1 > 0$ in any field.

6. If $0 < x < y$ then $0 < y^{-1} < x^{-1}$.

7. If $0 < x < y$ then $x^2 < y^2$.

8. If $x \leq y$ and $z \leq w$ then $x + z \leq y + w$.

Proof. 1. $x > 0$, so by order additivity $x + (-x) > 0 + (-x)$, and thus $0 > -x$.

2. If $y < z$ then $y + (-y) < z + (-y)$ by order additivity, so $0 < z - y$. Then since $x > 0$, order multiplicativity gives us $0 < x(z - y)$, and distributivity gives $0 < xz - xy$. Finally, order additivity gives us $0 + xy < xz - xy + xy$, and thus $xy < xz$.

3. Exercise

4. If $x > 0$ then by order positivity, $x \cdot x > 0$.

   If $x < 0$ then by the previous result, we have $x \cdot x > 0 \cdot 0 - 0$.

5. $1 = 1^2$, and $1^2 > 0$ by the previous result.

6. We know by field axioms that $\frac{1}{x} \neq 0, \frac{1}{y} \neq 0$. We first want to prove that both of them are positive. Suppose $\frac{1}{x} < 0$. Then $-\frac{1}{x} > 0$, and since $x > 0$, by order multiplicativity we have $-\frac{1}{x} \cdot x > 0$. Thus $-1 > 0$, which is a contradiction.

A similar argument shows that $\frac{1}{y} > 0$. So now we just need to show that $\frac{1}{y} > \frac{1}{x}$. But we know that $\frac{1}{x}\frac{1}{y} > 0$ by order multiplicativity; and since $x < y$, order multiplicativity tells us that $\frac{1}{x}\frac{1}{y}x < \frac{1}{x}\frac{1}{y}y$. Then muliplicative identities tells us that $\frac{1}{y} < \frac{1}{x}$.

7. Exercise

8. Exercise

□

Proposition 1.10. *Let $x, y \in F$ where F is an ordered field. Then $xy > 0$ if and only if either $x, y > 0$ or $x, y < 0$. $xy < 0$ if and only if either $x < 0, y > 0$ or $x > 0, y < 0$.*

Proof. If either $x = 0$ or $y = 0$, then $xy = 0$. So let's assume both x and y are nonzero.

If $x, y > 0$ then $xy > 0$ by definition of ordered field. If $x, y < 0$ then by proposition 1.9 we have $xy > x0 = 0$.

Now suppose x and y have opposite signs. Without loss of generality, assume $x > 0, y < 0$. Then by proposition 1.9 we have $xy < x0 = 0$. □

Example 1.11. $\mathbb{R}$ is an ordered field, as is $\mathbb{Q}$.

The set $\mathbb{Z}/p\mathbb{Z}$ of integers modulo p is a field, but cannot be made into an ordered field. For suppose it were an ordered field. We will see that any square in an ordered field must be positive; since $1 = 1 \cdot 1$, we know that $1 > 0$. Then by positive additivty, we have $2 = 1 + 1 > 0 + 0 = 0, 3 = 1 + 1 + 1 > 0 + 0 + 0 = 0, \ldots$. Adding p copies of 1 gives us 0, and thus we have $0 > 0$, which is a violation of the trichotomy principle.

Exercise 1.12. *Prove that $\mathbb{C}$ cannot be an ordered field. (Hint: is $i > 0$, $i = 0$, or $i < 0$?)*

Definition 1.13. The *absolute value function* is defined by the formula

$$|a| = \begin{cases} a & a > 0 \\ 0 & a = 0 \\ -a & a < 0 \end{cases}$$

Proposition 1.14. *1. $|a| \geq 0$ for all $a \in \mathbb{R}$, and $|a| = 0$ if and only if $a = 0$.*

2. $|ab| = |a| \cdot |b|$ for all $a, b \in \mathbb{R}$.

3. $|a|^2 = a^2$ for all $a \in \mathbb{R}$.

Lemma 1.15 (Triangle Inequalities)**.** *Let $a, b \in \mathbb{R}$. Then*

1. $|a+b| \leq |a| + |b|$
2. $|a-b| \geq ||a| - |b||$.

Proof. 1. We know that $\pm a \leq |a|$ and $\pm b \leq |b|$. Thus $a+b \leq |a|+|b|$, and $-(a+b) \leq |a|+|b|$, by order additivity. Thus $|a+b| \leq |a|+|b|$.

2. This is really just the triangle inequality rearranged. In particular, we notice that $|a| = |a-b+b| \leq |a-b| + |b|$ by the triangle inequality. Adding $-|b|$ to both sides gives $|a| - |b| \leq |a-b| + |b| + (-|b|)$ by order additivity, and inverses and identity give us $|a| - |b| \leq |a-b|$.

   We can make the same argument with a and b switched, which gives us $|b| - |a| \leq |b-a| = |a-b|$. Thus these two statements together give us $||a| - |b|| \leq |a-b|$.

□

We sometimes want to turn expressions involving absolute value into expressions that don't. The following proposition is useful for this:

Proposition 1.16. *$|x-a| < \epsilon$ if and only if $a - \epsilon < x < a + \epsilon$.*

Proof. $|x-a| < \epsilon$ if and only if $x - a < \epsilon$ and $a - x < \epsilon$. The first inequality is equivalent to $x < a + \epsilon$ by order addition; the second is equivalent to $a - \epsilon < x$. □

1.3 The Least Upper Bound Property

We now understand what an ordered field is, but we've identified at least two: $\mathbb{Q}$ and $\mathbb{R}$. (In fact there are infinitely many ordered fields that contain $\mathbb{Q}$ and are contained in $\mathbb{R}$; an example is $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$).

If we want to understand what makes $\mathbb{R}$ special, we need one extra property. Formally this property is called "completeness", a term which we will return to later in the course. Here we can give a much simpler characterization.

Definition 1.17. Suppose F is an ordered field, and $S \subset F$. We say that a is an *upper bound* for S if $s \leq a$ for all $s \in S$. If S has an upper bound, we say that it is *bounded above*.

We say that y is a *least upper bound* of S if y is an upper bound for S, and if a is also an upper bound for S, then $y \leq a$. We suptimes write that $y = \sup S$.

Lemma 1.18. *Let F be an ordered field and $S \subset F$. If S has an least upper bound, then that upper bound is unique.*

Proof. Suppose that a and b are both least upper bounds of S. Then since a is a least upper bound and b is an upper bound, then by definition of least upper bound $a \leq b$.

But b is a least upper bound and a is an upper bound, so by definition of least upper bound $b \leq a$.

Since $a \leq b$ and $b \leq a$, we know that $a = b$. □

Exercise 1.19. *Let F be an ordered field and $S \subset F$. Let y be a least upper bound of S, and let $x < y$. Prove that there is an $s \in S$ such that $x < s$.*

Example 1.20. Let $S = \{x : x \leq 0\}$ be a subset of $\mathbb{R}$. Then we claim that 0 is the least upper bound for S.

First we prove that 0 is an upper bound for S. If $x \in S$ then $x \leq 0$ by definition of S.

Now we prove that 0 is a least upper bound—that is, we prove that if y is an upper bound for S, then $0 \leq y$. But suppose that y is an upper bound for S. Then since $0 \in S$, we have that $0 \leq y$. Thus 0 is the least upper bound for S.

Example 1.21. Let $S = \{x : x < 0\}$ be a subset of $\mathbb{R}$. Then we claim that 0 is a least upper bound for S.

If $x \in S$ then $x < 0$, so $x \leq 0$, so 0 is an upper bound for S.

Now we want to prove that 0 is the least upper bound. We can't do the same thing we did last time, because 0 is not an element of S. Instead we do a proof by contradiction.

Let y be an upper bound for S, and suppose $y < 0$. By properties of ordered fields, since $1 > \frac{1}{2}$ we know that $y < y/2$, and since $\frac{1}{2} > 0$ we know that $y/2 < 0$. Thus $y/2 \in S$ and $y < y/2$, contradicting our assumption that y was an upper bound for S.

Thus if y is an upper bound for S, we have $y \geq 0$, so 0 is the least upper bound.

Example 1.22. Let $S = \{x : x^2 < 2\}$ be a subset of $\mathbb{Q}$. Then S is bounded above, for instance, $x < 2$ for all $x < S$. But S has no least upper bound in $\mathbb{Q}$.

Definition 1.23. Let F be an ordered field and let S be a subset of F. We say that S has the *Least Upper Bound Property* if, whenever T is a non-empty subset of S and T is bounded above, then T has a least upper bound in S.

Now we can completely characterize the real numbers: $\mathbb{R}$ is the unique ordered field that satisfied the Least Upper Bound property. Whenever we have a set of real numbers that is bounded above, there is a real number that is the least upper bound for that set.

Lemma 1.24 (Archimedean Property)**.** *If $x \in \mathbb{R}$, then there exists a $n \in \mathbb{N}$ such that $n > x$.*

Proof. Suppose this is false; then there is some real number x such that $n \leq x$ for all $n \in \mathbb{N}$. This would mean that $\mathbb{N}$ is bounded above, and by the Least Upper Bound property it must have a least upper bound a.

But if n is a natural number, then so is $n+1$, so we see that $n+1 \leq a$ for all $n \in \mathbb{N}$. Then $n \leq a-1$ for all $n \in \mathbb{N}$, which means that $a-1$ is an upper bound for $\mathbb{N}$; but $a-1 < a$, contradicting the assumption that a is a least upper bound.

Thus no least upper bound can exist; so the set of integers cannot be bounded above by x. This proves our lemma. □

Exercise 1.25. *Prove that for any real number $\epsilon > 0$, there is a $n \in \mathbb{N}$ such that $1/n < \epsilon$.*

Proposition 1.26. *For any $x \in R$, there is an integer $n \in \mathbb{Z}$ such that $n \leq x < n+1$.*

Proof. By the Archimedean property, there is an N such that $N > |x|$. Consider the set of integers $S = \{n : -N < n < N\}$. This is a finite set, so we can take the largest n in S such that $n \leq x$.

Then we claim $n+1 > x$. Otherwise, we would have $n+1 < N$, so it would be in S, and n wouldn't be the largest element of S that is less than or equal to x. □

Corollary 1.27. *For any $x \in \mathbb{R}$ and any positive integer N, there is an integer n such that $\frac{n}{N} \leq x < \frac{n+1}{N}$.*

Proposition 1.28. *For every $x \in \mathbb{R}$ and $\epsilon > 0$, there is a rational number $r \in \mathbb{Q}$ such that $|x - r| < \epsilon$.*

Proof. By our exercise 1.25, we know there is a positive integer N with $1/N < \epsilon$. Then by corollary 1.27 there is some integer n with $n/N \leq x < (n+1)/N$, so we have $0 < x - \frac{n}{N} < \frac{n+1}{N} - \frac{n}{N} = \frac{1}{N}$.

Since these numbers are all positive, we can take absolute values, and we get $0 < |x - \frac{n}{N}| < \frac{1}{N}$. Thus we take our rational number $r = \frac{n}{N}$. □

1.4 Constructing the Real Numbers

So what do the real numbers look like? We know the real numbers contain all the rational numbers; this gives us an ordered field. And the real numbers satisfy the Least Upper Bound property. So what extra numbers does this give us?

Proposition 1.29. *Let x be a positive real number. Then there is a unique positive real number y such that $y^2 = x$.*

Proof. If $0 < y_1 < y_2$ then $y_1^2 < y_2^2$, so any positive square root is unique. We just need to show that a positive square root exists.

Let $S = \{a \in \mathbb{R} : 0 \leq a^2 < x\}$. Then S is non-empty, since $0 \in S$, and S is bounded above by $\max\{1, x\}$, since $a < a^2$ for $a \geq 1$. By the Least Upper Bound principle, we know that S has a least upper bound; let $y = \sup(S)$. We claim that $y^2 = x$.

First we observe that $y > 0$. We know that $(\min\{1, x\})^2 \leq \min\{1, x\} \cdot 1 = \min\{1, x\} < x$, so $\min\{1, x\} \in S$; and since $1, x > 0$ we know that $\min\{1, x\} > 0$. Thus $y \geq \min\{1, x\} > 0$.

Now for any real number ϵ with $0 < \epsilon < y$, we see that $0 < y - \epsilon < y < y + \epsilon$, and so $(y - \epsilon)^2 < y^2 < (y + \epsilon)^2$. But we can also see that since $y - \epsilon \in S$, then $(y - \epsilon)^2 < x$, and since $y + \epsilon \notin S$ we know that $(y + \epsilon)^2 \geq x$. We conclude that $(y - \epsilon)^2 < x < (y + \epsilon)^2$. Then order additivity gives us

$$\begin{aligned}
(y - \epsilon)^2 < y^2 < (y + \epsilon)^2 \\
(y - \epsilon)^2 < x < (y + \epsilon)^2 \\
(y - \epsilon)^2 - (y + \epsilon)^2 < y^2 - x < (y + \epsilon)^2 - (y - \epsilon)^2 \\
|y^2 - x| < (y + \epsilon)^2 - (y - \epsilon)^2 = 4y\epsilon.
\end{aligned}$$

But this inequality holds for any ϵ with $0 < \epsilon < y$, so for any $a > 0$ we can choose ϵ so that $4y\epsilon < a$. Thus $|y^2 - x| < a$ for all $a > 0$, and so $|y^2 - x| = 0$. Thus $y^2 = x$ as claimed. □

As a corollary, this tells us that there are real numbers which are not rational numbers: $\sqrt{2}$ is a real number, but is not rational.

It also tells us that the order of the reals is more or less uniquely specified. We just saw that the set of positive real numbers is precisely the set of squares of real numbers. That is, $\{x : x \geq 0\} = \{y^2 : y \in \mathbb{R}\}$. Since knowing the set of positive real numbers is enough to determine the order completely, this means that the Least Upper Bound property allows only one possible order structure.

But how can we be sure we've found all the real numbers? Here we can turn to infinite decimals.

Definition 1.30. If a_0 is any integer, n a positive integer, and $a_1, \ldots, a_n$ are elements of the set $\{0, \ldots, 9\}$, then we define the finite decimal

$$a_0.a_1 \ldots a_n = a_0 + \frac{a_1}{10} + \cdots + \frac{a_n}{10^n}.$$

Remark 1.31. If $a_0 < 0$ this is actually not the usual way we interpret a finite decimal; but it's much more convenient for what we're doing. The difference doesn't matter to anything terribly important.

If $m < n$, we see that

$$\begin{aligned} a_0.a_1 \dots a_m &\leq a_0.a_1 \dots a_m + a_{m+1}10^{-m-1} + \dots + a_n 10^{-n} \\ &\leq a_0.a_1 \dots a_m + 9 \cdot 10^{-m-1} + \dots + 9 \cdot 10^{-n} \\ &< a_0.a_1 \dots a_m + 10^{-m} \end{aligned}$$

where we obtain the last inequality by adding 10^{-n} to the previous expression. This means that for any finite decimal, we have the bounds

$$a_0.a_1 \dots a_m \leq a_0.a_1 \dots a_n < a_0.a_1 \dots a_m + 10^{-m}.$$

This allows us to define infinite decimals:

Definition 1.32. If a_0 is any integer and $a_1, a_2, \dots$ is a sequence of elements of $\{0, \dots, 9\}$, then we define the *infinite decimal* $a_0.a_1a_2\dots$ to be the least upper bound of the set $\{a_0.a_1 \dots a_n : n \in \mathbb{N}\}$.

For this definition to make sense, we need to check that this set *has* a least upper bound. But we know the set is bounded above, for example by $a_0 + 1$; thus by the Least Upper Bound principle it has a least upper bound in $\mathbb{R}$.

Proposition 1.33. *Every real number can be represented by an infinite decimal.*

Proof. Let $x \in \mathbb{R}$. Then by corollary 1.27, taking $N = 10^{-m}$, there is some finite decimal $a_0.a_1 \dots a_m$ such that

$$a_0.a_1 \dots a_m < x < a_0.a_1 \dots a_m + 10^{-m}.$$

We can do this for any m, and if $n > m$ then the first m terms of both finite decimals will be the same. Thus we can define an infinite decimal by taking these $a_0.a_1 \dots a_m$ for each $m \in \mathbb{N}$, and then x is equal to this infinite decimal. □

Thus we've seen that every real number can be given by an infinite decimal. And this means that the entire field $\mathbb{R}$ is uniquely specified. We've seen that there is only one order we can choose compatible with the Least Upper Bound principle; and given that order, the set of real numbers is precisely the set of infinite decimals. This justifies our definition that $\mathbb{R}$ is "the" complete ordered field.

Remark 1.34. There are two other definitions or "constructions" of the reals you will sometimes see. Both of them construct sets of rational numbers, and define real numbers to be equivalence classes of these sets of rational numbers.

Later in this course we will see how we could define the reals to be equivalence classes of "Cauchy sequences" of rational numbers. This just means that a real number is the limit of some sequence of rational numbers—and if a sequence looks like it should converge, but there is no rational number it converges to, we call that a real number.

The most rigorous construction is via the use of Dedekind cuts. Here we define real numbers to be equivalence classes of partitions of the rational numbers into two sets, where the first is strictly smaller than the second and contains no greatest element. So the square root of two would be the partition (A, B) where $A = \{a \in \mathbb{Q} : a^2 < 2 \text{or} a < 0\}$, and $B = \{b \in \mathbb{Q} : b^2 \geq 2 \text{and} b \geq 0\}$.

We can then define addition and multiplication on these partitions, and check that they satisfy the field axioms, and the order axioms, and the least upper bound property. But this is tedious and not terribly enlightening, so we'll avoid it.

If you want to learn more about this, you can see p.186 of Rogers and Boman, which is available free online and linked in the syllabus.

2 Metric Spaces

2.1 Metrics

Definition 2.1. A metric space is a set E and a function $d : E \times E \to \mathbb{R}$ such that for all $p, q, r \in E$:

1. (Non-negativity) $d(p, q) \geq 0$, and $d(p, q) = 0$ if and only if $p = q$;
2. (Symmetry) $d(p, q) = d(q, p)$;
3. (Triangle inequality) $d(p, r) \leq d(p, q) + d(q, r)$

We say that d is a *metric* on $\mathbb{R}$.

Example 2.2. The most important and fundamental example of a metric space is $\mathbb{R}$. We define our metric to be $d(a, b) = |a - b|$. All three properties of the definition follow directly from the equivalent properties of the absolute value function.

Example 2.3. Let $E = \mathbb{R}^n$ be the set of ordered n-tuples of real numbers, and define

$$d((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

Then this is a metric space, and we call d the *Euclidean metric*. This is exactly the definition of distance we get doing normal three-dimensional geometry from the Pythagorean Theorem. (Some books, like Rosenlicht, call this space E^n to stand for "Euclidean space").

It's easy to verify the first two properties hold. Non-negativity holds because a sum of squares is always ≥ 0, and is equal to zero if and only if all of the terms are zero. And symmetry holds because $(x_i - y_1)^2 = (y_i - x_i)^2$.

Verifying the triangle inequality is a bit trickier, and depends on the Cauchy-Schwarz inequality, which you may recall from Linear Algebra.

Lemma 2.4 (Cauchy-Schwarz Inequality)**.** *Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be real numbers. Then*

$$|x_1 y_1 + \cdots + x_n y_n| \leq \sqrt{x_1^2 + \cdots + x_n^2}\sqrt{y_1^2 + \cdots + y_n^2}.$$

This may look more familiar in the form

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

Now we can prove that the Euclidean metric satisfies the triangle inequality:

$$\begin{aligned}(x_1+y_1)^2+\cdots+(x_n+y_n)^2 &= (x_1^2+\cdots+x_n^2)+2(x_1y_1+\cdots+x_ny_n)+(y_1^2+\cdots+y_n^2)\\ &\leq (x_1^2+\cdots+x_n^2)+2\sqrt{x_1^2+\cdots+x_n^2}\sqrt{y_1^2+\cdots+y_n^2}\\ &\quad +(y_1^2+\cdots+y_n^2)\\ &= \left(\sqrt{x_1^2+\cdots+x_n^2}+\sqrt{y_1^2+\cdots+y_n^2}\right)^2.\end{aligned}$$

Taking square roots gives

$$\sqrt{(x_1+y_1)^2+\cdots+(x_n+y_n)^2} \leq \sqrt{x_1^2+\cdots+x_n^2}+\sqrt{y_1^2+\cdots+y_n^2}.$$

Now we can compute

$$\begin{aligned}d(\vec{x},\vec{z}) &= \sqrt{(x_1-z_1)^2+\cdots+(x_n-z_n)^2}\\ &= \sqrt{((x_1-y_1)+(y_1-z_1))^2+\cdots+((x_n-y_n)+(y_n-z_n))^2}\\ &\leq \sqrt{(x_1-y_1)^2+\cdots+(x_n-y_n)^2}+\sqrt{(y_1-z_1)^2+\cdots+(y_n-z_n)^2}\\ &= d(x,y)+d(y,z).\end{aligned}$$

Example 2.5. Let $E=\mathbb{R}^n$ and let $d(\vec{x},vecy)=\max\{|x_i-y_i|\}$. We can check this is a metric on E.

1. Let $\vec{x},\vec{y}\in\mathbb{R}^n$. Then $d(\vec{x},\vec{y})\geq 0$ since $|x_i-y_i|\geq 0$ for any real numbers x_i,y_i. $d(\vec{x},\vec{y})=0$ if and only if $|x_i-y_i|=0$ for every i, which happens if and only if $x_i=y_i$ for every i.

2. $d(\vec{x},\vec{y})=\max\{|x_i-y_i|\}$. But $|x_i-y_i|=|y_i-x_i|$, so $\max\{|x_i-y_i|\}=\max\{|y_i-x_i|\}=d(\vec{y},\vec{x})$.

3. Let $\vec{x},\vec{y},\vec{z}\in\mathbb{R}^n$. We know that $d(\vec{x},\vec{y})=\max\{|x_i-y_i|\}$ and $d(\vec{y},\vec{z})=\max\{|y_i-z_i\}$. For each i, we have $|x_i-z_i|\leq|x_i-y_i|+|y_i-z_i|$.

   Then $d(\vec{x},\vec{z})=\max\{|x_i-z_i|\}=|x_j-z_j|$ for some specific j. But we know that

$$\begin{aligned}|x_j-z_j| &\leq |x_j-y_j|+|y_j-z_j|\\ &\leq \max\{|x_i-y_i|\}+\max\{|y_i-z_i|\}=d(\vec{x},\vec{y})+d(\vec{y},vecz).\end{aligned}$$

Example 2.6. Let E be any set and for every $x, y \in E$, define

$$d(x,y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases}$$

This is called the *discrete metric* and E is a *discrete metric space*.

1. It's clear that $d(x,y) \geq 0$ since $d(x,y) \in \{0,1\}$. Further, $d(x,y) = 0$ if and only if $x = y$ by definition.

2. $d(x,y) = d(y,x)$ since $x = y$ if and only if $y = x$.

3. Let $x, y, z \in E$. If $x = z$ then $d(x,z) = 0$, and $d(x,y) + d(y,z) \geq 0$, so $d(x,z) \leq d(x,y) + d(y,z)$.

   If $x \neq z$ then $d(x,z) = 1$. But at least one of $x \neq y$ or $y \neq z$, so $d(x,y) + d(y,z) \geq 1$. Thus $d(x,z) \leq d(x,y) + d(y,z)$.

Example 2.7. Let $E = \mathcal{C}([0,1], \mathbb{R})$ be the space of continuous functions from $[0,1]$ to $\mathbb{R}$. For $f, g \in E$, define

$$d(f,g) = \int_0^1 (f(t) - g(t))^2 \, dt.$$

Then this is a metric on E, called the L^2 *metric*.

Proposition 2.8 (Reverse Triangle Inequality)**.** *Let (E,d) be a metric space, and let $p, q, r \in E$. Then $|d(p,r) - d(q,r)| \leq d(p,q)$.*

Proof. Geometrically, this is the claim that the difference of any two sides of a triangle is less than the length of the third side—otherwise the third side wouldn't reach all the way.

Formally we can prove this from the triangle inequality. By the triangle inequality we have

$$\begin{aligned} d(p,r) &\leq d(p,q) + d(q,r) & d(q,r) &\leq d(q,p) + d(p,r) \\ d(p,r) - d(q,r) &\leq d(p,q) & d(q,r) - d(p,r) &\leq d(q,p) \\ d(p,r) - d(r,q) &\leq d(p,q) & -(d(p,r) - d(r,q)) &\leq d(p,q) \end{aligned}$$

and combining these two inequalities gives the desired result.

□

2.2 Open and Closed Sets

To understand these metrics, it's helpful to figure out what a circle looks like. A circle captures the idea of equal distances, and everything inside a circle is relatively close to the center. We want to generalize this to any metric.

Definition 2.9. Let E be a metric space, $x_0 \in E$, and $r > 0$ a real number. We define the *open ball* centered at x_0 with radius r as

$$B(x_0, r) = B_r(x_0) = \{x \in E : d(x_0, x) < r\}.$$

We define the *closed ball* centered at x_0 with radius r as

$$\overline{B}(x_0, r) = \overline{B}_r(x_0) = \{x \in E : d(x_0, x) \leq r\}.$$

If our metric space is $\mathbb{R}^3$, then an open or closed ball is just a literal three-dimensional ball; an open ball doesn't contain the spherical boundary, and a closed ball does. In $\mathbb{R}^2$, an open ball is the interior of a circle, and the closed ball is the circle including its boundary. In some other metrics, the balls are more unusual.

In $\mathbb{R}$ these concepts are even more familiar. The open ball $B_r(x_0)$ is the open interval $(x_0 - r, x_0 + r)$, and the closed ball $\overline{B}_r(x_0)$ is the closed interval $[x_0 - r, x_0 + r]$.

Example 2.10. If E is a metric space under the discrete metric, then $B_r(x_0) = \{x_0\}$ for $r \leq 1$ and $B_r(x_0) = E$ for $r > 1$.

$\overline{B}_r(x_0) = \{x_0\}$ for $r < 1$ and $\overline{B}_r(x_0) = E$ for $r \geq 1$.

Example 2.11. If $E = \mathbb{R}^2$ under the sup metric, then $B_r(x_0)$ is a square centered at x_0 with side length $2r$.

Example 2.12. If $E = \mathcal{C}([0, 1], \mathbb{R})$ is the space of continuous functions from $[0, 1]$ to $\mathbb{R}$ under the L^2 metric, then the open ball centered at the 0 function of radius 1 is the set of all functions f such that $\int_0^1 f(x)^2 \, dx < 1$.

We see a clear distinction between open balls and closed balls: closed balls contain their boundaries, and open balls do not. (And we can imagine a set that contains *some* of its boundary and is thus neither open nor closed). We'd like to extend this distinction to all sets, not just the ones that are shaped like sphers.

Definition 2.13. Let E be a metric space and $U \subset E$. We say that U is an *open set* if for each $x \in U$, there is some $r > 0$ such that $B_r(x) \subset U$.

Example 2.14. Consider the space $E = \mathbb{R}^2$ with the sup metric. Let $U = \{(x, y) : x > 0\}$. Let's prove that U is open.

Let $(x, y) \in U$. We claim that $B_x(x, y) \subset U$. To prove this, let $(x_1, y_1) \in B_x(x, y)$; we want to show that $(x_1, y_1) \in U$, that is, that $x_1 > 0$.

Since $(x_1, y_1) \in B_x(x, y)$, we know that $d((x_1, y_1), (x, y)) < x$. Thus $\max\{|x_1 - x|, |y_1 - y|\} < x$, and in particular $|x_1 - x| < x$.

From here we can make one of two arguments. One is that $-x < x_1 - x < x$, so $0 < x_1 < 2x$. The other is that by the reverse triangle inequality, we have $x > |x - x_1| \geq x - x_1$, and thus $x_1 > 0$. Either way, we see that $x_1 > 0$, so $(x_1, y_1) \in U$.

This proves that $B_x(x, y) \subset U$. Since $(x, y) \in U$ was arbitrary, we conclude that U is open.

Proposition 2.15. *Let E be a metric space. Then:*

1. *$\varnothing$ and E are open sets.*
2. *An arbitrary union of open sets is open.*
3. *A* finite *intersection of open sets is open.*

Proof. 1. For every $x \in \varnothing$, anything is true, since there are no $x \in \varnothing$. Thus for every $x \in \varnothing$ there is an $r > 0$ such that $B_r(x) \subset \varnothing$. Thus $\varnothing$ is open.

   For every $x \in E$ and every $r > 0$, $B_r(x) \subset E$ by definition. Thus E is open.

2. Let $\{U_\alpha\}$ be a collection of open sets, and let $x \in \bigcup U_\alpha$. Then $x \in U_\beta$ for some β, and since U_β is open, there is a $r > 0$ such that $B_r(x) \subset U_\beta$. But $U_\beta \subset \bigcup U_\alpha$, so $b_r(x) \subset \bigcup U_\alpha$.

3. Let $U_1, \ldots, U_n$ be open sets and let $x \in \bigcap U_i$. Then $x \in U_i$ for each i, and thus there is a $r_i > 0$ such that $B_{r_i}(x) \subset U_i$.

   Let $r = \min\{r_1, \ldots, r_n\}$. Then $B_r(x) \subset B_{r_i}(x)$ for every i, so $B_r(x) \subset U_i$ for each i. Thus $B_r(x) \subset \bigcap U_i$. We conclude that $\bigcap U_i$ is open.

□

Notice that while openness is preserved for any union, it is only preserved for finite intersections. (We needed finiteness when we took the minimum of the set—an infinite collection might have no minimum, and the greatest lower bound might be zero, which doesn't work). In your homework you will look at an example of an infinite intersection of open sets that isn't open.

Proposition 2.16. *Let E be a metric space, let $x \in E$, and $r > 0$. Then $B_r(x)$ is an open set.*

Proof. We need to prove that any point in $B_r(x)$ has an open ball around it that is inside $B_r(x)$. We're essentially going to use the triangle inequality to do this: if $y \in B_r(x)$ and you're close enough to y, you also have to be reasonably close to x.

So let $y \in B_r(x)$. Then set $s = d(x, y)$; we know that $s < r$ by definition of an open ball. We claim that $B_{r-s}(y) \subset B_r(x)$.

Let $z \in B_{r-s}(y)$. Then $d(y, z) < r - s$. But $d(x, y) = s$, so by the triangle inequality we have $d(x, z) \leq d(x, y) + d(y, z) < s + r - s = r$. Thus $z \in B_r(x)$. □

Definition 2.17. If E is a metric space, we say that $V \subset E$ is *closed* if the complement of V in E, denoted V^C, is open.

Exercise 2.18. *If E is a metric space, $x \in E$, and $r > 0$, then $\overline{B}_r(x)$ is a closed set.*

Proposition 2.19. *1. $\varnothing$ and E are closed sets.*

2. Any intersection of closed sets is closed.

3. Any finite *union of closed sets is closed.*

Proof. This is basically a corollary to proposition 2.33.

1. E is the complement of $\varnothing$ and $\varnothing$ is the complement of E.
2. Let $\{V_\alpha\}$ be a collection of closed sets. For each V_α, the complement U_α is open, so $\bigcup U_\alpha$ is open. $\bigcap V_\alpha$ is the complement of $\bigcup U_\alpha$.
3. Let $V_1, \ldots, V_n$ be closed. For each V_i, the complement U_i is open, so $U_1 \cap \cdots \cap U_n$ is open. $V_1 \cup \cdots \cup V_n$ is the complement of $U_1 \cap \cdots \cap U_n$.

□

Remark 2.20. There's one important subtlety to be aware of: whether a set is open depends not only on the set, but on the metric space it's a part of. This is really clear when we change the metric: the set $\{0\} \subset \mathbb{R}$ is open in the discrete metric, but not in the Euclidean metric.

But especially weird things happen on boundaries. Let $E = [0, 1]$ be a metric space under the metric $d(x, y) = |x - y|$ inherited from the real line. Then it's pretty clear that $B_{1/2}(0)$ is open. But $B_{1/2}(0) = [0, 1/2)$, so in this metric space, $[0, 1/2)$ is open. In contrast, in $\mathbb{R}$, the set $[0, 1/2)$ is neither open nor closed.

Definition 2.21. Let (E, d) be a metric space. We say a subset $U \subset E$ is *bounded* if there is some open or closed ball $B_r(x)$ or $\overline{B}_r(x)$ such that $U \subset B_r(x)$.

Example 2.22. A subset of $\mathbb{R}$ is bounded if and only if it is bounded above and below. If $S \subset \mathbb{R}$ with a as a lower bound and b as an upper bound, then $S \subset [a, b]$. But $[a, b]$ is a closed ball centered at $(a + b)/2$.

Example 2.23.

Conceptually, an open set is a set that doesn't contain any of its boundary; a closed set is a set that contains its entire boundary. We can't use this as a definition since we haven't defined what a "boundary" is yet. But we *can* somewhat justify this for $\mathbb{R}$ where the boundary of a set is just a maximum or minimum element.

Proposition 2.24. *A non-empty closed subset of $\mathbb{R}$ that is bounded above contains a greatest element.*

A non-empty closed subset of $\mathbb{R}$ that is bounded below contains a least element.

Proof. We'll prove the first claim; the proof of the second is identical.

Let S be a non-empty closed subset of $\mathbb{R}$ that is bounded above. Then S has a least upper bound, so let $a = \sup(S)$. If $a \in S$ then it is the greatest element, so instead suppose $a \in S^C$ the complement of S.

S^C is open by definition, so there is some $r > 0$ such that $B_r(a) \in S^C$. That is, $(a - r, a + r) \in S^C$, so no element of S can be in $(a - r, a + r)$. This implies that $a - r$ is an upper bound for S, but $a - r < a = \sup(S)$, which is a contradiction. So $a \in S$ is a largest element. □

2.3 Convergent Sequences

In this section we want to define the fundamental idea of calculus, which is the idea of convergence to a limit.

Recall a sequence is just an ordered list of things. (If we're being extremely formal, we say a sequence of elements of S is a function from the natural numbers to S; the nth element of the sequence is $f(n)$).

Intuitively, we want to use the word "limit" to describe a point that a sequence gets really close to. People sometimes say "gets closer and closer", but this is wrong for two reasons. First, because a sequence can oscillate around a point and not get strictly closer. Second, because it's entirely possible to get continually closer to a value but not get anywhere near

it; we'd never say that the limit of the sequence $x_n = 1/n$ is -3, even though the values of the sequence do get closer and closer to -3.

We can close these loopholes by giving a formal definition.

Definition 2.25. Let (E, d) be a metric space and $x_1, x_2, \dots$ be a sequence of elements of E. We say that the sequence (x_n) *converges* to a limit $L \in E$, or that $lim_{n\to\infty} x_n = L$, if for every $\epsilon > 0$ there exists a $N \in \mathbb{N}$ such that, whenever $n > N$, then $d(x_n m L) < \epsilon$. Symbolically:

$$\forall \epsilon > 0 \quad \exists N \in \mathbb{N} \ni (n > N \Rightarrow d(x_n, L) < \epsilon)$$

Notice that N depends on ϵ, and n depends on N. We don't get any control over the value of ϵ, but we can pick any value of N depending on ϵ to make this definition work.

Convergence depends on the specific metric space. First, because it often depends on the specific metric used. (Very few sequences converge in the discrete metric). Second, because it depends on the points in the space; the sequence $3, 3.1, 3.14, \dots$ has no limit in the rationals and thus doesn't converge.

Example 2.26. Let's prove that $\lim_{n\to\infty} \frac{3}{n} = 0$.

Let $\epsilon > 0$, and set $N = \frac{3}{\epsilon}$. Then if $n > N$, we have

$$d(3/n, 0) = \left|\frac{3}{n}\right| = \frac{3}{n} < \frac{3}{N} = \frac{3}{3/\epsilon} = \epsilon.$$

Thus by definition, $\lim_{n\to\infty} \frac{3}{n} = 0$.

From this definition, we can prove lots of nice properties for limits. The first is that they're unique—the same sequence can't have more than one limit.

Proposition 2.27. *Let (E, d) be a metric space. Then a sequence (x_n) in E has at most one limit.*

Proof. This proof uses a common trick which I think of as the $\epsilon/2$ or ϵ/n trick. Heuristically, we want to show that if x_n gets arbitrarily close to L and also to K, then L and K have to be arbitrarily close together, and thus identical. We want to show that $d(L, K) < \epsilon$; to do this, we show that two other distances are less than $\epsilon/2$.

Suppose $\lim_{n\to\infty} x_n = L$ and $\lim_{n\to\infty} x_n = K$. Let $\epsilon > 0$. Then by definition of limit, there is a $N_1 \in \mathbb{N}$ such that if $n > N_1$ then $d(x_n, L) < \epsilon/2$. And there is a $N_2 \in \mathbb{N}$ such that if $n > N_2$ then $d(x_n, K) < \epsilon/2$.

Let $N = \max\{N_1, N_2\}$. Then if $n < N$, we know that $d(x_n, L) < epsilon/2$ and $d(x_n, K) < \epsilon/2$. Adding these inequalities tells us that

$$d(x_n, L) + d(x_n, K) < \epsilon$$

and the triangle inequality tells us that

$$d(L, K) \leq d(x_n, L) + d(x_n, K).$$

Thus $d(L, K) < \epsilon$.

But ϵ was an arbitrary positive real number, so $d(L, K) = 0$ and thus $L = K$. □

An important property of limits of sequences is that you can distort them in a lot of ways without changing the limit. One example is that for any finite $k \in \mathbb{N}$, you can change the first k terms of the sequence without changing the limit; you just have to make sure that you take $N > k$. But we can also make infinite changes if we're careful.

Definition 2.28. If $x_1, x_2, \dots$ is a sequence and $n_1, n_2, \dots$ is a strictly increasing sequence of integers, we say the sequence of elements $x_{n_1}, x_{n_2}, \dots$ is a *subsequence* of (x_n).

Proposition 2.29. *If a sequence (x_n) converges to a limit L, then any subsequence of (x_n) also converges to L.*

Proof. Suppose $\lim_{n\to\infty} x_n = L$. Let $\epsilon > 0$. Then there is a $N \in \mathbb{N}$ such that if $n > N$ then $d(x_n, L) < \epsilon$.

Since $n_k \geq k$ for any $k \in \mathbb{N}$, if $k > N$ then we know that $n_k > N$, and so $d(x_{n_k}, L) < \epsilon$.

Thus by definition of limit, $\lim_{k\to\infty} x_{n_k} = L$. □

Exercise 2.30. *Let (E, d) be a metric space, and suppose $x_1, x_2, \dots$ is a sequence that converges to x in E (that is, $\lim_{n\to\infty} x_n = x$). Then the sequence $x_1, x, x_2, x, x_3, \dots$ converges to x.*

Remark 2.31. We can reframe the idea of limits in terms of open sets. It's clear enough that $\lim_{n\to\infty} x_n = L$ if for every open ball $B_\epsilon(L)$ centered at L, there is some $N \in \mathbb{N}$ such that $x_n \in B_\epsilon(L)$ for all $n > N$—this is just the original definition rephrased.

It's not too much harder to convince yourself that $\lim_{n\to\infty} x_n = L$ if and only if, for every open set U with $L \in U$, there is some $N \in \mathbb{N}$ such that $x_n \in U$ for all $n > N$.

(In the field of topology, we have a definition of "open set" but not of "metric" or distance. In topology we can take this to be the definition of limit).

We've seen how to prove a specific sequence converges. We can also prove that one doesn't:

Example 2.32. Let $(x_n) = 1, 0, 1, 0, \ldots$ be a sequence of real numbers. Then (x_n) has no limit.

Suppose $\lim_{n\to\infty} x_n = x$. Then for every $\epsilon > 0$ there is some $N \in \mathbb{N}$ so that if $n > N$, then $d(x_n, L) < \epsilon$.

But for some even $n_1 > N$, we have $x_{n_1} = 0$, so $d(0, L) < \epsilon$. And for some odd $n_2 > N$, we have $x_{n_2} = 1$, so $d(1, L) < \epsilon$. By the triangle inequality, we have

$$d(0, 1) \leq d(1, L) + d(0, L) < 2\epsilon$$

and thus $1 < 2\epsilon$ for every $\epsilon > 0$. Since this is false, we have a contradiction, so no limit can exist.

2.4 Closures, Boundaries and Interiors

We said earlier that we want to think of a closed set as "containing its boundary," and an open set as not doing that. Sequences and their convergence lets us define this more carefully.

We first need to see how sequences behave in open and closed sets.

Proposition 2.33. *Let (E, d) be a metric space, and let $V \subset E$. Then V is closed if and only if every convergent sequence of points in V converges to a point in V.*

Proof. First, suppose V is closed. Let (x_n) be a sequence contained in V that converges to x, and suppose $x \notin V$. Then x is in the complement of V, which is open. Thus there is some ball $B_\epsilon(x) \subset V^C$.

By definition of convergence, there is some $N \in \mathbb{N}$ such that when $n > N$, we know that $d(x_n, x) < \epsilon$. Thus $x_n \in B_\epsilon(x) \subset V^C$, which is a contradiction since $x_n \in V$.

Now suppose V is not closed. This means that V^C is not open, and so there is some point $x \in V^C$ such that $B_\epsilon(x)$ is not a subset of V^C for any $\epsilon > 0$.

For each $n \in \mathbb{N}$, let $x_n \in B_{1/n}(x)$ such that $x \in V$. This is possible since $B_{1/n}(x) \not\subset V^C$. We claim that $\lim_{n\to\infty} x_n = x$.

Let $\epsilon > 0$. Then there is some $N \in \mathbb{N}$ with $1/N < \epsilon$. if $n < N$, then $x_n \in B_{1/n}(x)$ so $d(x_n, x) < 1/n < 1/N < \epsilon$. Thus $\lim_{n\to\infty} x_n = x$.

So if V is not closed, we can construct a sequence of points $x_n \in V$ that converges to a limit not in V. If no such sequence exists—if every sequence of points in V that converges has a limit in V—then V must be closed. This completes the proof. □

We now want to take sets and find ways to turn them into the closest closed and open sets we can.

Definition 2.34. Let (E, d) be a metric space, and $U \subset E$. We define the *closure* of U to be $\overline{U}$ the intersection of every closed set containing U.

It's clear that $\overline{U}$ contains U, since it's the intersection of sets which all contain U. It's also clear that $\overline{U}$ is closed, since any intersection of closed sets is closed. Less obvious is the folloiwng result:

Proposition 2.35. *Let (E, d) be a metric space and $U \subset E$. Then $\overline{U}$ is the set of the limits of all sequences in U that converge in E.*

Proof. First we show that the set of all limits of sequences in U is contained in the closure of U. Suppose (x_n) is a sequence in U, and $\lim_{n\to\infty} x_n = x$. Then for any closed set V containing U, we have $x_n \in V$ for all n, and by proposition 2.33 we have $x \in V$. Thus x is in every closed set containing U, and so in their intersection.

Now suppose $x \in \overline{U}$; we wish to show that x is the limit of some sequence of points in U. Let $n \in \mathbb{N}$. If $B_{1/n}(x) \cap U = \varnothing$, then $U \subset B_{1/n}(x)^C$; and since $B_{1/n}(x)^C$ is closed, this implies that $\overline{U} \subset B_{1/n}(x)^C$. But $x \in \overline{U}$, so this is impossible.

Thus $B_{1/n}(x) \cap U$ is non-empty. Now as in proposition 2.33 we can choose $x_n \in B_{1/n}(x) \cap U$, and $\lim_{n\to\infty} x_n = x$. □

The closure of U is the smallest closed set containing U; you can think of it as taking U and then adding the "boundary" (which we still haven't defined!).

Exercise 2.36. *Let (E, d) be a metric space, and let V be a closed subset of E. Prove that $\overline{V} = V$.*

Example 2.37. Consider $\mathbb{R}^2$ under the Euclidean metric. Then the closure of $B_1(0, 0)$ is the closed ball $\overline{B}_1(0, 0)$.

Clearly the closure is a subset of $\overline{B}_1(0, 0)$, since $\overline{B}_1(0, 0)$ is a closed set containing $B_1(0, 0)$, and the closure is the intersection of all of these.

Now suppose $x = (x_1, x_2) \in \overline{B}_1(0, 0)$. We want to prove that x is the limit of some sequence (x_n) that is contained in $\overline{B}_1(0, 1)$.

For each n, define $x_n = ((n-1)x/n, (n-1)y/n)$. Then

$$d(x_n, \vec{0}) = \sqrt{\frac{n-1}{n}}\sqrt{x_1^2 + x_2^2} \leq \sqrt{\frac{n-1}{n}} < 1$$

so $x_n \in B_1(0,0)$.

We claim that $\lim_{n\to\infty} x_n = x$. Let $\epsilon > 0$, and let $N > 1/\epsilon$. Then if $n > N$ we compute

$$\begin{aligned} d(x_n, x) &= \sqrt{(x_1(n-1)/n - x_1)^2 + (x_2(n-1)/n - x_2)^2} \\ &= \sqrt{(x_1/n)^2 + (x_2/n)^2} = \frac{1}{n}\sqrt{x_1^2 + x_2^2} = \frac{1}{n} d(x, \vec{0}) \\ &\leq \frac{1}{n} < \frac{1}{N} < \epsilon. \end{aligned}$$

We can also do the equivalent for open sets: take a set and remove the boundary.

Definition 2.38. Let (E,d) be a metric space, and $U \subset E$. We define the *interior* of U to be $\mathring{U}$ the union of every open subset of U.

As before, it's clear that this is an open set, since it is a union of open sets, and it's a subset of U, since it's a union of subsets.

Exercise 2.39. *If (E,d) is a metric space and $U \subset E$, prove the interior of U is the set of all points $x \in U$ such that some open ball containing x is also a subset of U.*

Lemma 2.40. *If U is a set in a metric space (E,d), then $\left(\mathring{U}\right)^C = \overline{U^C}$.*

Proof.

$$\left(\mathring{U}\right)^C = \left(\bigcup_{S\subset U \text{ open}} S\right)^C = \bigcap_{S\subset U \text{ open}} S^C = \bigcap_{V\supset U \text{ closed}} V = \overline{U}.$$

□

Definition 2.41. The *boundary* of a set U is $\partial U = \overline{U} \cap \overline{U^C}$.

Equivalently we can write that $\partial U = \overline{U} \setminus \mathring{U}$, since

$$\overline{U} \setminus \mathring{U} = \overline{U} \cap \left(\mathring{U}\right)^C = \overline{U} \cap \overline{U^C}.$$

Proposition 2.42. *Let (E,d) be a metric space and $S \subset E$. Then $x \in \partial S$ if and only if every ball centered at x contains points in S and also points in S^C.*

Proof. Suppose $x \in \partial S$. Then $x \in \overline{S}$, so x is the limit of some sequence x_n of points in S such that $\lim_{n\to\infty} x_n = x$. Similarly, $x \in \overline{S^C}$, so x is the limit of some sequence of points y_n in S^C such that $\lim_{n\to\infty} y_n = x$.

Then for every $\epsilon > 0$, there is some $N \in \mathbb{N}$ so that $n > N$ implies $x_n \in B_\epsilon(x)$. And there is some $M \in \mathbb{N}$ so that $m > M$ implies $y_m \in B_\epsilon(x)$. Thus any open ball centered at x contains (infinitely many) points in S and also (infinitely many) points in S^C.

Conversely, suppose any open ball centered at x contains points in S and also in S^C. Then for every $n \in \mathbb{N}$ there is a $x_n \in B_{1/n}(x) \cap S$ and a $y_n \in B_{1/n}(x) \cap S^C$. Then $\lim_{n\to\infty} x_n = x$, so $x \in \overline{S}$. But $\lim_{n\to\infty} y_n = x$ so $x \in \overline{S^C}$. □

3 Special Types of Metric Spaces

3.1 Limits in the Real Numbers

The metric space given by the real numbers has a few properties that we can't generalize to all metric spaces, but that are still quite important.

The first important property of the real numbers is that they form a field—that is, we can do arithmetic with them. (Many of these results generalize to "normed vector spaces", which are vector spaces that are also metric spaces in a compatible way. Normed vector spaces are mostly outside the scope of this course, but would be important to a second course in analysis).

Proposition 3.1. *Let $a_n, b_n \in \mathbb{R}$ such that $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b$. Then*

1. $\lim_{n\to\infty}(a_n + b_n) = a + b$
2. $\lim_{n\to\infty}(a_n - b_n) = a - b$
3. $\lim_{n\to\infty}(a_n b_n) = ab$
4. *If $\lim_{n\to\infty} b_n \neq 0$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{a}{b}$.*

Proof. 1. The goal is to show that if a_n is close to a, and b_n is close to b, then $a_n + b_n$ must be close to $a + b$.

Let $\epsilon > 0$. Then there is a N_1 such that $|a_n - a| < \epsilon/2$ for all $n > N_1$, and there is a N_2 such that $|b_n - b| < \epsilon/2$ for all $n > N_2$.

Let $N = \max\{N_1, N_2\}$. Then if $n < N$, we have

$$\begin{aligned} |a_n + b_n - (a + b)| &= |(a_n - a) + (b_n - b)| \\ &\leq |a_n - a| + |b_n - b| && \text{by the Triangle Inequality} \\ &< \epsilon/2 + \epsilon/2 = \epsilon && \text{because } n > N_1, N_2 \end{aligned}$$

2. Basically identical proof.

3. Exercise. (Hint: the quantity $a_n b$ may be helpful).

4. This one is a little trickier.

   First we prove that $\lim_{n\to\infty} \frac{1}{b_n} = \frac{1}{b}$. We see that

$$> \left|\frac{1}{b_n} - \frac{1}{b}\right| = \frac{|b - b_n|}{|b \cdot b_n|}.$$

The numerator can be made as small as we like since $\lim_{x\to\infty} b_n = b$; but we also need to make sure the *denominator* doesn't get too small.

But since b_n is close to b, we should be able to ensure that the denominator is bigger than $|b \cdot b/2| = b^2/2 > 0$. Then we just need to also make sure the numerator is smaller than $\epsilon b^2/2$.

Let $\epsilon > 0$. Then since $\epsilon b^2/2 > 0$, there is a N_1 such that $|b_n - b| < \epsilon b^2/2$ whenever $n > N_1$. And since $|b/2| > 0$, there is a N_2 such that $|b_n - b| < |b/2|$ whenever $n < N_2$.

Let $N = \max\{N_1, N_2\}$. Then if $n > N$, we have

$$\begin{aligned}
|b_n - b| &< |b/2| && n > N_2 \\
|b/2| &> |b_n - b| \geq |b| - |b_n| && \text{Reverse Triangle Inequality} \\
|b_n| &> |b| - |b/2| = |b/2| \\
|bb_n| &> |b^2/2| = b^2/2 \\
\frac{1}{|bb_n|} &< \frac{1}{b^2/2} \\
\left|\frac{1}{b_n} - \frac{1}{b}\right| &= \frac{|b_n - b|}{|bb_n|} < \frac{|b_n - b|}{b^2/2} \\
&< \frac{\epsilon b^2/2}{b^2/2} = \epsilon && n > N_1.
\end{aligned}$$

Thus $\lim_{n\to\infty} \frac{1}{b_n} = \frac{1}{b}$.

Now we can prove the original claim. We see that $\lim_{n\to\infty} \frac{a_n}{b_n} = \lim_{n\to\infty} a_n \frac{1}{b_n}$, and by the previous result this is equal to

$$\left(\lim_{n\to\infty} a_n\right)\left(\lim_{n\to\infty} \frac{1}{b_n}\right) = a\frac{1}{b} = \frac{a}{b}.$$

□

The other important property of the reals is that they are ordered and have the least upper bound property. This tells us a few things about sequence convergence.

Exercise 3.2. *If* $\lim_{n\to\infty} a_n = a$ *and* $\lim_{n\to\infty} b_n = b$*, and* $a_n \leq b_n$ *for all* n*, then* $a \leq b$*.*

Definition 3.3. We say a sequence (x_n) is *monotone increasing* or *monotonically increasing* if $x_n \leq x_{n+1}$ for every $n \in \mathbb{N}$. A sequence is *monotonically decreasing* if $x_n \geq x_{n+1}$ for every $n \in \mathbb{N}$.

Example 3.4. The sequence $(1/n)$ is monotonically decreasing.

The sequence $1, 1, 2, 2, 3, 3, \ldots$ is monotonically increasing.

The sequence $1, 1, 1, 1, \ldots$ is both monotonically increasing and monotonically decreasing.

Proposition 3.5 (Monotone Convergence)**.** *Let (x_n) be a monotonically increasing sequence of real numbers that is bounded above. Then (x_n) converges.*

Proof. This is a straightforward application of the Least Upper Bound property. The set $\{x_n\}$ is bounded above and non-empty, so it has a least upper bound; set $x = \sup\{x_n\}$. We claim that $\lim_{n\to\infty} x_n = x$.

Let $\epsilon > 0$. Then since x is the least upper bound of $\{x_n\}$, there is some x_N such that $x_N > x - \epsilon$. Suppose $n > N$. Then by monotonicity, $x_n > x_N > x - \epsilon$, but since x is an upper bound, $x \geq x_n$. Thus

$$\begin{aligned} x - \epsilon < x_n &\leq x \\ -\epsilon < x_n - x &\leq 0 < \epsilon \\ |x_n - x| &< \epsilon. \end{aligned}$$

Thus by definition, $\lim_{n\to\infty} x_n = x$.

□

Exercise 3.6. *Let $S \subset \mathbb{R}$ be nonempty and bounded above. Prove there is a monotone sequence (x_n) such that $x_n \in S$ and $\lim_{n\to\infty} x_n = \sup S$.*

Remark 3.7. Proposition 3.5 tells us that a sequence must have a limit, but doesn't tell us what that limit is. (Well, it tells us that the limit is the supremum, but if we don't know the supremum already that doesn't really help). It's still extremely useful to show that various limits have to (or can't) exist.

This monotone convergence property is actually equivalent to the Least Upper Bound property: if any monotonically increasing sequence converges, then any set must have a least upper bound. This is the second version of "completeness" we'll see in this course. We'll discuss the third, and most general, next week.

Unfortunately, most sequences are not monotone sequences. Fortunately, we can kind of fake it.

Proposition 3.8. *If $a_n = \sup\{x_k : k \geq n\}$, then (a_n) is monotone decreasing.*

Similarly, if $b_n = \inf\{x_k : k \geq n\}$, then the sequence (b_n) is monotone increasing.

Proof. If $m > n$ then $\{x_k : k \geq m\} \subset \{x_k : k \geq n\}$, so $\inf\{x_k : k \geq m\} \geq \inf\{x_k : k \geq n\}$. Thus $a_m \geq a_n$. □

Example 3.9. The sequence $(1/n)$ is already monotone decreasing. We see that for every n, $a_n = \sup\{1/k : k \geq n\} = 1/n$. But $b_n = \inf\{1/k : k \geq n\} = 0$, so the sequence a_n is constantly 0.

If $(x_n) = 0, 1, 0, 1, \ldots$, then $b_n = \inf\{0, 1\} = 0$ and $a_n = \sup\{0, 1\} = 1$. These sequences are distinct and constant.

If $(x_n) = 1, 1, 1/2, 1, 1/3, 1, 1/4, 1, \ldots$, then

$$a_n = \sup\{1/k : k \geq n/2\} \cup \{1\} = 1$$
$$b_n = \inf\{1/k : k \geq n/2\} \cup \{1\} = 0.$$

Notice that not only are the sequences (a_n) and (b_n) different, they don't necessarily have the same limit. But their limits tell us about the limit of our original sequence (x_n).

Definition 3.10. Let (x_n) be a sequence of real numbers. We define the *limit inferior* of the sequence to be

$$\liminf_{n\to\infty} x_n = \underline{\lim}_{n\to\infty} x_n = \lim_{n\to\infty} b_n.$$

We define the *limit superior* to be

$$\limsup_{n\to\infty} x_n = \overline{\lim}_{n\to\infty} x_n = \lim_{n\to\infty} a_n.$$

The basic idea here is that the lim inf is the smallest value that gets hit infinitely often, and the lim sup is the largest value that gets hit infinitely often. These don't always exist—what if the sequence increases without bound?—but they almost always exist.

Proposition 3.11. *Let (x_n) be a bounded sequence. Then*

1. $\liminf x_n$ *and* $\limsup x_n$ *both exist.*
2. $\liminf_{n\to\infty} x_n = \sup_{n\in\mathbb{N}} \left\{ \inf_{k\geq n} x_k \right\}$ *and* $\limsup_{n\to\infty} x_n = \inf_{n\in\mathbb{N}} \left\{ \sup_{k\geq n} x_k \right\}$.
3. $\liminf x_n \leq \limsup x_n$.

Proof. 1. (a_n) is a bounded decreasing sequence, and (b_n) is a bounded increasing sequence. By the Monotone Convergence Theorem, both limits exist.

2. $\liminf_{n\to\infty} x_n = \lim_{n\to\infty} b_n$ by definition. But (b_n) is monotone increasing, so this limit is $\sup(b_n)$. Since $b_n = \inf\{x_k : k \geq n\}$, the proposition follows.

3. For each n, we have $b_n \leq a_n$ since a_n is the supremum of a set and b_n is the infimum of that same set. (So, e.g., $b_n \leq x_n \leq a_n$). Thus by HW 5 problem 1, we have $\lim_{n\to\infty} b_n \leq \lim_{n\to\infty} a_n$.

□

Example 3.12. Let

$$x_n = \begin{cases} \frac{n+1}{n} & n \text{ odd} \\ 0 & n\text{even} \end{cases}$$

We have $b_n = \inf\{x_k : k \geq n\} = 0$, so $\liminf x_n = \lim_{n\to\infty} b_n = 0$.

We have $a_n = \sup\{x_k : k \geq n\} = \frac{n+1}{n}$ if n is odd, and $a_n = \frac{n+2}{n+1}$ if n is even. Then we claim $\limsup x_n = \lim_{n\to\infty} a_n = 1$. For $|x_n - 1| = \frac{1}{n}$ if n is odd, and $\frac{1}{n+1}$ if n is even.

If $\epsilon > 0$, we can take N so that $N > 1/\epsilon$ by the Archimedean property. Then if $n > N$, we have $|x_n - x| < \frac{1}{n} < \frac{1}{N} < \epsilon$.

The lim sup and lim inf are not actually limits of the sequence. But in many ways they behave the same. They follow the same limit laws that we proved for sequences in proposition 3.1. And they are *almost* limits of the sequence.

Proposition 3.13. *Let (x_n) be a bounded sequence. Then there exists a subsequence (x_{n_k}) such that $\lim_{k\to\infty} x_{n_k} = \limsup_{n\to\infty} x_n$.*

Similarly, there exists a subsequence (x_{m_k}) such that $\lim_{k\to\infty} x_{m_k} = \liminf_{n\to\infty} x_n$.

Proof. We have $a_n = \sup\{x_k : k \geq n\}$ and set $x = \limsup x_n = \lim a_n$. We define x_{n_k} as follows:

Let $n_1 = 1$, that is, $x_{n_1} = x_1$. Define the rest of the sequence inductively: if we have defined x_{n_k}, we can choose some $m > n_k$ such that $|a_{n_k+1} - x_m| < \frac{1}{k+1}$, since $a_{n_k+1} = \sup\{x_m : m \geq n_{k+1}\}$. We define $n_{k+1} = m$ so that $x_{n_{k+1}} = n_m$.

We want to show that $\lim_{k\to\infty} x_{n_k} = x$. We first want to show that x_{n_k} gets close to a_{n_k}, and then we can use the fact that a_{n_k} gets close to x.

We know that $a_{n_k+1} \geq a_{n_k}$ since they're both supremums, and the first set contains the second. We also know that $a_{n_k} \geq x_{n_k}$, again since a_{n_k} is a supremum. Then we can calculate

$$\begin{aligned} |a_{n_k} - x_{n_k}| &= a_{n_k} - x_{n_k} \\ &\leq a_{n_{k-1}+1} - x_{n_k} < \frac{1}{k}. \end{aligned}$$

Now let $\epsilon > 0$. We know $\lim_{n\to\infty} a_n = x$, so $\lim_{k\to\infty} a_{n_k} = x$, and thus there is some N_1 such that if $k > N_1$ then $|a_{n_k} - x| < \epsilon/2$. There is also some N_2 so that $1/N_2 < \epsilon/2$, and thus if $k > N_2$ then $|a_{n_k} - x_{n_k}| < 1/k < 1/N_2 < \epsilon/2$.

Then if $k > \max\{N_1, N_2\}$ we have

$$\begin{aligned} |x - x_{n_k}| &= |x - a_{n_k} + a_{n_k} - x_{n_k}| \\ &\leq |x - a_{n_k}| + |a_{n_k} - x_{n_k}| \\ &< \epsilon/2 + \epsilon/2 = \epsilon. \end{aligned}$$

Thus we've constructed a subsequence (x_{n_k}) whose limit is $x = \limsup x_n$.

□

We want to show that the limit superior and limit inferior tell us about the limit of our sequence if it exists. But before we can do that, we need one more result that should be familiar from Calculus 1.

Exercise 3.14 (Squeeze Theorem)**.** *Let $(a_n), (b_n), (x_n)$ be sequences of real numbers such that $a_n \leq x_n \leq b_n$ for all $n \in \mathbb{N}$. Suppose (a_n) and (b_n) converge, and $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$. Then (x_n) converges, and $\lim_{n\to\infty} x_n = \lim_{n\to\infty} a_n$.*

Remark 3.15. Notice that this is a bit stronger than just showing that if all three sequences converge, then the limit of x_n is between the limits of a_n and b_n. In particular, you have to show that if (x_n) is trapped between a_n and b_n then it has to approach a particular limit.

Proposition 3.16. *Let (x_n) be a bounded sequence of real numbers. Then (x_n) converges if and only if $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n$. If (x_n) converges, then the limit is equal to the lim sup.*

Proof. Suppose $\lim_{n\to\infty} x_n = x$. Then by proposition 3.13 there is a subsequence (x_{n_k}) that converges to $\liminf_{n\to\infty} x_n$. But we know that $\lim_{k\to\infty} x_{n_k} = \lim_{n\to\infty} x_n$ if the latter limit exists, so $\lim_{n\to\infty} x_n = \liminf_{n\to\infty} x_n$. Similarly, $\lim_{n\to\infty} x_n = \limsup_{n\to\infty} x_n$, and thus the lim sup and lim inf are equal.

Conversely, suppose $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n$. Then we have $b_n \leq x_n \leq a_n$ for each n, and we have

$$\begin{aligned} \lim_{n\to\infty} b_n &= \liminf_{n\to\infty} x_n \\ &= \limsup_{n\to\infty} x_n = \lim_{n\to\infty} a_n. \end{aligned}$$

Thus by the Squeeze Theorem 3.14, we know that (x_n) converges, and

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} a_n = \limsup_{n\to\infty} x_n$$

□

The ability to make arguments using the least upper bound property and thee lim sup and lim inf in the real numbers is really important and useful. Over the next couple of sections we want to find ways to generalize these properties to other metric spaces.

3.2 Completeness

In this section we're working towards the idea of *completeness*, which generalizes the Least Upper Bound property. Recall the Least Upper Bound property tells us the real numbers don't have any "holes"—that everywhere we would hope to find a real number, there is one actually there.

To generalize this to metric spaces, we first have to ask where we *want* to find points. We can't use the same trick we used in the reals, because we don't have an order. Instead we do something like generalizing the Monotone Convergence Theorem: we want to see that every sequence that should converge does.

So what kind of sequences should converge?

Definition 3.17. Let (E, d) be a metric space, and let (x_n) be a sequence in E. We say that (x_n) is a *Cauchy sequence* if for every $\epsilon > 0$ there is a $N \in \mathbb{N}$ so that if $n, m > N$, then $d(x_n, x_m) < \epsilon$.

This definition looks similar to the definition of convergence, but is subtly and importantly different. A sequence converges if the values all get arbitrarily close to some limit point. A sequence is Cauchy if the values all get arbitrarily close to *each other*.

The basic idea here is that if the points of a sequence are getting arbitrarily close together, they should be gathering at one point—and we would like that point to be a limit.

Example 3.18. We claim the sequence $(1/n)$ is Cauchy under the absolute value metric. Let $\epsilon > 0$, and choose N so that $1/N < \epsilon/2$. Then if $m, n < N$, we have $|1/m - 1/n| \leq |1/m| + |1/n| < \frac{1}{N} + \frac{1}{N} < \epsilon$. Thus $(1/n)$ is Cauchy.

Notice what information we *didn't* use: we needed to know the metric, but we didn't need to know the space. We probably assumed that $(1/n)$ was a sequence in the real numbers,

in which case it is convergent. If we think of it as a sequence of rational numbers, it also converges.

But if we think of it as a sequence in $(0,1)$, it doesn't converge—because there's nothing there for it to converge to! But even though we can't say the sequence converges, we can still say it's Cauchy. Because the property of being Cauchy is internal to the sequence.

Example 3.19. We claim the sequence $0, 1, 0, 1, \dots$ is not Cauchy. Let $\epsilon = 1$ and let $N \in \mathbb{N}$. Then there is some even $m > N$ so that $x_m = 1$, and there is some odd $n > N$ so that $x_n = 0$. Then $|x_m - x_n| = 1 \not< 1$.

Example 3.20. Consider the sequence $(x^n) \subset \mathcal{C}([0,1], \mathbb{R})$ of continuous functions on the unit interval, under the L_1 metric. Is this Cauchy?

Let $\epsilon > 0$ and choose $N > 2/\epsilon$. Then if $m, n > N$, we have

$$\begin{aligned} d(x^m, x^n) &= \int_0^1 |x^m - x^n| \, dx \leq \int_0^1 |x^m| + |x^n| \, dx = \int_0^1 x^m + x^n \, dx \\ &= \frac{x^{m+1}}{m+1} + \frac{x^{n+1}}{n+1}\bigg|_0^1 = \frac{1}{m+1} + \frac{1}{n+1} < \frac{1}{N} + \frac{1}{N} < \epsilon. \end{aligned}$$

Thus this sequence is Cauchy.

Does this sequence converge? It does, in fact—just weirdly. We claim the limit is 0. Let $\epsilon > 0$ and let $N > 1/\epsilon$. Then if $n > N$, we have

$$\begin{aligned} d(x^n, 0) &= \int_0^1 |x^n - 0| \, dx = \int_0^1 x^n \, dx \\ &= \frac{x^{n+1}}{n+1}\bigg|_0^1 = \frac{1}{n+1} \\ &< \frac{1}{N} < \epsilon. \end{aligned}$$

Thus $\lim_{n\to\infty} x^n = 0$.

(The weird part is: $1^n = 1$ for any n, but $0 = 0$, so the limit of the values at 1 is not the value of the limit at 1. We'll discuss this weirdness more later on in the course.)

Proposition 3.21. *Every convergent sequence is Cauchy.*

Proof. Suppose $\lim_{n\to\infty} x_n = x$. Let $\epsilon > 0$. Then there is some $N \in \mathbb{N}$ so that if $n > N$, then $d(x_n, x) < \epsilon/2$.

Now suppose $m, n > N$. Then $d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) < \epsilon/2 + \epsilon/2 = \epsilon$. Thus (x_n) is Cauchy. □

Exercise 3.22. *Prove that any Cauchy sequence is bounded.*

Exercise 3.23. *Let (x_n) be a Cauchy sequence. Then any subsequence of (x_n) is Cauchy.*

Proposition 3.24. *Let (x_n) be a Cauchy sequence, and suppose (x_{n_k}) is a convergent subsequence. Then (x_n) converges.*

Proof. Suppose $\lim_{k\to\infty} x_{n_k} = x$. Let $\epsilon > 0$. Then there is some $N_1 \in \mathbb{N}$ so that if $k > N_1$, then $d(x_{n_k}, x) < \epsilon/2$. And there is some $N_2 \in \mathbb{N}$ so that f $n, m > n$, then $d(x_n, x_m) < \epsilon/2$.

Let $N = \max\{N_1, N_2\}$, and let $k > N$. Then we know that $n_k \geq k > N$, so we know that $d(x_{n_k}, x) < \epsilon/2$. And we also have $k, n_k > N$, so $d(x_k, x_{n_k}) < \epsilon/2$. Thus $\epsilon > d(x_{n_k}, x) + d(x_k, x_{n_k}) \geq d(x, x_k)$ by the triangle inequality, and thus $\lim_{k\to\infty} x_k = x$. □

Definition 3.25. We say a metric space (E, d) is *complete* if every Cauchy sequence converges.

Example 3.26. $\mathbb{R}$ is complete under the absolute value metric.

Suppose (x_n) is a Cauchy sequence. Then (x_n) is bounded, and thus has an upper bound, and thus has a lim sup. Let $x = \limsup x_n$. We claim that $\lim_{n\to\infty} x_n = x$.

Let $\epsilon > 0$. By definition of Cauchy, there is a $N_1 \in \mathbb{N}$ so that $|x_m - x_n| < \epsilon/3$ whenever $m, n > N_1$.

Now consider the sequence $(a_n) = (\sup\{x_k : k \geq n\})$. We know that $\lim_{n\to\infty} a_n = \limsup x_n = x$, and thus there is a N_2 so that if $n > N_2$, then $|a_n - x| < \epsilon/3$.

Now let $N > N_1, N_2$, and suppose $n > N$. Then we know that $|a_n - x| < \epsilon/3$. Since $a_n = \sup\{x_k : k \geq n\}$, there is some $k \geq n$ such that $|a_n - x_k| < \epsilon/3$. And since $k > N$, we know that $|x_n - x_k| < \epsilon/3$. Adding these three inequalities together gives us

$$\epsilon > |x_n - x_k| + |x_k - a_n| + |a_n - x| \geq |x_n - x|$$

by the triangle inequality, and thus $\lim_{n\to\infty} x_n = x$.

Thus any Cauchy sequence converges, and so $\mathbb{R}$ is complete.

Remark 3.27. Rosenlicht has a slightly different proof that doesn't involve the work we did defining lim sup. It's substantially the same, though, just taking the upper bound of the set of lower bounds—which is, of course, the lim inf.

Example 3.28. The rational numbers $\mathbb{Q}$ are not complete. For instance, the sequence $3, 3.1, 3.14, 3.141, \ldots$ is Cauchy, since if $n, m > N$, they are the same to $N-1$ decimal places and thus $|x_n - x_m| < 10^{1-N}$. But it doesn't converge in $\mathbb{Q}$, since $\pi \notin \mathbb{Q}$.

Similarly, we saw in example 3.18 that the set $(0, 1)$ is not complete: the sequence $(1/n)$ is Cauchy, but it does not converge in $(0, 1)$.

Exercise 3.29. *If E is any metric space under the discrete metric, then it is complete.*

Proposition 3.30. *A closed subset of a complete metric space is complete.*

Proof. Let (E, d) be a complete metric space, and V a closed subset of E.

Suppose (x_n) is a Cauchy sequence in V. Then by definition of completeness, x_n converges in E, and thus there is an $x \in E$ so that $\lim_{n\to\infty} x_n = x$. But then x is the limit of a sequence in V, so $x \in V$ and thus x_n converges in V. □

Example 3.31. Any closed interval in $\mathbb{R}$ is complete.

In contrast, we saw that $(0, 1)$ was not complete in example 3.18. In fact, no non-trivial open interval will be complete.

There are many metric spaces that are complete; in general we prefer to only work in complete metric spaces. The complex numbers $\mathbb{C}$ are complete, as are the various spaces of funcitons L_1, L_2, L_∞ that we have discussed at various points (as long as you define them carefully).

But one more complete metric space will be a very important example in this course, so we write out the proof carefully:

Proposition 3.32. $\mathbb{R}^n$ *is complete for any* $n \in \mathbb{N}$.

We'll prove this under the sup metric, but the proof is similar for the sum or Euclidean metrics. In fact, it's not too hard to show that a sequence converges under the sup metric if and only if it converges under the Euclidean metric, if and only if it converges under the sum metric.

The hardest part of this proof is the notation. We'll prove the result in $\mathbb{R}^3$ to keep the notation simple and avoid having to write double subscripts on everything, but the proof is identical for the general case in $\mathbb{R}^n$.

Proof. Let $(p_n) = ((x_n, y_n, z_n))$ be a Cauchy sequence in $\mathbb{R}^3$. We claim that the three sequences of real numbers $(x_n), (y_n), (z_n)$ are all Cauchy, and then we will use the fact that $\mathbb{R}$ is complete to show that $\mathbb{R}^3$ is complete.

Let $\epsilon > 0$. Then there is a $N \in \mathbb{N}$ so that $d(p_n, p_m) < \epsilon$ if $n, m > N$. But $d(p_n, p_m) = \sup\{|x_n - x_m|, |y_n - y_m|, |z_n - z_m|\}$, and thus we have $|x_n - x_m|, |y_n - y_m|, |z_n - z_m| < \epsilon$. So by definition the sequences $(x_n), (y_n), (z_n)$ are all Cauchy sequences of real numbers.

Since $\mathbb{R}$ is complete, we know that there are $x, y, z \in \mathbb{R}$ so that $\lim_{n\to\infty} x_n = x, \lim_{n\to\infty} y_n = y, \lim_{n\to\infty} z_n = z$. Thus there is some $M_x, M_y, M_z \in \mathbb{N}$ so that if $n > M_x$ then $|x_n - x| < \epsilon$; we define M_y and M_z similarly.

Let $M = \max\{M_x, M_y, M_z\}$. Then if $n > M$ we have $|x_n - x| < \epsilon, |y_n - y| < \epsilon, |z_n - z| < \epsilon$, and thus

$$d(p_n, (x, y, z)) = \sup\{|x_n - x|, |y_n - y|, |z_n - z|\} < \epsilon$$

and thus $\lim_{n\to\infty} p_n = (x, y, z)$.

Thus any Cauchy sequence in $\mathbb{R}^3$ converges in $\mathbb{R}^3$, and so $\mathbb{R}^3$ is complete.

□

A final note on completeness: if (E, d) is a metric space, we define the *completion* of E to be the smallest metric space containing E that is complete. We can construct such a thing by taking the set of Cauchy sequences in E, and declaring two sequences equivalent if they go to "the same place"—which we can define by, say requiring that the sequence $x_1, y_1, x_2, y_2, \ldots$ still be Cauchy. Then we can define a metric on the set of equivalence classes of Cauchy sequences, and it is a complete metric space.

This is the third construction of the reals from the rationals: the reals are the completion of the rationals. So we can define the reals to be equivalence classes of Cauchy sequences of rationals. (In fact, this is almost what we did: we defined the reals to be the set of equivalence classes of infinite decimals, which is just a very specific type of Cauchy sequence).

3.3 Compactness

In this section we want to generalize the idea of lim sup and lim inf. Those operators are useful because they work on any bounded sequence: in $\mathbb{R}$, every bounded sequence "almost" has a limit. We want to see how far we can generalize that property.

Definition 3.33. Let (E, d) be a metric space, and $S \subset E$. We say that a point $x \in E$ is an *accumulation point* (or *cluster point*) of S if any open ball centered at x contains infinitely many points of S.

Example 3.34. 0 and 1 are accumulation points of $(0, 1)$. $(1/2$ is also an accumulation point of $(0, 1)$, in fact).

0 is an accumulation point of the set $\{1/n : n \in \mathbb{N}\}$.

Every real number is an accumulation point of $\mathbb{Q}$.

In $\mathbb{R}^2$ with the Euclidean metric, the set of accumulation points of $B_1(0, 0)$ is $\overline{B}_1(0, 0)$.

Proposition 3.35. *Let (E, d) be a metric space, and $S \subset E$. x is an accumulation point of S if and only if every open ball centered at x contains a point of $S \setminus \{x\}$.*

Proof. If x is an accumulation point of S, then every ball centered at x contains infinitely many points of S. Thus it contains at least two points, so it contains at least one point of $S \setminus \{x\}$.

Conversely, suppose every open ball centered at x contains a point of $S \setminus \{x\}$. Let $r > 0$, and suppose $B_r(x)$ contains only finitely many points (not including x) $y_1, \dots, y_n$. Then we can set $d = \min\{d(y_1, x), \dots, d(y_n, x)\}$. Then $B_d(x)$ contains no points of S (except possibly x), since otherwise we have a $y \in B_d(x) \subset B_r(x)$ with $d(y, x) < d$. But we know that every open ball contains some point in $S \setminus \{x\}$, so this is a contradiction. □

Proposition 3.36. *Let (E, d) be a metric space, and $S \subset E$. x is an accumulation point of S if and only if x is the limit of a sequence of points in $S \setminus \{x\}$.*

Proof. Suppose x is an accumulation point of S. We construct a sequence as follows: for each $n \in \mathbb{N}$, we know that $B_{1/n}(x)$ contains a point of $S \setminus \{x\}$, so take x_n to be such a point.

Then if $\epsilon > 0$ and $1/N < \epsilon$, if $n > N$ we know that $x_n \in B_{1/n}(x)$ so $d(x_n, x) < 1/n < 1/N < \epsilon$. Thus $\lim_{n\to\infty} x_n = x$. Thus x is the limit of a sequence in $S \setminus \{x\}$.

Conversely, suppose $x = \lim_{n\to\infty} x_n$ where $x_n \in S \setminus \{x\}$ for each n. Then for any $r > 0$, there is a $N \in \mathbb{N}$ so that $x_n \in B_r(x)$ for all $n > N$. Thus each open ball centered at x contains a point of $S \setminus \{x\}$, and thus x is an accumulation point for S. □

Exercise 3.37. *Let (E, d) be a metric space, and let $V \subset E$. Prove that V is closed if and only if it contains all its accumulation points.*

We're particularly interested in sequences here, so we want a similar definition for cluster points of sequences.

Definition 3.38. We say that x is an accumulation point of the sequence (x_n) every open ball centered at x contains x_n for infinitely many natural numbers n.

Example 3.39. 0 is an accumulation point for the sequence $(1/n)$.

The sequence $0, 1, 0, 1, 0, 1$ has two accumulation points: 0 and 1. (Notice that the set $\{0, 1, 0, 1, \dots\}$ is just the same as the set $\{0, 1\}$, and thus has no accumulation points at all as a set).

The sequence $0, 1, 2, 0, 1, 2$ has three accumulation points.

Proposition 3.40. *Let (x_n) be a sequence. x is an accumulation point of (x_n) if and only if it is the limit of some subsequence.*

Proof. Suppose x is the limit of a subsequence of x_n. Then there is some subsequence x_{n_k} so that $\lim_{k\to\infty} x_{n_k} = x$. Let $r > 0$; we need to show that $B_r(x)$ contains infinitely many x_n.

But by definition of convergence, there is some $N \in \mathbb{N}$ so that $d(x_{n_k}, x) < r$ for all $k > N$. Thus for every $k > N$, we know that$x_{n_k} \in B_r(x)$. There are infinitely many $k > N$, which proves the claim.

Conversely, suppose x is an accumulation point of (x_n). We need to construct a subsequence that approaches x. We do this just as we did in proposition 3.36; the details are left as an exercise.

□

Definition 3.41. We say a metric space (E, d) is *sequentially compact* if every sequence in E has an accumulation point in E.

Proposition 3.42. *A metric space is sequentially compact if and only if every infinite set has an accumulation point.*

Proof. Suppose (E, d) is a sequentially compact metric space, and $S \subset E$ is infinte. Let (x_n) be any sequence of distinct points in S. Then since (E, d) is sequentially compact, x_n has an accumulation point x. Let $r > 0$. Then there is some subsequence x_{n_k} and $N_r \in \mathbb{N}$ so that $x_{n_k} \in B_r(x)$ for all $k > N_r$, and thus B_r contains infinitely many points of S. Thus x is an accumulation point of S.

Conversely, suppose (E, d) is a metric space such that every infinite set has an accumulation point, and let (x_n) be a sequence. We need to show that (x_n) has an accumulation point.

We divide this into two cases. First, suppose the set $\{x_n\}$ is finite. Then there is at least one element x that appears infinitely many times in the sequence. We take the subsequence $x, x, x, \ldots$, and this subsequence has x for a limit. Thus x is an accumulation point of the sequence.

Now, suppose the set $\{x_n\}$ is infinite. Then the set has an accumulation point x. We define a subsequence converging to x as follows: for each k, we can find infinitely many points x_m so that $x_m \in B_{1/k}(x)$. Thus in particular we can find one such that $m > n_{k-1}$; we set $n_k = m$.

Now we see that $\lim_{k\to\infty} x_{n_k} = x$. If $\epsilon > 0$, we set $1/N < \epsilon$, and then if $k > N$ we have $d(x_{n_k}, x) < 1/k < 1/N < \epsilon$. Thus (x_n) has a convergent subsequence, and thus an accumulation point. □

Corollary 3.43. *A metric space is sequentialy compact if and only if every sequence has a convergent subsequence.*

Proof. Suppose E is sequentially compact. Then every sequence has an accumulation point, which is the limit of some subsequence by proposition 4.7.

Conversely, suppose every sequence has a convergent subsequence. Then the limit of that convergent subsequence is an accumulation point of the sequence. □

Proposition 3.44. *Let (E, d) be a metric space, and let $S \subset E$ be a (non-empty) sequentially compact subset. Then S is closed and bounded.*

Proof. First we show S is closed. Let (x_n) be a convergent sequence in S that converges to some point $x \in E$. We know that (x_n) has an accumulation point in S, which we call y. We want to show that $x = y$.

Let $\epsilon > 0$. Then since $x = \lim_{n\to\infty} x_n$, there is some $N \in \mathbb{N}$ so that $d(x, x_n) < \epsilon/2$ for all $n > N$. But y is an accumulation point for (x_n), so there are infinitely many m such that $d(x_m, y) < \epsilon/2$. Then we can choose some m that is also greater than N; then we have $d(x_m, y) < \epsilon/2$ and $d(x_m, x) < \epsilon/2$.

By the triangle inequality, we see that $d(x, y) < \epsilon$. But this holds for any $\epsilon > 0$, so $d(x, y) = 0$ and thus $x = y$. So if (x_n) converges in E, its limit must be in S; thus S is closed.

Now we show that S is bounded. Let $x \in S$. If S is not bounded, then for each n there is a $x_n \in S$ such that $d(x, x_n) > n$. Then (x_n) is a sequence in S, and so it has an accumulation point y.

Let $r = d(x, y)$, and let $n > 2r$. Then $d(x, x_n) \leq d(x, y) + d(y, x_n)$ by the triangle inequality; but $2r < n < d(x, x_n)$ and $d(x, y) = r$, so we have $2r < r + d(y, x_n)$ and thus $r < d(y, x_n)$. But then $d(y, x_n) > r$ for all $n > 2r$, and thus $B_r(y)$ contains only finitely many x_n, which contradicts the claim that y is an accumulation point. Thus we have a contradiction, and thus S must be bounded. □

Proposition 3.45. *Let (E, d) be a sequentially compact metric space, and let $S \subset E$ be closed. Then S is sequentially compact.*

Proof. Let (x_n) be a sequence in S. Then (x_n) has an accumulation point in E; let x be an accumulation point of (x_n). Then x is the limit of some subsequence (x_{n_k}). But (x_{n_k}) is a sequence in S, and thus $\lim_{k\to\infty} x_{n_k} = x \in S$. Therefore (x_n) has an accumulation point in S. □

We've now seen some properties that any compact set must have—they all have to be closed and bounded. The converse of this theorem isn't quite true. But it's close!

Exercise 3.46. *If (x_n) is a bounded sequence of real numbers, then $\liminf_{n\to\infty} x_n$ is an accumulation point of (x_n), and $\limsup_{n\to\infty} x_n$ is an accumulation point of (x_n).*

Lemma 3.47. *Let V be a closed and bounded subset of $\mathbb{R}$. Then V is sequentially compact.*

Proof. Let (x_n) be a sequence in V. Then (x_n) is bounded, and so $\limsup_{n\to\infty} x_n$ is an accumulation point of x_n. Since V is closed, then $\limsup_{n\to\infty} x_n \in V$. So (x_n) has an accumulation point in V. □

Theorem 3.48 (Bolzano-Weierstrass)**.** *If V is a closed and bounded subset of $\mathbb{R}^n$, then V is sequentially compact.*

Proof. For notational reasons we'll only write down the proof for $\mathbb{R}^3$.

Let V be a closed and bounded subset of $\mathbb{R}^3$, and let (x_n, y_n, z_n) be a sequence in V. Then (x_n) is a bounded sequence in $\mathbb{R}$, so it has a convergent subsequence (x_{n_k}).

Now consider the sequence y_{n_k} which is a subsequence of (y_n). This is a bounded sequence in $\mathbb{R}$, and so it has some convergent subsequence y_{m_k}. Similarly, we now consider the sequence z_{m_k}, and it is a bounded sequence in $\mathbb{R}$ and so has a convergent subsequence z_{ℓ_k}.

Now consider the sequence $(x_{\ell_k}, y_{\ell_k}, z_{\ell_k})$ in V. We know that x_{ℓ_k} is a convergent sequence (since it is a subsequence of a subsequence of a convergent sequence), so there is some $x \in \mathbb{R}$ such that $\lim_{k\to\infty} x_{\ell_k} = x$, and some N_x so that if $k > N_x$ then $|x_{\ell_k} - x| < \epsilon$. Similarly, there is a $y \in \mathbb{R}$ and N_y so that if $k > N_y$ then $|y_{\ell_k} - y| < \epsilon$, and a $z \in \mathbb{R}$ and N_z so that if $k > N_z$ then $|z_{\ell_k} - z| < \epsilon$.

Let $N = \max\{N_x, N_y, N_z\}$. Then if $k > N$, we have that

$$d_{\sup}((x_{\ell_k}, y_{\ell_k}, z_{\ell_k}), (x, y, z) = \sup\{|x_{\ell_k} - x|, |y_{\ell_k} - y|, |z_{\ell_k}, z|\} < \epsilon.$$

Thus (x_n, y_n, z_n) has a convergent subsequence. And since V is closed, the limit must be an element of V. Thus every sequence in V has a convergent subsequence, and so by definition V is sequentially compact. □

We've shown that every closed and bounded subset of $\mathbb{R}^n$ is compact. But this isn't true in any metric space:

Example 3.49. Let S be any infinite subset of a discrete metric space. Then S is bounded since $S \subset B_2(x)$ for any $x \in S$. And S is closed since any subset of a discrete metric space is closed. But S is not compact, since we can take a sequence of distinct points of S and this sequence will have no accumulation points.

This example, like most discrete metric space examples, is a little weird. But you can even find counterexamples in a reasonable (normed vector space) metric space.

Example 3.50. We define the metric space $\ell_\infty(\mathbb{R})$ to be the set of bounded sequences of elements of R, and define the sup metric by

$$d_{\sup}((x_n),(y_n)) = \sup\{|x_n - y_n|\}.$$

Let $V = \overline{B}_1((0))$ be the closed ball of radius 1 centered at the sequence of all zeroes. This set is clearly closed and bounded.

Now consider the sequence defined by $e_n = (0,0,0,\dots,0,1,0,\dots)$ to be the sequence whose nth element is 1, and all of whose other elements are 0. This is a sequence in V. But for all distinct $m,n \in \mathbb{N}$, we see that $d_{\sup}(e_m,e_n) = 1$, so there is no x such that $d_{\sup}(e_m,x) < 1/2$ and also $d_{\sup}(e_n,x) < 1/2$. Thus the sequence (e_n) has no accumulation point; so V is not compact.

In general, it's surprisingly difficult to be compact in an infinite dimensional space. But it's totally possible; and a lot of work in functional analysis has to do with proving that some subset of a space of functions is compact. (This allows us to prove that differential equations have solutions, for instance).

There is one other way of thinking about compactness. It is easier to prove things with, but a bit harder to visualize.

Definition 3.51. Let (E,d) be a metric space and $S \subset E$. We say S is *(topologically) compact* if, whenever S is contained in union of open sets $S \subset \bigcup U$, then S is contained in the union of some finite collection of those open sets.

Example 3.52. A closed interval $[a,b] \subset \mathbb{R}$ is (topologically) compact.

Proposition 3.53. *Any topologically compact metric space is sequentially compact.*

Proof. Let (E,d) be topologically compact. By proposition 3.42, it's sufficient to prove that every infinite set has an accumulation point.

So suppose S is a subset of E with no accumulation point. Then for each $x \in E$, we can find some open ball $B_{r_x}(x)$ that contains only finitely many points of S.

Clearly, $E = \bigcup_{x\in E} B_{r_x}(x)$ since every point of E is in that union. By compactness, E is contained in some finite union of these balls; but each ball contains only finitely many points of S, so the finite union contains only finitely many points of S. But $S \subset E \subset \bigcup B_{r_x}(x)$, so S must be finite.

□

Remark 3.54. The converse of this theorem is also true: any sequentially compact space is topologically compact. This is a bit more irritating to prove, though. The basic idea is that if your space is not compact, then you can find a bad open cover, build a sequence that has only finitely many points in any of the open sets. But this sequence consequently has no accumulation point.

4 Continuous Functions

4.1 Limits of Functions

Definition 4.1. Let E, F be metric spaces, and let $f : E \to F$ be a function between them. Suppose a is a cluster point of E. Then we write $\lim_{x\to a} f(x) = b$, and say that b is the limit of f at a, if for every $\epsilon > 0$, there is a $\delta > 0$ such that if $0 < d(x, a) < \delta$ then $d(f(x), b) < \epsilon$.

Example 4.2. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = 3x$. Then we claim $\lim_{x\to 2} f(x) = 6$.

Let $\epsilon > 0$ and let $\delta = \epsilon/3$. If $0 < d(x, 2) < \delta$, then $d(3x, 6) = |3x-6| = 3|x-2| < 3\delta = \epsilon$.

Example 4.3. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$. Then we claim $\lim_{x\to 1} f(x) = 1$.

Let $\epsilon > 0$ and let $\delta \leq \underline{1, \epsilon/3}$. If $d(x, 1) < \delta$ then

$$\begin{aligned} d(x^2, 1) = |x^2 - 1| &= |x - 1| \cdot |x + 1| < \delta|x + 1| \\ &= \delta|x - 1 + 2| \leq \delta(|x - 1| + 2) < \delta(\delta + 2) \\ &\leq \delta(1 + 2) \leq 3(\epsilon/3) = \epsilon. \end{aligned}$$

Example 4.4. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) = \frac{x^2 y}{x^2+y^2}$. We claim that $\lim_{(x,y)\to(0,0)} f(x, y) = 0$.

Let $\epsilon > 0$. If $0 < d((x, y), (0, 0)) < \delta$, $x^2 + y^2 \neq 0$ so the function is defined. Then

$$\begin{aligned} d\left(\frac{x^2 y}{x^2 + y^2}, 0\right) = \frac{|x^2 y|}{x^2 + y^2} &= |y| \frac{x^2}{x^2 + y^2} \\ &\leq |y| \leq d((x, y), (0, 0)) < \delta. \end{aligned}$$

So if we take $\delta = \epsilon$, then $d((x, y), (0, 0)) < \delta$ implies that $|f(x, y) - 0| < \epsilon$.

The definition of a limit of a function is obviously similar to the definition of a limit of a sequence, but not identical. We'd like to avoid having to take the results we already proved about sequences, and prove them all over again for functions. Fortunately, we can prove one result relating sequences to functions, and then use it to bring over all our old sequence results.

Lemma 4.5. *Let $f : E \to F$, and let a be a cluster point of E. Then $\lim_{x\to a} f(x) = b$ if and only if, whenever (x_n) is a sequence that converges to a, then the sequence $f(x_n)$ converges to b.*

Proof. Suppose $\lim_{x\to a} f(x) = b$, and suppose (x_n) is a sequence in E such that $\lim_{n\to\infty} x_n = a$. Let $\epsilon > 0$. Then there is a $\delta > 0$ so that if $0 < d(x, a) < \delta$ then $d(f(x), b) < \epsilon$.

Since $\lim_{n\to\infty} x_n = a$, there is a $N \in \mathbb{N}$ so that if $n > N$ then $d(x_n, a) < \delta$, and thus $d(f(x_n), b) < \epsilon$. Thus $\lim_{n\to\infty} f(x_n) = b$.

Conversely, suppose that $\lim_{x\to a} f(x) \neq b$. Then this means there is some $\epsilon > 0$ such that, for any $\delta > 0$, there is some x such that $0 < d(x, a) < \delta$ but $d(f(x), b) \geq \epsilon$.

In particular, for each $n \in \mathbb{N}$ there is some x_n such that $0 < d(x_n, a) < 1/n$ but $d(f(x), b) \geq \epsilon$. By construction we see that $\lim_{n\to\infty} x_n = a$. But $d(f(x_n), b) \geq \epsilon$ for every n, so the sequence $(f(x_n))$ does not converge to b. This proves the claim.

□

Example 4.6. We claim $\lim_{x\to 0} \sin(1/x)$ does not exist.

Suppose that the limit does exist, and set $L = \lim_{x\to 0} \sin(1/x)$. First consider the sequence given by $x_n = \frac{1}{\pi/2+2n\pi}$. We see that $\lim_{n\to\infty} x_n = 0$, so $\lim_{n\to\infty} \sin(1/x_n) = L$. But $\sin(1/x_n) = \sin(\pi/2 + 2n\pi) = 1$, so $L = \lim_{n\to\infty} 1 = 1$.

Now consider the sequence $y_n = \frac{1}{3\pi/2+2n\pi}$. Then $\lim_{n\to\infty} y_n = 0$ and so

$$L = \lim_{n\to infty} \sin(1/y_n) = \lim_{n\to\infty} \sin(3\pi/2 + 2n\pi) = \lim_{n\to\infty} -1 = -1,$$

which is a contradiction.

Alternately, we could have argued as follows: Let $x_n = \frac{2}{n\pi}$. Then $\lim_{n\to\infty} x_n = 0$, so

$$L = \lim_{x\to 0} \sin(1/x) = \lim_{n\to\infty} \sin(1/x_n) = \lim_{n\to\infty} \sin(n\pi/2).$$

But $(\sin(n\pi/2)) = 1, 0, -1, 0, 1, \ldots$ has no limit, which is a contradiction.

Notice that this is essentially the same argument you would have seen in calculus 1 to prove that this limit does not exist. But we can make the argument much more simply by using the language of sequences.

We wish to pay some special attention to limits in the reals. (As with sequences, much of this work can be generalized to $\mathbb{R}^n$, but we won't do so here).

Proposition 4.7. *Let (E, d) be a metric space, and suppose f, g are functions from E to $\mathbb{R}$. If $\lim_{x\to a} f(x) = L_1$ and $\lim_{x\to a} g(x) = L_2$, then*

- $\lim_{x\to a} f(x) + g(x) = L_1 + L_2$
- $\lim_{x\to a} f(x) - g(x) = L_1 - L_2$
- $\lim_{x\to a} f(x)g(x) = L_1L_2$
- *If $L_2 \neq 0$ then $\lim_{x\to a} f(x)/g(x) = L_1/L_2$.*

Proof. We could prove these directly, as we did with proposition 3.1 about limits of real sequences. But we can also just combine our result related sequence convergence to function limits with proposition 3.1 and avoid having to do any actual work.

Whenever (x_n) is a sequence converging to a, we know that $\lim_{n\to\infty} f(x_n) = \lim_{x\to a} f(x) = L_2$ and $\lim_{n\to\infty} g(x_n) = \lim_{x\to a} g(x) = L_2$. Thus

$$\lim_{n\to\infty} f(x_n) + g(x_n) = \lim_{n\to\infty} f(x_n) + \lim_{n\to\infty} g(x_n) = L_1 + L_2$$

by proposition 3.1.

So we showed that whenever (x_n) converges to a, then $\lim_{n\to\infty} f(x_n) + g(x_n) = L_1 + L_2$. This proves that $\lim_{x\to a} f(x) + g(x) = L_1 + L_2$ by Lemma 4.5.

This proves the first statement; the others follow similarly.

□

4.2 Continuity

We are now ready to define the most important type of function we'll be talking about for the rest of the course.

Definition 4.8. Let E, F be metric spaces, and let $f : E \to F$. Let $a \in E$ be an accumulation point of E. We say that f is *continuous at* a if $\lim_{x\to a} f(x) = f(a)$.

We say f is *continuous* (or continuous on its domain, or continuous on E) if f is continuous at every accumulation point of E.

Example 4.9. Any rational function is continuous on its domain, by Proposition 4.7.

Example 4.10. Let $f(x) = \begin{cases} x^3 & x < 0 \\ x^2 & x \geq 0 \end{cases}$. We claim $f(x)$ is continuous.

For $a < 0$, we know that $\lim_{x\to a} f(x) = \lim_{x\to a} x^3 = a^3 = f(a)$; for $a > 0$ we know that $\lim_{x\to a} f(x) = \lim_{x\to a} x^2 = a^2 = f(a)$.

So we just need to show that $\lim_{x\to 0} f(x) = f(0) = 0$. Let $\epsilon > 0$. We can do this explicitly or more abstractly.

Explicit argument: If $\delta < \sqrt{\epsilon}$ and $d(x, 0) < \delta$ then we have $d(x^2, 0) = x^2 < \delta^2 < \epsilon$. And if $\delta < \sqrt[3]{\epsilon}$ then $d(x^2, 0) = |x^3| < \delta^3 < \epsilon$. So set $\delta < \sqrt{\epsilon}, \sqrt[3]{\epsilon}$. Then if $d(x, 0) < \delta$, then $d(f(x), 0) < \epsilon$ since either $f(x) = x^2$ or $f(x) = x^3$. This proves that $\lim_{x\to 0} f(x) = 0$.

Abstract argument: We know that $\lim_{x\to 0} x^3 = 0^3 = 0$, so there is a δ_1 so that if $d(x, 0) < \delta_1$ then $d(x^3, 0) < \epsilon$. And we know that $\lim_{x\to 0} x^2 = 0^2 = 0$, so there is a δ_2 so that if $d(x, 0) < \delta_2$ then $d(x^2, 0) < \epsilon$.

Let $\delta = \min\{\delta_1, \delta_2\}$. If $d(x,0) < \delta$ then $d(x^2,0) < \epsilon$ and $d(x^3,0) < \epsilon$, and thus $d(f(x),0) < \epsilon$. This proves that $\lim_{x\to 0} f(x) = 0$.

Example 4.11. Define a function $\chi_{\mathbb{Q}} : \mathbb{R} \to \mathbb{R}$ by $\chi_{\mathbb{Q}}(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$. We claim that f is discontinuous at every real number.

First, suppose $a \notin \mathbb{Q}$. Let $\epsilon = 1$. For any $\delta > 0$, we can find a rational number x such that $d(x,a) < \delta$. In particular, there is a $N \in \mathbb{N}$ such that $1/N < \delta$, and there is a $n \in \mathbb{N}$ so that $n/N < a < (n+1)/N$. Then $d(n/N, a) < 1/N < \delta$.

Then for any $\delta > 0$, there is a x with $d(x,a) < \delta$ but $d(f(x), f(a)) = d(1,0) = 1 \not< \epsilon$. Thus $\lim_{x\to a} f(x) \neq f(a)$ and thus f is not continuous at a.

Now suppose $a \in \mathbb{Q}$. Let $\epsilon = 1$. For every $\delta > 0$, we claim there is an irrational number x such that $d(x,a) < \delta$: Let b be any positive rational number. Then if $N > b/\delta$ we have $b/N < \delta$ and $a + b/N$ is irrational. Let $x = a + b/N$. Then $d(x,a) < \delta$, but $f(x) = 0$ and $f(a) = 1$ so $d(f(x), f(a)) = 1 \not< \epsilon$. Thus $\lim_{x\to a} f(x) \neq f(a)$ and f is not continuous at a.

Exercise 4.12. *If $f : E \to F$ is a continuous function and (x_n) is a convergent sequence in E such that $\lim_{n\to\infty} x_n = x$, then prove that $\lim_{n\to\infty} f(x_n) = f(x)$.*

We can rephrase the definition of continuity in terms of open balls or open sets. Recall the definition of inverse image:

Definition 4.13. Let $f : E \to F$ be a function, and let $U \subset F$. We define $f^{-1}(U) = \{x \in E : f(x) \in U\}$.

This definition makes sense even if f is not invertible.

Example 4.14. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $x \mapsto x^2$. Then

$$f^{-1}(\{0\}) = \{0\} \qquad f^{-1}(\{1\}) = \{1, -1\}$$
$$f^{-1}(\{-1\}) = \varnothing \qquad f^{-1}(-2, -1) = \varnothing$$
$$f^{-1}([0,1]) = [-1,1] \qquad f^{-1}((0,+\infty)) = \mathbb{R} \setminus \{0\}.$$

Proposition 4.15. *Let $f : E \to F$ be a function of metric spaces. f is continuous (at every point in E) if and only if, whenever $U \subset F$ is open, then $f^{-1}(U) \subset E$ is also open.*

Proof. Suppose f is continuous, and suppose $U \subset F$ is open. We need to show that $f^{-1}(U)$ is open. So let $a \in f^{-1}(U)$. We want to find an open ball containing a that is contained in $f^{-1}(U)$.

Then $f(a) \in U$, and U is open, so there is a ϵ such that $B_\epsilon(f(a)) \subset U$. But f is continuous, so $\lim_{x\to a} f(x) = f(a)$. Thus there is a $\delta > 0$ so that whenever $d(x,a) < \delta$, then $d(f(x), f(a)) < \epsilon$.

So we claim that $B_\delta(a) \subset f^{-1}(U)$. If $x \in B_\delta(a)$, then $d(x,a) < \delta$, so $d(f(x), f(a)) < \epsilon$, and thus $f(x) \in B_\epsilon(f(a)) \subset U$. But since $f(x) \in U$ we see that $x \in f^{-1}(U)$ by definition. Thus $B_\delta(a) \subset f^{-1}(U)$. We can find such a $B_\delta(a)$ for any $a \in f^{-1}(U)$, so we see that $f^{-1}(U)$ is open.

Conversely, suppose we know that $f^{-1}(U)$ is open for any open U. Let $a \in E$; we want to show that $\lim_{x\to a} f(x) = f(a)$.

Let $\epsilon > 0$. Then the ball $B_\epsilon(f(a))$ is open in F, so $f^{-1}(B_\epsilon(f(a)))$ is open in E. Clearly $a \in f^{-1}(B_\epsilon(f(a)))$ since $f(a) \in B_\epsilon(f(a))$, and thus there is a $\delta > 0$ such that $B_\delta(a) \subset f^{-1}(B_\epsilon(f(a)))$.

Then if $d(x,a) < \delta$, we have $x \in B_\delta(a) \subset f^{-1}(B_\epsilon(f(a)))$ and thus $f(x) \in B_\epsilon(f(a))$. So $d(f(x), f(a)) < \epsilon$. Thus $\lim_{x\to a} f(x) = f(a)$, and f is continuous at a. □

Corollary 4.16. *If $f : E \to \mathbb{R}$ is continuous and $a \in \mathbb{R}$, then the sets*

$$\{x : f(x) > a\} \qquad \{x : f(x) < a\}$$

are open.

Exercise 4.17. *If $f : E \to F$ and $g : F \to G$ are continuous functions of metric spaces, prove that $g \circ f$ is continuous. (Hint: use the topological result about open sets, not the limit definition).*

We're particularly interested in functions involving the real numbers. We can make two easy observations here.

Exercise 4.18. *Let $p_i : \mathbb{R}^n \to \mathbb{R}$ be the projection into the ith coordinate given by $p_i(x_1, \dots, x_n) = x_i$. Prove that p_i is a continuous function.*

Proposition 4.19. *Let E be any metric space, and $f : E \to \mathbb{R}^n$. Then f is continuous at $a \in E$ if and only if $p_i \circ f$ is continuous at a for each i.*

Proof. If f is continuous, then $p_i \circ f$ is a composition of continuous functions, and thus continuous.

Conversely, suppose $p_i \circ f$ is a continuous function. Let $\epsilon > 0$. Then for each i, there is a δ_i so that if $d(x,a) < \delta_i$, then $|p_i(f(x)), p_i(f(a))| < \epsilon$.

Let $\delta = \min\{\delta_i\}$ (which is well-defined since n is finite). Then if $d(x,a) < \delta$, for each i we have $d(x,a) < \delta_i$ and thus $|p_i(f(x)), p_i(f(a))| < \epsilon$. Then

$$d(f(x), f(a)) = \sup\{|p_i(f(x)) - p_i(f(a))|\} < \epsilon.$$

Thus $\lim_{x\to a} f(x) = f(a)$. This is true for any $a \in E$, so f is continuous at a. □

Remark 4.20. This proof is pretty trivial in the sup metric. It's a bit trickier in the Euclidean or sum metrics, but essentially the same—you just have to use $\epsilon/\sqrt{n}$ or ϵ/n.

Example 4.21. The function $f : \mathbb{R}^3 \to \mathbb{R}$ given by $f(x,y,z) = x^3y + xz^2 - yz$ is continuous.

The function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x,y) = \frac{x^2}{x^2+y^2}$ is continuous everywhere except the origin (where it is not defined).

4.3 Compact Sets and the Extreme Value Theorem

We said that continuous functions behave well with respect to metric spaces. They also behave well with regard to specifically compact spaces.

Proposition 4.22. *Let $f : E \to F$ be a continuous function on metric spaces. If E is compact, then the image $f(E) = \{f(x) : x \in E\}$ is also compact.*

Proof. We'll prove this for sequential compactness.

Suppose $(f(x_n))$ is a sequence in $f(E)$. (Every sequence looks like this, since $f(E) = \{f(x) : x \in E\}$. Then x_n is a sequence in E, and since E is compact, it has a convergent subsequence (x_{n_k}), and there is some x such that $\lim_{k\to\infty} x_{n_k} = x$.

But since f is continuous, we know that $\lim_{n\to\infty} f(x_{n_k}) = f(x)$, and thus $(f(x_n))$ has a convergent subsequence. Since every sequence in $f(E)$ has a convergent subsequence, we know that $f(E)$ is sequentially compact. □

Alternate topological proof. We can also prove this result using the topological definition. I like this proof better, but it's a bit more abstract. (It's not actually any more complicated; but it looks more complicated because we've spent much less time working with these concepts).

Suppose $f(E) \subset \bigcup U_\alpha$. Then for each $x \in E$, we have $f(x) \in U_\alpha$ for some α, and thus $x \in \bigcup f^{-1}(U_\alpha)$. So we have $E \subset \bigcup f^{-1}(U_\alpha)$.

But since f is continuous, we know that $f^{-1}(U_\alpha)$ is open for each α. So we've written E as a subset of a union of open sets; thus there is some finite collection $f^{-1}(U_{\alpha_1}), \dots, f^{-1}(U_{\alpha_n})$ so that $E \subset f^{-1}(U_{\alpha_1}) \cup \dots \cup f^{-1}(U_{\alpha_n})$.

But then we have that for each $x \in E$, $f(x) \in U_{\alpha_i}$ for some i. Thus $f(E) \subset U_{\alpha_1} \cup \cdots \cup U_{\alpha_n}$. So whenever $f(E)$ is contained in a union of open sets, it is contained in the union of some finite collection of them, and thus $f(E)$ is compact.

□

This seems like a weird technical result, and to some extent it is. But it's a technical result that has a lot of easy and powerful implications.

Definition 4.23. Let $f : E \to F$ be a function where F is a metric space. We say f is *bounded* if the set $f(E)$ is bounded, that is, contained in some ball.

Corollary 4.24. *Let $f : E \to F$ be a continuous function of metric spaces. If E is compact, then f is a bounded function.*

We often want to require our functions to be bounded. This allows us to talk about their suprema, as we do in the metric space L_∞ which we will discuss soon. It also is a necessary ingredient to many proofs that integrals are well-defined and other similar results.

Corollary 4.25 (Extreme Value Theorem)**.** *Let $f : E \to \mathbb{R}$ be a continuous function, and E a non-empty compact metric space. Then f attains a maximum value at some point, and a minimum value at some point.*

Proof. $f(E)$ is a compact set, and thus closed and bounded. Since it is bounded, it has an infimum and a supremum; since it is closed, the infimum and supremum must be elements of $f(E)$. Thus there exist x_1, x_2 such that $f(x_1) = \sup f(E)$ and $f(x_2) = \inf f(E)$, and thus f attains a maximum and a minimum value. □

Remark 4.26. The compactness condition is really important here. For instance, $f(x) = x$ is continuous, but doesn't achieve a maximum on the set $(0, 1)$, which is bounded.

4.4 Connected sets and the Intermediate Value Theorem

In high school you may have heard the following "definition" of continuous functions: f is continuous if you can draw its graph without picking up your pen. This is a fairly inadequate definition for our purposes, both because it doesn't generalize well to higher dimensions and other metric spaces, and because it isn't terribly precise. However, we'd still like to prove that it is *true*.

To do this we need to introduce the idea of connectedness. Intuitively, we say a set is connected if it isn't made up of two independent pieces; we can't separate it without cutting or breaking or tearing something. WE can formalize that idea like this:

Definition 4.27. Let (E, d) be a metric space. We say that E is *connected* if it is not the disjoint union of two non-empty open sets.

Example 4.28. The metric space $(0, 1) \cup (2, 3)$ is not connected.

The metric space $[0, 1] \cup [2, 3]$ is not connected, because $[0, 1]$ and $[2, 3]$ are open in that metric space.

No discrete metric space with more than one point is connected, since any subset of the discrete metric space is open.

Lemma 4.29. *Any subset $S \subset \mathbb{R}$ that contains two distinct points a, b, and does not contain all of $[a, b]$, is not connected.*

Proof. Suppose $c \in (a, b)$ and $c \notin S$. Then we can take $T_1 = \{x \in S : x < c\}$ and $T_2 = \{x \in S, x > c\}$. Then T_1 and T_2 are disjoint open subsets of S, and $S = T_1 \cup T_2$. □

It's pretty difficult to prove a metric space is connected from this definition, but there's a slightly different way to look at it that is much easier to deal with.

Proposition 4.30. *A metric space (E, d) is disconnected if and only if it contains a non-trivial set (other than $\varnothing$ or E) that is both closed and open.*

Proof. If S is both closed and open, then S^C is also open, and thus $E = S \cup S^C$ is a disjoint union of open sets. Since $S \neq \varnothing$ it is nonempty, and since $S \neq E$ we know S^C is non-empty.

Conversely, suppose $E = S \cup T$ is a disjoint union of non-empty open sets. Then S is open, and since T is open S is also closed. Further, $S \neq \varnothing$ because it is non-empty, and $S \neq E$ since T is non-empty. □

Example 4.31. $\mathbb{R}$ is connected because no subset can be closed and open simultaneously. Similarly, $\mathbb{R}^n$ is connected.

Lemma 4.32. *Any closed interval in $\mathbb{R}$ is connected.*

Proof. Let $[a, b] \subset \mathbb{R}$, and suppose we have a non-empty set $S \subsetneq [a, b]$ that is both closed and open. We can assume without loss of generality that $b \notin S$, since if $b \in S$ we can consider the non-trivial closed and open set S^C instead.

Then S is closed and bounded, and thus has a maximum element $c < b$. But S is open, so it must contain some open ball centered at c, which will contain elements of $[a, b]$ which are larger than c, contradicting the maximality of c. □

Lemma 4.33. *Let $S \subset \mathbb{R}$ have the property that if $a, b \in S$, then $[a, b] \subset S$. Then S is connected.*

Proof. Suppose S is the disjoint union of two non-empty open sets A, B. Then there exist $a \in A, b \in B$, and we can assume without loss of generality that $a < b$. Then $A \cap [a, b]$ and $B \cap [a, b]$ are open in $[a, b]$, and thus $[a, b]$ is disconnected, contradicting the previous lemma. □

Corollary 4.34. *$\mathbb{R}$ is connected.*

Any open interval in $\mathbb{R}$ is connected.

Remark 4.35. We can use this to prove that if a metric space is *path-connected*—that is, any two points are connected by some parametrized curve—then the space is also connected. If we can write the space as a union of two disjoint open sets, then we can write the path as the union of two disjoint open sets, which then allows us to write the pre-image of the path as a union of two disjoint open sets. But the pre-image is a closed interval, so this is a contradiction.

This proves, among other things, that $\mathbb{R}^n$ is connected. It's also a bit outside the scope of this course.

Now let's apply our theory of continuous functions to connected sets. We know intuitively that continuous functions shouldn't break things apart, and thus they shouldn't turn connected sets into disconnected sets. Fortunately, this turns out to be correct.

Lemma 4.36. *Let $f : E \to F$ be a continuous function. If E is connected, then so is $f(E)$.*

Proof. Suppose we can write $f(E) = S \cup T$ where S, T are disjoint non-empty open sets. Then for every $x \in E$ we have either $f(x) \in S$ or $f(x) \in T$, but not both. So $E \subset f^{-1}(S) \cup f^{-1}(T)$ is a disjoint union.

Further, if $y \in S$ then there is some $x \in E$ such that $f(x) = y$, so $f^{-1}(S)$ is non-empty, and a similar argument shows that $f^{-1}(T)$ is non-empty. But f is continuous, so $f^{-1}(S)$ and $f^{-1}(T)$ are both open; and thus E is the disjoint union of non-empty open sets, and thus is disconnected, which is a contradiction. □

Corollary 4.37 (Intermediate Value Theorem)**.** *Let (E, d) be a metric space, and $f : E \to \mathbb{R}$ a continuous function. If $a, b \in E$ with $f(a) < f(b)$, and $y \in (f(a), f(b))$, then there is a $c \in E$ such that $f(c) = y$.*

Proof. Since f is continuous and E is connected, we know $f(E)$ is connected. Since $f(a), f(b) \in f(E)$, by lemma 4.29 we know that $[f(a), f(b)] \subset f(E)$. But then $y \in f(E)$ so there is some $c \in E$ such that $f(c) = y$. □

Corollary 4.38 (Intermediate Value Theorem for Real Functions)**.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. If y is between $f(a)$ and $f(b)$ then there is a $c \in (a, b)$ such that $f(c) = y$.*

We can get one more corollary out of this, which becomes very important in numerical analysis and other computational branches of math.

Definition 4.39. Let $f : E \to E$ be a function. We say that x is a *fixed point* of f if $f(x) = x$.

Exercise 4.40. *Suppose $f : [0, 1] \to [0, 1]$ is continuous. Show that f has a fixed point. That is, show there is an $x \in [0, 1]$ such that $f(x) = x$.*

5 Integrals of Real Functions

5.1 Riemann Sums

Definition 5.1. Let $a, b \in \mathbb{R}$ with $a < b$. We define a *partition* of the closed interval $[a, b]$ to be a finite set of numbers $\{x_0, \dots, x_n\}$ such that $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$. The *width* of a partition is

$$\max\{x_i - x_{i-1} : 1 \leq i \leq n\}.$$

Definition 5.2. If $f : [a, b] \to \mathbb{R}$, and $P = \{x_0, \dots, x_n\}$ is a partition of $[a, b]$, then we define a *Riemann sum for f corresponding to P* to be

$$\sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1})$$

where $x_{i-1} \leq x_i^* \leq x_i$ for each $1 \leq i \leq n$.

Definition 5.3. Let $f : [a, b] \to \mathbb{R}$. We say f is *Riemann Integrable on* $[a, b]$ if there is a number $I \in \mathbb{R}$ so that, for any $\epsilon > 0$, there is a $\delta > 0$ such that, if S is a Riemann sum corresponding to a partition of width less than δ, then $|S - I| < \epsilon$. In this case we say that I is the *Riemann Integral of f* and write $I = \int_a^b f(x)\, dx$.

Remark 5.4. If f is defined on some larger set containing $[a, b]$, this definition still works; we define $\int_a^b f(x)\, dx$ to be the integral of the restriction of f to the domain $[a, b]$.

The Riemann integral is not a limit of a sequence or a function; it is an entirely different type of limit. (Technically, all of these limits are generalized by the concept of a *topological net*, but we're not really going to go there). The big difference here is that there are lots of different partitions of any given width, and they're not in any particular "order" with respect to each other. For our purposes we can only ask that the width be small enough, and we have to ask that this is enough.

Exercise 5.5. *Prove that the Riemann integral is unique. That is, if $f : [a, b] \to \mathbb{R}$ and both I and J satisfy the definition of Riemann integral of f, then prove that $I = J$.*

Example 5.6. Let $f : [a, b] \to \mathbb{R}$ be defined by $f(x) = c$ for some fixed $c \in \mathbb{R}$. We claim that $\int_a^b f(x)\, dx = c(a - b)$.

Let $\epsilon > 0$. Then any Riemann sum for f is

$$\sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) = \sum_{i=1}^{n} c(x_i - x_{i-1}) = c\sum_{i=1}^{n}(x_i - x_{i-1}) = c(x_n - x_0) = c(b - a).$$

Thus we can take δ to be any positive real number, and for any partition of width less than δ we have $|S - c(b-a)| = 0 < \epsilon$ and thus $c(b-a) = \int_a^b f(x)\,dx$ by definition.

Exercise 5.7. *Let $c \in (a,b)$, and let $f : [a,b] \to \mathbb{R}$ be defined by $f(c) = 1$ and $f(x) = 0$ if $x \neq c$. Prove that $\int_a^b f(x)\,dx = 0$.*

Example 5.8. Let $c, d \in [a,b]$ with $a < c < d < b$. Define $f(x) = \chi_{(c,d)}(x) = \begin{cases} 1 & x \in (c,d) \\ 0 & x \notin (c,d) \end{cases}$.

We claim that $\int_a^b f(x)\,dx = d - c$.

Let $P = \{x_0, x_1, \dots, x_n\}$ be a partition of width less than δ. There exist r, s such that $x_r \leq c \leq x_{r+1}$ and $x_s \leq d \leq x_{s+1}$. Then for any x_i^* with $i \leq r$ or $i > s+1$, we have $f(x_i^*) = 0$, and for any x_i^* with $r+1 < i \leq s$ we have $f(x_i^*) = 1$.

So if S is a Riemann sum corresponding to P, we have

$$\begin{aligned}
S &= \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) \\
&= \sum_{i=1}^{r} f(x_i^*)(x_i - x_{i-1}) + f(x_{r+1}^*)(x_{r+1} - x_r) + \sum_{i=r+2}^{s} f(x_i^*)(x_i - x_{i-1}) \\
&\quad + f(x_{s+1}^*)(x_{s+1} - x_s) + \sum_{i=s+2}^{n} f(x_i^*)(x_i - x_{i-1}) \\
&= \sum_{i=1}^{r} 0(x_i - x_{i-1}) + f(x_{r+1}^*)(x_{r+1} - x_r) + \sum_{i=r+2}^{s} 1(x_i - x_{i-1}) \\
&\quad + f(x_{s+1}^*)(x_{s+1} - x_s) + \sum_{i=s+2}^{n} 0(x_i - x_{i-1}) \\
&= f(x_{r+1}^*(x_{r+1} - x_r) + f(x_{s+1}^*)(x_{s+1} - x_s) + (x_s - x_{r+1}).
\end{aligned}$$

Thus in particular, we see that $x_s - x_{r+1} \leq S \leq x_{s+1} - x_r$ and so, subtracting $d - c$ through the chain of inequalities, we get

$$x_s - d - (x_{r+1} - c) \leq S - (d - c) \leq x_{s+1} - d - (x_r - c).$$

Since the partition has width less than δ, we know that $d - x_s, x_{r+1} - c, x_{s+1} - d, x_r - c < \delta$, and thus we have $-2\delta < S - (d-c) < 2\delta$, and so $|S - (d-c)| < 2\delta$.

So if $\epsilon > 0$, we can take $\delta < \epsilon/2$, and then if P is a partition of width less than δ, we have $|S - (d-c)| < 2\delta < \epsilon$. Thus by definition $\int_a^b f(x)\,dx = d - c$.

Example 5.9. Let $\chi_{\mathbb{Q}} : [a,b] \to \mathbb{R}$ be the characteristic function of the rationals, defined by $\chi_{\mathbb{Q}}(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$ We claim that $\int_a^b \chi_{\mathbb{Q}}(x)\,dx$ does not exist.

Let $P = \{x_0, \dots, x_n\}$ be any partition of $[a, b]$. Each interval $[x_{i-1}, x_i]$ contains a rational number, so we can take each x_i^* to be rational. Then the corresponding Riemann sum is

$$S_1 = \sum_{i=1}^{n} \chi_{\mathbb{Q}}(x_i^*)(x_i - x_{i-1}) = \sum_{i=1}^{n} 1(x_i - x_{i-1}) = x_n - x_0 = b - a.$$

But it is also true that each interval $[x_{i-1}, x_i]$ contains an irrational number, so we can take each x_i^* to be irrational. Then the corresponding Riemann sum is

$$S_2 = \sum_{i=1}^{n} \chi_{\mathbb{Q}}(x_i^*)(x_i - x_{i-1}) = \sum_{i=1}^{n} 0 \cdot (x_i - x_{i-1}) = 0.$$

Then it is clear that no Riemann integral exists. Suppose $I = \int_a^b \chi_{\mathbb{Q}}(x)\,dx$ exists, and let $\epsilon = (b-a)/2$. Then for any $\delta > 0$, we can find a partition of width less than δ, and by the above argument we have $|(b-a) - I| < \epsilon$ and $|0 - I| < \epsilon$, which by the triangle inequality gives us $|b - a| < 2\epsilon = (b-a)/2$ which is a contradiction.

5.2 Integral Properties and Linearity

In this section, we want to establish and prove a few important properties of the integral. These properties will hold for any integrals that do in fact exist; but we don't yet have an easy way of showing that integrals do exist. That will come in the next two sections.

Proposition 5.10 (Linearity)**.** *The Riemann Integral is a linear function from the vector space of integrable functions on $[a, b]$ to $\mathbb{R}$. That is,*

1. *If $f, g : [a, b] \to \mathbb{R}$ are integrable functions, then $f + g$ is integrable, and*

$$\int_a^b f(x) + g(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$$

2. *If $f : [a, b] \to \mathbb{R}$ is an integrable function and $c \in \mathbb{R}$, then $cf(x)$ is integrable and*

$$\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx.$$

Proof. 1. Let $\epsilon > 0$. Then there are δ_1, δ_2, so that if a partition $P = \{x_0, \dots, x_n\}$ has width less than δ_1 then $\left|\sum_{i=1}^n f(x_i^*)(x_i - x_{i-1}) - \int_a^b f(x)\,dx\right| < \epsilon/2$, and similarly if P has width less than δ_2 then $\left|\sum_{i=1}^n g(x_i^*)(x_i - x_{i-1}) - \int_a^b g(x)\,dx\right| < \epsilon/2$.

Let $\delta = \min \delta_1, \delta_2$. Then if P has width less than δ, we see that

$$\begin{aligned}
&\left|\sum_{i=1}^{n} \left(f(x_i^*) + g(x_i^*)\right)(x_i - x_{i-1}) - \left(\int_a^b f(x)\,dx + \int_a^b g(x)\,dx\right)\right| \\
&\quad = \left|\sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) - \int_a^b f(x)\,dx + \sum_{i=1}^{n} g(x_i^*)(x_i - x_{i-1}) - \int_a^b g(x)\,dx\right| \\
&\quad \leq \left|\sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) - \int_a^b f(x)\,dx\right| + \left|\sum_{i=1}^{n} g(x_i^*)(x_i - x_{i-1}) - \int_a^b g(x)\,dx\right| \\
&\quad < \epsilon/2 + \epsilon/2 + \epsilon.
\end{aligned}$$

Thus by definition, $\int_a^b f(x) + g(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$.

2. Exericse.

□

Proposition 5.11. *If $f : [a,b] \to \mathbb{R}$ is integrable, and $f(x) \geq 0$ for all $x \in [a,b]$, then $\int_a^b f(x)\,dx \geq 0$.*

Proof. For any $\epsilon > 0$ we can find a Riemann sum $S = \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1})$ such that $|S - \int_a^b f(x)\,dx| < \epsilon$. Then we have

$$\begin{aligned}
\epsilon &> S - \int_a^b f(x)\,dx \\
\int_a^b f(x)\,dx &> S - \epsilon.
\end{aligned}$$

But clearly $S \geq 0$, so $\int_a^b f(x)\,dx > -\epsilon$.

This is true for any $\epsilon > 0$, so we can conclude that $\int_a^b f(x)\,dx \geq 0$.

□

Corollary 5.12. *If $f, g : [a,b] \to \mathbb{R}$ are integrable, and $f(x) \leq g(x)$ for all $x \in [a,b]$, then*

$$\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx.$$

Corollary 5.13. *If $f : [a,b] \to \mathbb{R}$ is integrable, and $m, M \in \mathbb{R}$ with $m \leq f(x) \leq <$ for all $x \in [a,b]$, then*

$$m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a).$$

5.3 Existence of Integrals

Now that we understand some properties of the integral, we would like to know when the integral is well-defined. In this section and the next we will prove results about which functions are integrable. The primary result is that any continuous function defined on a closed interval is integrable, which we will see in section 5.4. But first we will prove some other results about integrability.

We begin with a technical lemma that completely but awkwardly characterizes the integrable functions. This lemma is essentially an application of completeness from section 3.2, and is the only way we can really prove that integrals have to converge in the abstract.

Lemma 5.14. *Let $f : [a, b] \to \mathbb{R}$ be a real-valued function. Then f is integrable if and only if: for any $\epsilon > 0$ there is a $\delta > 0$ so that if S_1, S_2 are two Riemann sums corresponding to partitions of width less than δ, then $|S_1 - S_2| < \epsilon$.*

Proof. First, suppose f is integrable and $\int_a^b f(x)\,dx = I$. If $\epsilon > 0$, there exists a $\delta > 0$ so that if S is a Riemann sum corresponding to a partition of width less than δ, then $|S - I| < \epsilon/2$.

So suppose S_1, S_2 are two Riemann sums for f corresponding to partitions of width less than δ. Then

$$|S_1 - S_2| = |S_1 - I - (S_2 - I)| \leq |S_1 - I| + |S_2 - I| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Conversely, suppose $f : [a, b] \to \mathbb{R}$, and for every $\epsilon > 0$ there is a $\delta > 0$ so that whenever S_1, S_2 are Riemann sums corresponding to partitions of width less than δ, then $|S_1 - S_2| < \epsilon$. We wish to show that f is integrable.

We begin by finding a plausible integral, by constructing a sequence which is Cauchy and thus convergent. For each $n \in \mathbb{N}$, choose some partition of width $1/n$ and some Riemann sum S_n corresponding to that partition. Then (S_n) is a sequence of real numbers.

We claim that S_n is Cauchy. Let $\epsilon > 0$. Then there is a $\delta > 0$ so that whenever S_1, S_2 are Riemann sums corresponding to partitions of width less than δ, then $|S_1 - S_2| < \epsilon$. So let $N > 1/\delta$, and then if $n, m > N$ then $1/n, 1/m < \delta$ and so $|S_n - S_m| < \epsilon$.

Then (S_n) is a Cauchy sequence of real numbers, so it has a limit, which we will call I. We claim that $I = \int_a^b f(x)\,dx$. Let $\epsilon > 0$. Then there is some $\delta_1 > 0$ so that whenever S, T are Riemann sums corresponding to partitions of width less than δ_1, then $|S - T| < \epsilon/2$, and there is some $N \in \mathbb{N}$ so that if $n > N$ then $|S_n - I| < \epsilon/2$.

Let $\delta = \min\{\delta_1, 1/N\}$, and suppose S is a Riemann sum corresponding to a partition of width less than δ. If $n > N$ we know that $|S - S_n| < \epsilon/2$, and $|S_n - I| < \epsilon/2$. Then

$$|S - I| = |S - S_n + S_n - I| \leq |S - S_n| + |S_n - I| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Thus $I = \int_a^b f(x)\,dx$ by definition.

□

Remark 5.15. This sort of completeness result really is necessary to make integrals converge. If we consider integrals in the rationals, then $\int_1^2 \frac{dx}{x}$ is perfectly well-defined and seems like the sort of thing that ought to converge. But of course it doesn't, since it would converge to $\ln(2) \notin \mathbb{Q}$.

Before we move on to continuous functions, I want to focus on another type of function whose integrals always exist, and see how much we can learn from those.

Definition 5.16. We say $f : [a, b] \to \mathbb{R}$ is a *step function* if there is some partition $P = \{x_0, \dots, x_n\}$ for $[a, b]$ so that f is constant on each open interval (x_i, x_{i+1}).

Example 5.17. Any constant function is a step function.

The function $f(x) = \begin{cases} 0 & x < 1 \\ 3 & 1 < x < 2 \\ 1 & 2 < x \end{cases}$ is a step function.

The function $\lfloor x \rfloor$ that gives the greatest integer less than or equal to x is a step function.

Clearly step functions should be integrable—the standard Riemann sum picture with rectangles is not just an approximation here as long as you choose your partition sensibly. We still need to prove that step functions are integrable, though, since the integral needs to converge for *any* partition, not just for sensible ones.

A class doing this in greater depth would define the concept of a *refinement* of a partition, which is a partition with more points that includes all the points of the starting partition. With that concept this proof is easy direclty from the definition. But there's a far easier argument that we can make using the the ideas of section 5.2.

Lemma 5.18. *Any step function is integrable.*

Further, if $f : [a, b] \to \mathbb{R}$ is a step function defined on the partition $\{x_0, \dots, x_n\}$ so that $f(x) = c_i$ for all $x_{i-1} < c_i < x_i$, then

$$\int_a^b f(x)\,dx = \sum_{i=1}^{n} c_i(x_i - x_{i-1}).$$

Proof. We can prove this directly, but that's a huge pain. It's much easier to use results we already have to prove this.

For each i, define $\phi_i(x) : [a, b] \to \mathbb{R}$ by $\phi_i(x) = \begin{cases} 1 & x_{i-1} < x < x_i \\ 0 & \text{otherwise} \end{cases}$ (This means that $\phi_i(x) = \chi_{(x_{i-1}, x_i)}(x)$ is the characteristic function of the open interval (x_{i-1}, x_i), but that

notation is extremely cumbersome). It's easy to see (by example 5.8) that $\int_a^b \phi_i(x)\,dx = (x_i - x_{i-1})$.

It's not quite the case that $f(x) = \sum_{i=1}^n c_i\phi_i(x)$, but it's close. We see that $f(x) - \sum_{i=1}^n c_i\phi_i(x) = 0$ unless $x = x_i$ for some i; thus it is zero except at finitely many points. By exercise 5.7 we know that

$$\int_a^b \left(f(x) - \sum_{i=1}^n c_i\phi_i(x)\right)\,dx = 0.$$

But then by linearity, we we have

$$\begin{aligned}\int_a^b f(x)\,dx &= \int_a^b \left(f(x) - \sum_{i=1}^n c_i\phi_i(x) + \sum_{i=1}^n c_i\phi_i(x)\right)\,dx \\ &= \int_a^b \left(f(x) - \sum_{i=1}^n c_i\phi_i(x)\right)\,dx + \int_a^b \left(\sum_{i=1}^n c_i\phi_i(x)\right)\,dx \\ &= 0 + \sum_{i=1}^n c_i \int_a^b \phi_i(x)\,dx \\ &= \sum_{i=1}^n c_i(x_i - x_{i-1}).\end{aligned}$$

Fact 5.19. *Let $f : [a,b] \to \mathbb{R}$. Then f is integrable if and only if for every $\epsilon > 0$ there exist step functions $f_1, f_2 : [a,b] \to \mathbb{R}$ such that*

$$f_1(x) \le f(x) \le f_2(x) \qquad \text{for every } x \in [a,b]$$

and

$$\int_a^b f_2(x) - f_1(x)\,dx < \epsilon.$$

This is effectively a variation on the Squeeze Theorem. This idea eventually grows into the concept of a Lesbesgue integral, where very much like this is taken to be the definition of an integral.

□

5.4 Uniform Continuity and Integrals of Continuous Functions

The main goal of this section is to prove that a continuous function is integrable. The basic idea is this: on a small subinterval, the distance between the maximum and minimum values of x are close together—since the point of continuity is that when inputs are close

together, so are outputs. So as partitions get smaller, the difference between the maximum and minimum Riemann sums will get smaller.

But this argument by itself doesn't quite work. As each subinterval gets smallerthe maximum and minimum values get closer together; but you would need to do this for each subinterval individually, and the number of subintervals goes to infinity so this doesn't actually work. We need some guarantee that we can control the function everywhere at once.

Definition 5.20. Let (E, d) and (F, d') be metric spaces, and let $f : E \to F$ be a function of metric spaces. We say f is *uniformly continuous* if, for every $\epsilon > 0$, there is a $\delta > 0$ so that if $x, y \in E$ and $d(x, y) < \delta$, then $d'(f(x), f(y)) < \epsilon$.

This definition is different from continuity in that we don't have a special point we're considering. We have to pick one δ that works everywhere.

Example 5.21. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x$. We claim f is uniformly continuous.

Let $\epsilon > 0$ and let $\delta = \epsilon$. If $|x - y| < \delta$ then $|f(x) - f(y)| = |x - y| < \delta = \epsilon$.

Example 5.22. Let $f : [0, 1] \to \mathbb{R}$ be defined by $f(x) = x^2$. We claim f is uniformly continuous.

Let $\epsilon > 0$, and set $\delta = \epsilon/2$. If $|x - y| < \delta$, then $|x + y| \leq 2$, so

$$|x^2 - y^2| = |x - y| \cdot |x + y| < \delta|x + y| \leq 2\delta = \epsilon.$$

Example 5.23. Let $f : (0, 1) \to \mathbb{R}$ be defined by $f(x) = 1/x$. Then f is continuous, since it's given by algebraic operations. We claim that f is not uniformly continuous. Heuristically, this is true because the same distance on the x-axis creates larger and larger distances on the y-axis as we move closer to $x = 0$.

To prove this, fix $\epsilon > 0$, and fix $\delta > 0$. We want to find x, y so that $|x - y| < \delta$ but $\left|\frac{1}{x} - \frac{1}{y}\right| \geq \epsilon$. There are lots of options here. If for simplicity we take $y = x/2$, then we need $0 < x < 2\delta$ and

$$\epsilon \leq \left|\frac{1}{x} - \frac{1}{x/2}\right| = \left|\frac{1}{x} - \frac{2}{x}\right| = \frac{1}{x},$$

which is equivalent to $x \leq 1/\epsilon$.

So if we take $x < \min\{1/\epsilon, 2\delta, 1\}$ and $y = x/2$, then we have $|x-y| < \delta$ and $|1/x-1/y| \geq \epsilon$. Since we can find these x, y for any δ, we know that no δ exists that "works" for every x, y, and thus f is not uniformly continuous.

Exercise 5.24. *Let $f : (0,1) \to \mathbb{R}$ be defined by $f(x) = \sin(1/x)$. Show that f is not uniformly continuous.*

However, the problem is, essentially, when we have to deal with infinitely many intervals. If we can somehow limit ourselves to finitely many intervals, we can make everything work out.

Proposition 5.25. *A continuous function on a compact set is uniformly continuous.*

Explicitly: let (E,d) and (F,d') be metric spaces, and $f : E \to F$ continuous. If E is compact, then f is uniformly continuous

Proof. We will use topological compactness again here. A proof using sequential compactness exists, but is much more complicated.

The basic idea is that since f is continuous on E, we can find a δ at each point. Since there are infinitely many points, we have infinitely many δ, and we can't just take the minimum one. But the point of compactness is to allow us to treat an infinite set sort of like a finite set; compactness allows us to narrow this down to finitely many δ, at which point we can take the minimum.

Let $\epsilon > 0$. We want to show that, for some $\delta > 0$, then whenever $d(x,y) < \delta$ we know that $d'(f(x), f(y)) < \epsilon$. So assume for contradiction no such δ exists.

Because f is continuous on E, for each $x \in E$, we can find a real number $\delta_x > 0$ so that if $d(x,y) < \delta_x$ then $d'(f(x), f(y)) < \epsilon/2$. Then for each x we can consider the ball of radius $\delta_x/2$ centered at x. Since $x \in B_{\delta_x/2}(x)$ for each $x \in E$, we know that $E = \bigcup_{x \in E} B_{\delta_x/2}(x)$.

We have now written E as an infinite union of open sets. Since E is compact, we can in fact choose finitely many of those sets and use them to cover all of E. So there exist $x_1, \dots, x_n \in E$ such that $E = \bigcup_{i=1}^n B_{\delta_{x_i}/2}(x_i)$. We set $\delta = \min\{\delta_{x_i}/2\}$.

Now suppose $x, y \in E$ and $d(x,y) < \delta$. We want to show that $f(x)$ and $f(y)$ are close together; we can do this by showing that x and y are both close to some x_i, and then using the definition of δ above to show that $f(x)$ and $f(y)$ are both close to $f(x_i)$.

Since $x \in E = \bigcup_{i=1}^n B_{\delta_{x_i}/2}(x_i)$, there is some i so that $x \in b_{\delta_{x_i}/2}(x_i)$, and thus $d(x, x_i) < \delta_{x_i}/2$. But we can show that y also has to be close to x_i; because

$$d(y, x_i) \leq d(y,x) + d(x, x_i) < \delta + \delta_{x_i}/2 < \delta_{x_i}/2 + \delta_{x_i}/2 = \delta_{x_i}.$$

Now we have shown that since x and y are close together, there must be some x_i they are both close to. Now we just use this to show that $f(x)$ and $f(y)$ have to be close together.

But we know that $d(f(x), f(x_i)) < \epsilon/2$ since $d(x, x_i) < \delta \leq \delta_{x_i}/2 < \delta_{x_i}$. And similarly we know that $d(y, x_i) < \delta_{x_i}$ and thus $d(f(y), f(x_i)) < \epsilon/2$. So

$$d(x, y) \leq d(x, x_i) + d(x_i, y) < \epsilon/2 + \epsilon/2 = \epsilon.$$

□

Now we can finally prove the convergence theorem for integrals of continuous functions.

Theorem 5.26. *Let $f : [a, b] \to \mathbb{R}$ be continuous. Then $\int_a^b f(x)\,dx$ exists.*

Proof. Recall our Cauchy criterion for proving integrals converge: lemma 5.14 says that f is integrable if and only if: for any $\epsilon > 0$ there is a $\delta > 0$ so that if S_1, S_2 are two Riemann sums corresponding to partitions of width less than δ, then $|S_1 - S_2| < \epsilon$. We combine this with our results on uniform continuity to show that any continuous function is integrable.

Since f is continuous on the compact set $[a, b]$, we know that f is in fact uniformly continuous. Thus for any $\epsilon > 0$ we can find a $\delta > 0$ so that if $|x - y| < \delta$ then $|f(x) - f(y)| < \frac{\epsilon}{2(b-a)}$.

Let S be a Riemann sum corresponding to a partition $P = \{x_0, \ldots, x_n\}$, and S' be a Riemann sum corresponding to a partition $Q = \{y_0, \ldots, y_m\}$, with $\text{width}(P), \text{width}(Q) < \delta$. We want to show that $|S - S'| < \epsilon$.

Our approach is to define another Riemann sum T that corresponds to another (specific) partition. Then we will show that $|S - T|, |S' - T| < \epsilon/2$. Once we have show this, the triangle inequality whill tell us that $|S - S'| < \epsilon$, and thus by our "completeness" lemma we will know that f is integrable.

Let $P \cup Q = \{z_0, \ldots, z_\ell\}$ be the partition obtained by taking all the points that are in either P or Q. Then it is clear that $\text{width}(P \cup Q) < \delta$. Let T be a Riemann sum corresponding to $P \cup Q$.

We can write S as a not-quie-Riemann sum corresponding to $P \cup Q$, by simply splitting

each interval of P up into the intervals of $P \cup Q$. Then we get something like

$$\begin{aligned} S &= \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) \\ &= \sum_{i=1}^{n} f(x_i^*)(z_{j_i} - z_{j_{i-1}}) \\ &= \sum_{i=1}^{n} f(x_i^*) \left(\sum_{k=j_{i=1}}^{j_i} z_k - z_{k-1} \right) \\ &= \sum_{k=1}^{\ell} f(x_i^*)(z_k - z_{k-1}). \end{aligned}$$

This isn't technically a Riemann sum because x_i^* doesn't necessarily belong to every sub-subinterval.

But now, if T is any Riemann sum corresponding to $P \cup Q$ we have

$$\begin{aligned} |S - T| &= \left| \sum_{k=1}^{\ell} f(x_i^*)(z_k - z_{k-1}) \right| - \left| \sum_{k=1}^{\ell} f(z_k^*)(z_k - z_{k-1}) \right| \\ &\leq \left| \sum_{k=1}^{\ell} \left(f(x_i^*) - f(z_k^*) \right) (z_k - z_{k-1}) \right| \\ &\leq \sum_{k=1}^{\ell} |f(x_i^*) - f(z_k^*)| \, (z_k - z_{k-1}). \end{aligned}$$

But since x_i^* and z_k^* are in the same subinterval of the partition P, and we know width $P < \delta$, then by uniform continuity we know that $|f(x_i^*) - f(z_k^*)| < \frac{\epsilon}{2(b-a)}$. Thus

$$\begin{aligned} |S - T| &\leq \sum_{k=1}^{\ell} \frac{\epsilon}{2(b-a)} (z_k - z_{k-1}) \\ &= \frac{\epsilon}{2(b-a)} \sum_{k=1}^{\ell} (z_k - z_{k-1}) \\ &= \frac{\epsilon}{2(b-a)} (b - a) = \epsilon/2. \end{aligned}$$

Nothing in this argument depended on the specific properties of S and T, so by the exact same argument we can conclude that $|S' - T| < \epsilon/2$. Thus the triangle inequality tells us that $|S - S'| < \epsilon$. We have shown that for any $\epsilon > 0$ there is a $\delta > 0$ such that if S, S' are two Riemann sums corresponding to partitions of width less than δ, then $|S - S'| < \epsilon$. Thus by lemma 5.14, we know that f is integrable.

□

5.5 The Fundamental Theorem of Calculus

In this section we want to use the integral to define new functions. In particular, if $f : [c, d] \to \mathbb{R}$ is integrable, and $a \in [c, d]$, then we can define a function $F(x) = \int_a^x f(t)\,dt$. Since f is integrable, this function is well-defined for any $x \neq a$, and if we define $\int_a^a f(t)\,dt = 0$ then F is defined on all of $[c, d]$. We call F an *indefinite integral* of f.

In fact F is differentiable, and we want to compute the derivative of F. But we need a couple intermediate results first.

Proposition 5.27. *If $a < b < c$ and $f : [a, c] \to \mathbb{R}$ is a function, then f is integrable on $[a, c]$ if and only if it is integrable on $[a, b]$ and $[b, c]$, and if it is integrable, we have the equality*

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx.$$

Remark 5.28. If we define $\int_a^a f(x)\,dx = 0$ and $\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$, then this result holds for any a, b, c such that at least two of the integrals exist.

Theorem 5.29 (Fundamental Theorem of Calculus)**.** *Let $U \subset \mathbb{R}$ be an open interval with $a \in U$, and let $f : U \to \mathbb{R}$ be continuous. Define $F : U \to \mathbb{R}$ by $F(x) = \int_a^x f(t)\,dt$. Then F is differentiable, and $F'(x) = f(x)$.*

Proof. Since f is continuous, we know that $F(x)$ is defined for any $x \in U$. W need to compute the derivative. So we have

$$\begin{aligned} F'(x) &= \lim_{h\to 0} \frac{1}{h}\left(\int_a^{x+h} f(t)\,dt - \int_a^x f(t)\,dt\right) \\ &= \lim_{h\to 0} \frac{1}{h}\int_x^{x+h} f(t)\,dt. \end{aligned}$$

We wish to show that this limit is equal to $f(x)$. We compute that

$$\begin{aligned} \left|\frac{1}{h}\int_x^{x+h} f(t)\,dt - f(x)\right| &= \left|\frac{1}{h}\int_x^{x+h} f(t)\,dt - \frac{1}{h}\int_x^{x+h} f(x)\,dt\right| \\ &= \left|\frac{1}{h}\int_x^{x+h} f(t) - f(x)\,dt\right| \\ &\leq \frac{1}{h}\int_x^{x+h} |f(t) - f(x)|\,dt. \end{aligned}$$

Since f is continuous at x, for any $\epsilon > 0$ we can find a δ so that if $|t - x| < \delta$ then $|f(t) - f(x)| < \epsilon$. Suppose $|h| < 0$. Then for any $t \in (x, x+h)$ we know that $|f(t) - f(x)| < \epsilon$,

and so

$$\begin{aligned}\frac{1}{h}\int_x^{x+h} |f(t)-f(x)|\,dt &< \frac{1}{h}\int_x^{x+h} \epsilon\,dt \\ &= \frac{1}{h}h\epsilon = \epsilon.\end{aligned}$$

Thus by definition, $F'(x) = f(x)$.

□

Corollary 5.30. *If $f : (c,d) \to \mathbb{R}$ is continuous, then there is a differentiable function $F : (c,d) \to \mathbb{R}$ such that $F'(x) = f(x)$.*

Proof. If $a \in (c,d)$, define $F(x) = \int_a^x f(t)\,dt$. □

Corollary 5.31. *If $F : (c,d) \to \mathbb{R}$ is differentiable and $F'(x) = f(x)$, and $[a,b] \subset (c,d)$, then $\int_a^b f(t)\,dt = F(b) - F(a)$.*

Proof. We know that

$$\frac{d}{dx}\left(\int_a^x f(t)\,dt - F(x)\right) = f(x) - f(x) = 0.$$

Thus $\int_a^x f(t)\,dt = F(x) + C$ for some fixed constant $C \in \mathbb{R}$. We can find our constant by plugging in a for x, so we get

$$F(a) + C = \int_a^a f(t)\,dt = 0$$

and thus $C = -F(a)$. So

$$\begin{aligned}\int_a^x f(t)\,dt &= F(x) - F(a) \\ \int_a^b f(t)\,dt &= F(b) - F(a).\end{aligned}$$

□

Proof of Proposition 5.27. □

Corollary 5.32 (Change of Variables)**.** *Let $U, V \subset \mathbb{R}$ be open intervals, $\phi : U \to V$ be continuously differentiable, and $f : V \to \mathbb{R}$ continuous. Then*

$$\int_{\phi(a)}^{\phi(b)} f(v)\,dv = \int_a^b f(\phi(u))\phi'(u)\,du.$$

Proof. Define $F : V \to \mathbb{R}$ by $F(y) = \int_{\phi(a)}^{y} f(v)\,dv$. Then $F'(y) = f(y)$. Then if we define $G : U \to \mathbb{R}$ by $G(x) = \int_{\phi(a)}^{\phi(x)} f(v)\,dv$, we see that $G(x) = F(\phi(x))$, and then by the chain rule we have

$$G'(x) = F'(\phi(x))\phi'(x) = f(\phi(x))\phi'(x).$$

But then G is an antiderivative for $f(\phi(x))\phi'(x)$, and thus

$$G(x) = \int_a^x f(\phi(u))\phi'(u)\,du + C$$

for some constant $C \in \mathbb{R}$. But we can compute that

$$G(a) = \int_{\phi(a)}^{\phi(a)} f(v)\,dv = 0$$
$$\int_a^a f(\phi(u))\phi'(u)\,du = 0$$

and thus $0 = G(a) = 0 + C$ and so $C = 0$. Thus

$$G(x) = \int_a^x f(\phi(u))\phi'(u))\,du.$$

□

5.6 Logarithm and Exponent

Definition 5.33. Define $\log(x) = \int_1^x \frac{dt}{t}$.

Proposition 5.34. *The function* $\log(x)$ *is a differentiable function on* $(0, \infty)$ *with* $\frac{d}{dx}\log(x) = \frac{1}{x}$. *It is strictly increasing,* $\log(1) = 0$ *and has image all of* $\mathbb{R}$, *and*

- $\log(xy) = \log(x) + \log(y)$
- $\log(x/y) = \log(x) - \log(y)$
- $\log(x^n) = n\log(x)$ *for* $n \in \mathbb{Z}_{\geq 0}$.

Proposition 5.35. *The derivative result follows directly from the definition and the fundamental theorem of calculus. It is increasing since* $\frac{d}{dx}\log(x) = \frac{1}{x} > 0$. *We know* $\log(1) = \int_1^1 \frac{dt}{t} = 0$.

Let $z = xy$. *Then*

$$\frac{d}{dx}\log(z) = \frac{1}{z}\cdot\frac{d}{dx}z = \frac{1}{xy}\cdot y = \frac{1}{x}.$$

This means that $\log(z) = \log(x) + C$ *for some* $C \in \mathbb{R}$. *Plugging in* $x = 1$ *gives us* $\log(y) = 0 + C$, *and thus* $\log(xy) = \log(x) + \log(y)$.

If $x = 1/y$ *then we have* $\log(1) = \log(y) + \log(1/y)$ *and thus* $\log(1/y) = -\log(y)$; *then*

$$\log(x/y) = \log(x \cdot 1/y) = \log(x) + \log(1/y) = \log(x) - \log(y).$$

The last result follows by induction on n.

6 Sequences of Functions

Definition 6.1. For each $n \in \mathbb{N}$, let $f_n : E \to F$ be a function of metric spaces. Then for each a we have a sequence $f_n(a)$ of points in F. If $f : E \to F$, we say that f_n *converges pointwise* to f if $\lim_{n\to\infty} f_n(a) = f(a)$ for each $a \in E$. We write $f = \lim_{n\to\infty} f_n$.

Example 6.2. The sequence of functions $f_n(x) = \frac{x}{n}$ converges pointwise to the contant zero function.

Example 6.3. Let $f_n(x) = \frac{1}{n}x + \left(1 - \frac{1}{n}\right)x^2$. Then $\lim_{n\to\infty} f_n = f$ where $f(x) = x^2$.

We'd like to be able to extend properties of our sequences to their limits. For example, we showed that the limit of a sequence of positive numbers is positive. The most important property of functions is continuity. Is the limit of a sequence of continuous functions continuous?

Example 6.4. Let $f_n : [0,1] \to \mathbb{R}$ be defined by $f_n(x) = x^n$. Then for each $x \in [0,1)$, we have $\lim_{n\to\infty} x^n = 0$, but $\lim_{n\to\infty} 1^n = 1$. So $\lim_{n\to\infty} f_n = \begin{cases} 0 & x \in [0,1) \\ 1 & x = 1 \end{cases}$

The assertion that the limit of a sequence of continuous functions is sometimes called "Cauchy's Wrong Theorem". The basic problem is that while each function in the sequence is continuous, the functions can get steeper and steeper as we get further into the sequence; each function has finite steepness and is continuous, but in the limit this steepness goes to infinity.

But this can only happen in this example because, while $f_n(x)$ goes to zero for every $x \neq 1$, it does this slower and slower as x gets closer to 1, allowing the function to stretch. If we find a way to prevent that, we can keep the function from stretching and becoming discontinuous in the limit.

Definition 6.5. For each n, let $f_n : E \to F$ be a function of metric spaces. We say that the sequence of functions (f_n) *converges uniformtly* to a function f if, for any $\epsilon > 0$, there is a $N \in \mathbb{N}$ so that if $n > N$, then $d(f_n(x), f(x)) < \epsilon$ for any $x \in E$.

Notice we've seen an idea like this before. A function is continuous if we can find a δ for any pair of ϵ and x; it is uniformly continuous if for each ϵ, we can find a δ that works for any x. Similarly, functions converge pointwise if for each x and ϵ you can find an N that works; they converge uniformly if you can find an N that works for any x.

Clearly, a sequence that converges uniformly also converges pointwise.

Example 6.6. If $f_n : \mathbb{R} \to \mathbb{R}$ is given by $f_n(x) = x/n$, then the convergence is not uniform; for any $\epsilon > 0, N \in \mathbb{N}$, we can take $x > (N+1)\epsilon$ and then $|f_{N+1}(x) - 0| = x/(N+1) > \epsilon$.

However, if we define $f_n : [0,1] \to \mathbb{R}$ instead, then the sequence converges uniformly. Let $\epsilon > 0$ and set $N > 1/\epsilon$. Then if $n < N$ we have

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

since $0 \leq x \leq 1$.

Example 6.7. Let $f_n : [0,1] \to \mathbb{R}$ be defined by $f_n(x) = x^n$. Then f_n converges to $f(x) = \begin{cases} 0 & x \in [0,1) \\ 1 & x = 1 \end{cases}$ pointwise, but the convergence is not uniform.

Let $\epsilon = 1/2$, and let $N \in \mathbb{N}$. We want to show that there is some $n > N$ and some $x \in [0,1]$ so that $d(f_n(x), f(x)) \geq \epsilon = 1/2$.

For each $x \neq 1$ we have $f(x) = 0$, so we just want to show that $f_n(x) \geq 1/2$. This is equivalent to $x^n \geq 1/2$ or $x \geq \sqrt[n]{1/2}$. So let $n < N$ and $x = \sqrt[n]{1/2}$. Then $d(f_n(x), f(x)) = |1/2 - 0| = 1/2 \geq \epsilon$.

Thus we have no N so that if $n > N$ then $d(f_n(x), f(x)) < \epsilon$ for any x; thus f_n does not converge to f uniformly.

Exercise 6.8. *If f_n converges to f uniformly, then f_n converges to f pointwise.*

Proposition 6.9. *Suppose $f_n : E \to F$ is a sequence of functions each continuous at some point $a \in E$, and f_n converges uniformly to some function $f : E \to F$. Then f is continuous at a.*

Proof. We want to show that f is continuous at a. This means that for any $\epsilon > 0$, there is some $\delta > 0$ so that if $d(x,a) < \delta$ then $d'(f(x), f(a)) < \epsilon$.

Since f_n converges to f uniformly, there is some $N \in \mathbb{N}$ so that if $n > N$, then $d'(f_n(x), f(x)) < \epsilon/3$ for any $x \in E$. So let $n < N$.

We know that f_n is continuous. So there is some δ so that if $d(x,a) < \delta$ then $d'(f_n(x), f_n(a)) < \epsilon/3$. But then we have

$$\begin{aligned} d'(f(x), f(a)) &\leq d'(f(x), f_n(x)) + d'(f_n(x), f_n(a)) + d'(f_n(a), f(a)) \\ &< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \end{aligned}$$

□

This idea of uniform convergence is really powerful. With pointwise convergence, we can make f_n work at any given point for large enough n, but we can't necessarily make things work for every point at the same time. With uniform convergence, we guarantee that we can make things work for the entire function at once.

(Recall we've seen this same effect when we talked about uniform continuity: it made a big difference that we could be continuous everywhere with the same δ).

But also, we've seen what's effectively uniform continuity before, in a completely different context.

Definition 6.10. Let $\mathcal{B}(E,F)$ be the set of bounded functions $f : E \to F$. We can define the sup metric by $d_\infty(f,g) = \sup\{d'(f(x),g(x)) : x \in E\}$.

Thus a sequence f_n converges to a function f uniformly if and only if it converges in the sup metric.

Proposition 6.11. *If F is complete, then $\mathcal{B}(E,F)$ is complete.*

Proof. Let f_n be a Cauchy sequence of functions. Then for each $\epsilon > 0$ there is an $N \in \mathbb{N}$ so that if $n, m > N$ then $d_\infty(f_n, f_m) < \epsilon$. We want to define a function f so that $\lim_{n\to\infty} f_n = f$.

Let $x \in E$. Then we have a sequence $(f_n(x)) \subset F$. For any $\epsilon > 0$, there is a $N \in N$ so that if $n, m > N$, then $d_\infty(f_n, f_m) < \epsilon$. Then

$$d'(f_n(x), f_m(x)) \leq \sup\{d'(f_n(x), f_m(x)) : x \in E\} < \epsilon.$$

Thus the sequence $(f_n(x))$ is Cauchy, and since it is a Cauchy sequence in the complete metric space F, it has a limit. So for each x, define $f(x) = \lim_{n\to\infty} f_n(x)$.

Now we just need to show that $\lim_{n\to\infty} f_n = f$ in the sup metric. Fix $\epsilon > 0$. There is a $N \in \mathbb{N}$ so that if $n, m > N$ then $d_\infty(f_n, f_m) < \epsilon/2$. If we fix $n > N$, then for all $m > N$ we have $d'(f_n(x), f_m(x)) < \epsilon/2$, and thus $f_m(x) \in \overline{B}(f_n(x))$. Thus the sequence $(f_m(x))$ is eventually contained in $\overline{B}(f_n(x))$, and since the closed ball is closed, the limit $\lim_{m\to\infty} f_m(x) \in \overline{B}(f_n(x))$.

Thus $d'(f(x), f_n(x)) \leq \epsilon/2$, and for any $m > N$ we have

$$d'(f_m(x), f(x)) \leq d'(f_m(x), f_n(x)) + d'(f_n(x), f(x)) < \epsilon/2 + \epsilon/2 = \epsilon.$$

Thus $\lim_{n\to\infty} f_n = f$ in the sup metric. So every Cauchy sequence converges, and $\mathcal{B}(E,F)$ is complete. □

When we work with this space at an advanced level, we make a minor technical tweak where we treat functions as equivalent if they differ at only finitely many points, and define the metric as the supremum ignoring finitely many points. (Even more technically, we ignore functions that differ on a "measure zero" set). But everything we said is true without worrying about that.

However, we can ignore all of this by specifying a bit. It's not obvious, but it is true, that if two *continuous* functions differ at only finitely many points, they are in fact identical. So if we need to take a specific representative of this equivalence class, and taking a continuous function is an option, we do that. In fact, we can say a bit more about this specific case:

Corollary 6.12. *If E is compact and F is complete, then $\mathcal{C}(E,F)$ is a complete metric space under the sup metric.*

Proof. Since E is compact, every continuous function is bounded. And by proposition 6.9, the limit of any continuous function is continuous. Thus $\mathcal{C}(E,F)$ is a closed subset of a complete metric space, and thus complete. □

Remark 6.13. There are other ways to talk about function convergence. Earlier in the course we talked about $L^1([a,b],\mathbb{R})$ with the metric $d_1(f,g) = \int_a^b |f(x)-g(x)|\,dx$. In fact, for any $p \geq 1$ we can define

$$d_p(f,g) = \sqrt[p]{\int_a^b |f(x)-g(x)|^p\,dx}.$$

These different metrics are analogous to the various metrics we defined on $\mathbb{R}^n$; but unlike in the case of $\mathbb{R}^n$, they are actually genuinely distict metrics with different convergence properties.

Understanding the L^p spaces is *far* beyond the scope of this course; I'll only say that we genuinely can't understand them fully without using the equivalence class idea I mentioned earlier, since otherwise we lose non-negativity of metrics. (A function that is zero except at one point has integral zero).

But the theory of what each of these different metrics does to convergence is deep and interesting, and useful in fields like statistics and differential equations.