Math 320-1: Real Analysis
Northwestern University, Lecture Notes

Written by Santiago Cañez

These are notes which provide a basic summary of each lecture for Math 320-1, the first quarter of "Real Analysis", taught by the author at Northwestern University. The book used as a reference is the 4th edition of *An Introduction to Analysis* by Wade. Watch out for typos! Comments and suggestions are welcome.

Contents

Lecture 1: Introduction to the Real Numbers

Today I gave a brief introduction to some concepts we'll be looking at this quarter, such as continuous functions and compact sets. Note that the book does not use the term "compact" until later on when it is looking at higher-dimensional spaces like $\mathbb{R}^2$ and $\mathbb{R}^3$; this is something you would study more closely in Math 320-2 but I think it is important to understand this concept this quarter as well. I will point out what results in the earlier chapters are really about compactness as we go on, even though the book won't phrase these results in the same way.

Facts about absolute values. For $\epsilon > 0$, the inequality $|x-a| < \epsilon$ means that the distance (on a number line) between x and a is less than ϵ. This can be rephrased as

$$-\epsilon < x - a < \epsilon.$$

Adding a to both sides, this is the same as

$$a - \epsilon < x < a + \epsilon.$$

Finally, this in turn means that x is in the open interval from $a-\epsilon$ to $a+\epsilon$:

$$x \in (a-\epsilon, a+\epsilon).$$

It will be incredibly useful to become comfortable manipulating inequalities involving absolute values in this manner.

Triangle Inequality. ***FINISH***

Archimedean Property of $\mathbb{R}$. The Archimedean Property of $\mathbb{R}$ says: for any $x \in \mathbb{R}$, there exists $N \in \mathbb{N}$ such that $x < N$. In words, given any real number we can find a positive integer larger than it. This is given as Theorem 1.16 in the book, although it is phrased in a slightly different manner. I'll leave it to you to understand why both versions are actually saying the same thing.

Proposition. *For any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $\frac{1}{N} < \epsilon$. In words, given any positive real number (no matter how small), there is a positive fraction of the form $\frac{1}{N}$, where N is a positive integer, which is smaller than it.*

Proof. Since $\epsilon \neq 0$, $\frac{1}{\epsilon}$ is a real number. By the Archimedean Property of $\mathbb{R}$, there exists $N \in \mathbb{R}$ such that $\frac{1}{\epsilon} < N$. Since ϵ and N are both positive, multiplying by ϵ and dividing by N does not change the direction of the inequality, so we get $\frac{1}{N} < \epsilon$ as desired. □

This is the version of the Archimedean Property which will be most useful for us, and says that fractions of the form $\frac{1}{N}$ can be made arbitrarily small.

Theorem. *Given $x, y \in \mathbb{R}$ with $x < y$, there exists $\frac{a}{b} \in \mathbb{Q}$ such that $x < \frac{a}{b} < y$. In words, given any two numbers (no matter how close they are to each other) we can always find a rational number strictly between them. This is what it means to say that $\mathbb{Q}$ is "dense" in $\mathbb{R}$. This is Theorem 1.18 in the book, although we give a different proof.*

Proof. Since $x < y$, $y - x$ is positive so $\frac{1}{y-x}$ is real. By the Archimedean Property, there exists $b \in \mathbb{N}$ such that

$$\frac{1}{y-x} < b.$$

Since $y - x > 0$, this implies that

$$1 < b(y - x), \text{ so } 1 < by - bx.$$

Since the distance between by and bx is greater than 1, there must be an integer between them so there exists $a \in \mathbb{Z}$ such that

$$bx < a < by.$$

Since b is positive, this gives

$$x < \frac{a}{b} < y$$

as desired. □

Lecture 2: Supremums and Infimums

Today we spoke about the notion of the "supremum" of a set. This will be an important concept which we will use throughout the course. We didn't get to defining "infimum" yet, but we'll continue with this on Monday. You can find notes I wrote a few years ago on this topic by going to `http://math.northwestern.edu/~scanez/archives/real-analysis/notes.php`.

Warm-Up. Show that if $-3 \leq x \leq 2$, then $|x^2 + x - 6| \leq 6|x - 2|$.

The key realization here is that the first absolute value factors as $|x + 3||x - 2|$. So if we want this to be $\leq |x - 2|$, it is enough to show that the $|x + 3|$ term is ≤ 6. One way to do this is by adding 3 throughout the given inequality $-3 \leq x \leq 2$, but we do it a different way here. Note that $|x + 3| \leq 6$ means the same thing as

$$-6 \leq x + 3 \leq 6,$$

which means the same thing as $-9 \leq x \leq 3$. So as long as this string of inequalities is true, so will the one we want. Luckily, our given restrictions on x do imply this string of inequalities, so we are good to go.

Proof. Since $-3 \leq x \leq 2$, it is certainly true that $-9 \leq x \leq 3$. Adding 3 throughout gives $-6 \leq x + 3 \leq 6$, so $|x + 3| \leq 6$. Thus

$$|x^2 + x - 6| = |x + 3||x - 2| \leq 6|x - 2|$$

as claimed. □

Definitions. Suppose that $S \subseteq \mathbb{R}$ is a nonempty set of real numbers. An *upper bound* of S is a real number u such that $s \leq u$ for all $s \in S$. We say that S is *bounded above* if it has an upper bound. When S is bounded above, the *supremum* of S is its "least upper bound". To be precise, b is the supremum of S when it satisfies the conditions: (i) b is an upper bound of S, and (ii) for any other upper bound u of S, $b \leq u$. We use the notation $b = \sup S$ for supremums.

Remark. We referred above to "the" supremum of S, without actually justifying the fact that if a set has a supremum, it has only one. This is proved in the book, and we will mention it in class on Monday.

Example 1. We claim that the supremum of $\{x \in \mathbb{R} \mid x^2 \leq 2\}$ is $\sqrt{2}$. First, note that after taking square roots of the inequality defining this set, we see that this set is just the closed interval $[-\sqrt{2}, \sqrt{2}]$. To show that $\sqrt{2}$ is indeed the supremum of this, we must show that it is an upper bound and that it is smaller than any other upper bound. First, any $x \in [-\sqrt{2}, \sqrt{2}]$ certainly satisfies $x \leq \sqrt{2}$ simply due to the definition of closed intervals. Thus $\sqrt{2}$ is an upper bound for this set.

To show that $\sqrt{2}$ is the least upper bound, suppose that u is another upper bound for $[-\sqrt{2}, \sqrt{2}]$. Then

$$s \leq u \text{ for any } s \in [-\sqrt{2}, \sqrt{2}].$$

But in particular, $\sqrt{2}$ itself is in $[-\sqrt{2}, \sqrt{2}]$ so u, being an upper bound, is larger than or equal to it: $\sqrt{2} \leq u$. This shows that $\sqrt{2}$ is an upper bound of $[-\sqrt{2}, \sqrt{2}]$ which is $\leq$ any other upper bound, so it is the supremum as claimed.

Alternate characterization of supremums. An upper bound b of $S \subseteq \mathbb{R}$ is the supremum of S if and only if for any $\epsilon > 0$, there exists $s \in S$ such that $b - \epsilon < s$.

The condition after the "if and only if" is a precise way of saying that nothing smaller than b can possibly be an upper bound of S: as ϵ varies through all positive numbers, $b - \epsilon$ varies through all possible numbers smaller than b, and no such number can be an upper bound of S since we can always find something in S larger than it. Since b is an upper bound of S with the property that nothing smaller than it can be an upper bound, b must be the least upper bound as claimed.

Example 2. We claim that the supremum of the open interval $(-\sqrt{2}, \sqrt{2})$ is also $\sqrt{2}$. This should hopefully be intuitively clear, but proving it is a little different than we did above for the closed interval. The difference is that in this case, the claimed supremum is no longer in the set in question, so the argument we gave before will no longer work. (Make sure you understand why not.)

Instead we use the alternate characterization of supremums given above. First, again it should be simple enough to see that $\sqrt{2}$ is an upper bound of $(-\sqrt{2}, \sqrt{2})$. To show that nothing smaller can be an upper bound, let $\epsilon > 0$. Our goal is to find some $s \in (-\sqrt{2}, \sqrt{2})$ such that $\sqrt{2} - \epsilon < s$. If you draw a picture of $\sqrt{2} - \epsilon$ and $\sqrt{2}$ on a number line, all we need is some number between; their midpoint, which is explicitly given by $\sqrt{2} - \frac{\epsilon}{2}$, works. This should be the number in $(-\sqrt{2}, \sqrt{2})$ which is larger than $\sqrt{2} - \epsilon$, showing that $\sqrt{2} - \epsilon$ cannot be an upper bound.

There is one slight issue with this, in that if ϵ is too large then $\sqrt{2} - \frac{\epsilon}{2}$ won't actually be in $(-\sqrt{2}, \sqrt{2})$. In particular, this happens if $\epsilon \geq 4\sqrt{2}$ since in this case

$$\sqrt{2} - \frac{\epsilon}{2} \leq \sqrt{2} - 2\sqrt{2} = -\sqrt{2}.$$

We get around this by restricting our values of ϵ to those which are $< 4\sqrt{2}$. We'll see why this is enough in the proof below.

Proof of claimed supremum. Anything in $(-\sqrt{2}, \sqrt{2})$ is stricly less than $\sqrt{2}$, so $\sqrt{2}$ is an upper bound for $(-\sqrt{2}, \sqrt{2})$. Now, let $0 < \epsilon < 4\sqrt{2}$. Then

$$-\sqrt{2} = \sqrt{2} - 2\sqrt{2} < \sqrt{2} - \frac{\epsilon}{2} < \sqrt{2}$$

so $s = \sqrt{2} - \frac{\epsilon}{2}$ is in $(-\sqrt{2}, \sqrt{2})$. This s also satisfies

$$\sqrt{2} - \epsilon < s$$

so for $0 < \epsilon < 4\sqrt{2}$ we have found an element of $(-\sqrt{2}, \sqrt{2})$ which is larger than $\sqrt{2} - \epsilon$. For $\epsilon \geq 4\sqrt{2}$, 0 is an element of $(-\sqrt{2}, \sqrt{2})$ which is larger than $\sqrt{2} - \epsilon$ since in this case $\sqrt{2} - \epsilon$ is negative. Thus, this shows that for any $\epsilon > 0$, $\sqrt{2} - \epsilon$ is not an upper bound of $(-\sqrt{2}, \sqrt{2})$, so $\sqrt{2}$ is the least upper bound as claimed. $\square$

Example 3. We claim that the supremum of $(-\sqrt{2}, \sqrt{2}) \cap \mathbb{Q}$ is also $\sqrt{2}$. Here, our set consists only of the *rational* numbers between $-\sqrt{2}$ and $\sqrt{2}$. We use the same idea as above, only now that the choice $s = \sqrt{2} - \frac{\epsilon}{2}$ no longer necessarily works since this might not be rational, as we need it to be in order to be in the set in question. We are saved this time by the fact that $\mathbb{Q}$ is dense in $\mathbb{R}$, which says that for any $\epsilon > 0$ we can for sure find a rational s such that

$$\sqrt{2} - \epsilon < s < \sqrt{2}.$$

There is the same issue as above that we have to be careful if $\epsilon \geq 4\sqrt{2}$, but the same way we got around that before works here. I'll leave it to you to write out a precise proof, mimicking the one for $(-\sqrt{2}, \sqrt{2})$.

Example 4. Denote the set $\left\{\frac{n}{n+1} \mid n \in \mathbb{N}\right\}$ by A. We claim that $\sup A = 1$, which we can guess based on the fact that the fractions above appear to be getting closer and closer to 1 as we plug in larger and larger values of n, or by considering the limit of the given expression as $n \to \infty$. (We'll talk about the relation between limits and supremums soon enough.)

All fractions we are looking at are positive and the numerator is always smaller than the denominator, so all such fractions are certainly smaller than 1. To show that 1 is the least upper bound, again take $\epsilon > 0$. We want a number of the form $\frac{n}{n+1}$ such that

$$1 - \epsilon < \frac{n}{n+1}.$$

Moving some terms around, we can rewrite this as

$$1 - \frac{n}{n+1} < \epsilon, \text{ or } \frac{1}{n+1} < \epsilon.$$

This last expression just comes from writing the left-hand side as a single fraction. Again, we are looking for a value n which makes this true, but now we see that the Archimedean Property of $\mathbb{R}$ precisely says that we can find such a value. Here is our proof.

Proof of claimed supremum. For any $n \in \mathbb{N}$, we have $n < n+1$, so $\frac{n}{n+1} < 1$ and hence 1 is an upper bound of S. To show that 1 is the least upper bound, let $\epsilon > 0$. By the Archimedean Property of $\mathbb{R}$ there exists $N \in \mathbb{N}$ such that $\frac{1}{N} < \epsilon$. But then also

$$\frac{1}{N+1} < \frac{1}{N} < \epsilon.$$

Since $\frac{1}{N+1} = 1 - \frac{N}{N+1}$, this gives

$$1 - \frac{N}{N+1} < \epsilon, \text{ or } 1 - \epsilon < \frac{N}{N+1}.$$

Thus $\frac{N}{N+1}$ is an element of S which is larger than $1-\epsilon$, so we conclude that $\sup A = 1$ as claimed. $\square$

Never underestimate the power of the Archimedean Property!

Lecture 3: Completeness of $\mathbb{R}$

Today we continued talking about supremums and infimums, and introduced the idea that $\mathbb{R}$ is "complete", which means that supremums of nonempty, bounded sets always exist.

Warm-Up. We claim that the supremum of

$$\left\{ \frac{3n^2}{n^2+n-1} \;\middle|\; n \in \mathbb{N} \text{ and } n \geq 2 \right\}$$

is 3. We come up with this value as a result of the fact that the given fraction is always smaller than or equal to 3 (as we will justify in a bit) and if you plug in larger and larger values of 3 the fraction appears to get closer and closer to 3, or by taking the limit of the given fraction as $n \to \infty$. (Again, we will talk about the relation between supremums and limits shortly.)

First, since $n^2+n-1 > n^2$ for $n \geq 2$, $\frac{1}{n^2+n-1} < \frac{1}{n^2}$ so

$$\frac{3n^2}{n^2+n-1} < \frac{3n^2}{n^2} = 3 \text{ for } n \geq 2.$$

Thus 3 is an upper bound of the given set. Now, let $\epsilon > 0$; we must show there is something in the given set which is larger than $3-\epsilon$. That is, we want $N \geq 2$ such that

$$3 - \epsilon < \frac{3N^2}{N^2+N-1}.$$

Rearranging terms, this is the same as

$$3 - \frac{3N^2}{N^2+N-1} < \epsilon, \text{ or } \frac{3N-3}{N^2+N-1} < \epsilon.$$

Since

$$\frac{3N-3}{N^2+N-1} \leq \frac{3N}{N^2+N-1} \leq \frac{3N}{N^2} = \frac{3}{N},$$

choosing N such that $\frac{3}{N} < \epsilon$ (which we can do by the Archimedean Property) gives us what we want. That is, for N such that $\frac{1}{N} < \frac{\epsilon}{3}$, we have

$$3 - \frac{3N^2}{N^2+N-1} = \frac{3N-3}{N^2+N-1} \leq \frac{3N}{N^2} = \frac{3}{N} < \epsilon,$$

so $3 - \epsilon < \frac{3N^2}{N^2+N-1}$ as required.

Example 1. For nonempty subsets A and B of $\mathbb{R}$, define $A+B$ to be the set

$$A + B = \{a + b \mid a \in A \text{ and } b \in B\}.$$

We claim that if A and B have supremums, $\sup(A+B) = \sup A + \sup B$. We give two proofs of this: one using the definition of supremum and one using the alternate characterization in terms of ϵ.

Proof #1. For any $a \in A$ and $b \in B$, we have $a \leq \sup A$ and $b \leq \sup B$ so

$$a + b \leq \sup A + \sup B.$$

This shows that anything in $A+B$ is $\le \sup A + \sup B$, so $\sup A + \sup B$ is an upper bound of $A+B$. Now, suppose that u is any upper bound of $A+B$. Then

$$a + b \le u \text{ for any } a \in A \text{ and } b \in B.$$

For a fixed $b \in B$, this means that

$$a \le u - b \text{ for any } a \in A.$$

Hence for a fixed $b \in B$, $u - b$ is an upper bound of A and thus

$$\sup A \le u - b \text{ for all } b \in B.$$

Rearranging terms, this gives

$$b \le u - \sup A \text{ for any } b \in B.$$

Thus $u - \sup A$ is an upper bound of B so $\sup B \le u - \sup A$. Therefore $\sup A + \sup B \le u$ and we conclude that $\sup A + \sup B$ is the supremum of $A+B$ since it is an upper bound which is smaller than any other upper bound. □

Proof #2. As above, $\sup A + \sup B$ is an upper bound of $A+B$. Let $\epsilon > 0$. By the alternate characterization of supremums, there exists $a \in A$ such that

$$\sup A - \frac{\epsilon}{2} < a$$

and there exists $b \in B$ such that

$$\sup B - \frac{\epsilon}{2} < b.$$

Then

$$\sup A + \sup B - \epsilon = \left(\sup A - \frac{\epsilon}{2}\right) + \left(\sup B - \frac{\epsilon}{2}\right) < a + b,$$

so $a+b$ is an element of $A+B$ which is larger than $\sup A + \sup B$. Hence nothing smaller than $\sup A + \sup B$ can be an upper bound for $A+B$, so $\sup A + \sup B$ is $\sup(A+B)$ as claimed. □

Definitions. Suppose that A is a nonempty subset of $\mathbb{R}$. We say that a real number t is a *lower bound* of A if

$$t \le a \text{ for all } a \in A.$$

We say that A is *bounded below* if it has a lower bound, and A is *bounded* if it is both bounded above and below.

The *infimum* (or "greatest lower bound") of A is a lower bound ℓ of A such that $t \le \ell$ for any other lower bound t of A. We use the notation $\inf A$ for infimums. As with supremums, if a set has an infimum it only has one, so infimums are unique.

Proposition. *Suppose that $A \subset \mathbb{R}$ is nonempty and bounded below. A lower bound ℓ of A is the infimum of A if and only if for any $\epsilon > 0$ there exists $a \in A$ such that $a < \ell + \epsilon$. This is an analog of the alternate characterization of supremums. The part after "if and only if" says that nothing larger than ℓ can be a lower bound of A since we can find something in A smaller than it.*

Completeness Axiom. The *completeness axiom* of $\mathbb{R}$ says that any nonempty set of real numbers which is bounded above has a supremum. So, to show that a set of a real numbers has a supremum all we need to do is show that it is bounded above. We say that $\mathbb{R}$ is "complete". Similarly, any set which is bounded below will have an infimum—I encourage you to think about how this follows directly from the corresponding fact for supremums. The book says a bit more about this.

The fact that $\mathbb{R}$ is complete is a crucially important property, and will underlie many things we do in this course. In particular, it is this property which allows us to visualize $\mathbb{R}$ as a continuous line with no "gaps". We will take this property as a given, but it is natural to think about why it should be true. To justify it precisely, we would have to start with a precise definition (or construction) of $\mathbb{R}$. I will write up some notes about this for those interested, but it is not something we will come back to. We will simply take the completeness axiom as a given and power through.

Example 2. Contrast the completeness property of $\mathbb{R}$ with the following example, which shows that $\mathbb{Q}$ is not complete. The set $S = \sup\{r \in \mathbb{Q} \mid r < \sqrt{2}\}$ of rational numbers is bounded above by the rational number 2 but has no supremum in $\mathbb{Q}$. Of course, $\sqrt{2}$ is also an upper bound of S in $\mathbb{R}$ but it is not an upper bound of S "in" $\mathbb{Q}$ since $\sqrt{2}$ is not rational; this is why I used 2 as an upper bound above. Similarly, of course S has a supremum in $\mathbb{R}$ (which is $\sqrt{2}$), but the point is that this supremum does not exist in $\mathbb{Q}$ itself, which is what it means to say that $\mathbb{Q}$ is not complete. If you try to "draw" $\mathbb{Q}$ as a line you will "gaps" all over the place.

Lecture 4: Sequences

Today we started talking about sequences of real numbers, and what it means for a sequence to converge.

Warm-Up 1. Suppose that A is a non-empty subset of $\mathbb{R}$ which is bounded below. We claim that $\inf A$ exists, and this follows from the completeness axiom of $\mathbb{R}$. Indeed, let L denote the set of all lower bounds of A:

$$L := \{\ell \in \mathbb{R} \mid \ell \text{ is a lower bound of } A\}.$$

This is non-empty since we are assuming that A is bounded below, and any element of A is an upper bound for L. By the completeness axiom, $\sup L$ exists; we claim that $\sup L = \inf A$.

Indeed, fix $a \in A$. Then $\ell \leq a$ for all $\ell \in L$ so a is an upper bound of L. Thus $\sup L \leq a$ for all $a \in A$, so $\sup L$ is a lower bound of A. Now, if ℓ is any other lower bound of A, $\ell \in L$ so $\ell \leq \sup L$. Thus $\sup L$ is $\geq$ any other lower bound of A so $\sup L = \inf A$ as claimed.

Warm-Up 2. Suppose that S and T are subsets of $\mathbb{R}$ such that $s \leq t$ for all $s \in S$ and $t \in T$. Then $\sup S \leq \inf T$.

Indeed, fix $t \in T$. Then $s \leq t$ for all $s \in S$ so t is an upper bound of S. Hence $\sup S \leq t$ for any $t \in T$. This means that $\sup S$ is a lower bound of T so $\sup T \leq \inf T$ as claimed.

Definition of convergent sequence. A *sequence* is simply an infinite list of real numbers:

$$x_1, x_2, x_3, \ldots.$$

We denote such a sequence by (x_n). Note that the book uses $\{x_n\}$ to denote a sequence, which I don't like since it confuses the notion of "sequence" with that of "set". In particular, the sequence (x_n) where $x_n = 1$ for all n consists of an infinite numbers of 1's, but the set $\{x_n\}$ consisting of all terms in the sequence only contains one element, namely the number 1.

We say that (x_n) *converges* to $x \in \mathbb{R}$ if for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$|x_n - x| < \epsilon \text{ for } n \geq N.$$

Intuitively, this definition says that given any interval around x—the given inequality says that x_n is in the interval $(x-\epsilon, x+\epsilon)$—no matter how small, eventually all the terms in the sequence (x_n) will be inside that interval. This rigorously captures the idea that the terms of (x_n) are "approaching" x as $n \to \infty$. We call x the *limit* of the sequence and use the notation $x = \lim_{n\to\infty} x_n$. We also use $x_n \to x$ to denote that (x_n) converges to x.

Remark. If a sequence converges, it can only converge to one thing. In other words, limits of convergent sequences are unique. Check the book for a proof of this, but it justifies us talking about *the* limit of a sequence.

Example 1. Consider the sequence (x_n) where $x_n = 3 - \frac{1}{n} + \frac{2}{n^3} - \frac{3}{n^5}$. We claim that $x_n \to 3$. To show this we let $\epsilon > 0$. We must find an index N such that

$$|x_n - 3| < \epsilon \text{ for } n \geq N.$$

This absolute value equals

$$\left| -\frac{1}{n} + \frac{2}{n^3} - \frac{3}{n^5} \right|,$$

which using the triangle inequality is $\leq \frac{1}{n} + \frac{2}{n^3} + \frac{3}{n^5}$. This expression is smaller than or equal to $\frac{6}{n}$, and now we see that the Archimedean Property gives us the N we need. Here is a full proof:

Proof that $x_n \to 3$. Let $\epsilon > 0$ and choose $N \in \mathbb{N}$ such that $\frac{6}{N} < \epsilon$. Then if $n \geq N$ we have

$$|x_n - 3| = \left| -\frac{1}{n} + \frac{2}{n^3} - \frac{3}{n^5} \right| \leq \frac{1}{n} + \frac{2}{n^3} + \frac{3}{n^5} \leq \frac{1}{n} + \frac{2}{n} + \frac{3}{n} \leq \frac{6}{N} < \epsilon.$$

Thus $x_n \to 3$ as claimed. □

Example 2. Suppose that $x_n = \frac{\sin n}{n}$. This converges to 0, as we now show. The key is that $|\sin n|$ is bounded by 1. Let $\epsilon > 0$ and choose $N \in \mathbb{N}$ such that $\frac{1}{N} < \epsilon$. If $n \geq N$ we have

$$|x_n - 0| = \frac{|\sin n|}{n} \leq \frac{1}{n} \leq \frac{1}{N} < \epsilon.$$

Hence $x_n \to 0$.

Remark. Yes, the Archimedean Property is often useful when showing sequences converge, just as it was when proving that the supremum or infimum of a set was what we said it was. However, note that examples were designed so that the Archimedean Property would apply; it will not always be the case that using this is how you should show a sequence converges. In particular, if you don't have an explicit formula for x_n as we do in these examples, the Archimedean Property may not be all that helpful.

More properties of convergent sequences. Check the book for more basic property of sequences, such as the facts that convergent sequences are always bounded, and that subsequences of convergent sequence are themselves convergent. (A subsequence (x_{n_k}) of a sequence (x_n) is a sequence formed by taking some terms from the original sequence, in the same order in which they appear in the original sequence.)

Relation to supremums and infimums. Suppose that u is an upper bound of a set S. Then $u = \sup S$ if and only if there exists a sequence of elements of S which converges to u. Similarly, given a lower bound t of S, $t = \inf S$ if and only if there exists a sequence of elements of S which converges to t.

Proof. We prove the statement about supremums; the infimum claim is very similar. First suppose that $u = \sup S$. We must construct a sequence (x_n) in S which converges to u. By the alternate characterization of supremums in terms of ϵ's, there exists $x_1 \in S$ such that

$$u - 1 < x_1 \leq u.$$

For the same reason, there exists $x_2 \in S$ such that

$$u - \frac{1}{2} < x_2 \leq u.$$

And so on, for any $n \in \mathbb{N}$ there exists $x_n \in S$ such that

$$u - \frac{1}{n} < x_n \leq u.$$

This constructs a sequence (x_n) in S and we claim that $x_n \to u$. Indeed, let $\epsilon > 0$ and choose $N \in \mathbb{N}$ such that $\frac{1}{N} < \epsilon$. By the choice of the terms x_n, for $n \geq N$ we have

$$|x_n - u| < \frac{1}{n} \leq \frac{1}{N} < \epsilon,$$

so we conclude that $x_n \to u$ as claimed.

Conversely, suppose that there exists a sequence (x_n) in S which converges to u. We use the alternate characterization of supremums. Let $\epsilon > 0$. Since $x_n \to u$ there exists $N \in \mathbb{N}$ such that

$$|x_N - u| < \epsilon.$$

But this inequality implies that

$$u - \epsilon < x_N.$$

Since $x_N \in S$, this shows that $u - \epsilon$ is not an upper bound of S, so u is the least upper bound as claimed. □

Lecture 5: Sequence Limit Theorems

Today we spoke about some well-known limit theorems involving sequences that you would have seen in a calculus course, only now we're interested in how to prove such things.

Warm-Up. Take any real number $x \in \mathbb{R}$. We claim that there is a sequence of rationals (r_n) converging to x. In practice, given an explicit x such a sequence is not hard to find. For instance, consider π whose decimal expansion looks like

$$\pi = 3.14159\ldots.$$

Consider the sequence formed by taking one more decimal place at each step:

$$r_1 = 3,\ r_2 = 3.1,\ r_3 = 3.14,\ r_4 = 3.141, \ldots .$$

Each of these numbers r_n is rational since they have a finite decimal expansion, and they will converge to π. Doing something similar for any real number, given its decimal expansion, will justify our claim.

However, this process depends on the fact that any real number indeed has a decimal expansion, which is not at all obvious. So, instead we give a proof of our claim that doesn't depend on this and uses only the fact that $\mathbb{Q}$ is dense in $\mathbb{R}$. The idea is to first consider $x - 1 < x$, and to use denseness to pick a rational r_1 such that $x - 1 < r_1 < x$. Then do the same for $x - \frac{1}{2} < x$ to get a rational r_2, then for $x - \frac{1}{3} < x$ to get r_3, and so on. The choice of these rationals as ever closer to x will guarantee the resulting sequence converges to x. Here is the proof; this technique of using $x - \frac{1}{n}$ is an example of what's commonly called a "$\frac{1}{n}$-trick".

Proof of claim. For any $n \in \mathbb{N}$ pick a rational $r_n \in \mathbb{Q}$ such that

$$x - \frac{1}{n} < r_n < x,$$

which exists by the denseness of $\mathbb{Q}$ in $\mathbb{R}$. This results in a sequence (r_n) of rational numbers which we claim converges to x. Indeed, let $\epsilon > 0$ and pick $N \in \mathbb{N}$ such that $\frac{1}{N} < \epsilon$. For $n \geq N$ we have

$$|r_n - x| < \frac{1}{n} \leq \frac{1}{N} < \epsilon$$

where the first inequality follows by the way in which r_n was chosen. Thus $r_n \to x$ as claimed. $\square$

Remark. We know from a previous homework problem that the irrationals are also dense in $\mathbb{R}$, so a slight modification of the above argument only choosing r_n to be irrational shows that for any $x \in \mathbb{R}$ there also exists a sequence of irrationals converging to it. This would be harder to show using decimal expansions.

Squeeze Theorem. Suppose we have sequences $(a_n), (b_n)$, and (c_n) such that

$$a_n \leq b_n \leq c_n \text{ for all } n.$$

If the two "outer" sequences a_n and c_n both converge to the same x, then $b_n \to x$ as well. This is no doubt something you probably saw in calculus, but here we are interested in proving this fact. The key is to relate the expression we want to make smaller than ϵ, namely $|b_n - x|$, to the ones we know we can somehow control: $|a_n - x|$ and $|c_n - x|$. Both of these we can make smaller than any positive number, and doing this in the right way will make our original expression smaller than ϵ. Here is the proof, which uses the useful technique of "picking the maximum of two indices."

Proof. Let $\epsilon > 0$. Since $a_n \to x$ we know there exists $N_1 \in \mathbb{N}$ such that

$$|a_n - x| < \epsilon \text{ for } n \geq N_1$$

and since $c_n \to x$ we know there exists $N_2 \in \mathbb{N}$ such that

$$|b_n - x| < \epsilon \text{ for } n \geq N_2.$$

Set $N = \max\{N_1, N_2\}$. Since $a_n \leq b_n \leq c_n$ we have

$$a_n - x \leq b_n - x \leq c_n - x.$$

For $n \geq N_1$ the term on the left is larger than $-\epsilon$ and for $n \geq N_2$ the term on the right is smaller than ϵ. Thus for $n \geq N = \max\{N_1, \mathbb{N}_2\}$ both of these inequalities hold so

$$-\epsilon < a_n - x \leq b_n - x \leq c_n - x < \epsilon.$$

Thus $|b_n - x| < \epsilon$ for $n \geq N$, so $b_n \to x$ as claimed. □

Addition Law. Suppose we have sequences $a_n \to a$ and $b_n \to b$. Then the sequences $(a_n + b_n)$ converges to $a + b$. In other words, assuming both sequences converge we have

$$\lim_{n\to\infty}(a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n.$$

Again the idea of the proof is to find a way to bound $|(a_n + b_n) - (a + b)|$ (the expression we have to make smaller than ϵ in order to show $a_n + b_n \to a + b$) in terms of $|a_n - a|$ and $|b_n - b|$, which we know we can control. This is simple: according to the triangle inequality we have

$$|(a_n + b_n) - (a + b)| = |(a_n - a) + (b_n - b)| \leq |a_n - a| + |b_n - b|.$$

Now we see that if we make the two terms on the right smaller than $\epsilon/2$ (which we know we can do past some indeces), the expression on the left will be smaller than ϵ. This is an example of what's called an "$\frac{\epsilon}{2}$-trick", and again "picking the maximum of indices" makes an appearance. Here is our proof.

Proof. Let $\epsilon > 0$. Since $a_n \to a$ and $b_n \to b$ there exist $N_1, N_2 \in \mathbb{N}$ such that

$$|a_n - a| < \frac{\epsilon}{2} \text{ for } n \geq N_1$$

and

$$|b_n - b| < \frac{\epsilon}{2} \text{ for } n \geq N_2.$$

Then if $n \geq N := \max\{N_1, N_2\}$ we have

$$|(a_n + b_n) - (a + b)| = |(a_n - a) + (b_n - b)| \leq |a_n - a| + |b_n - b| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

We conclude that $a_n + b_n \to a + b$ as claimed. □

Product Law. Supposing that $a_n \to a$ and $b_n \to b$, it follows that $a_n b_n \to ab$. In other words, given that a_n and b_n converge we have

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

In the proof we need to make $|a_n b_n - ab|$ smaller than ϵ. The tricky part is figuring how to bound this expression by the ones ($|a_n - a|$ and $|b_n - b|$) we know something about. Notice that we can rewrite $a_n b_n - ab$ as

$$a_n b_n - ab = a_n b_n - ab_n + ab_n - ab = (a_n - a)b_n + a(b_n - b),$$

since all we did was to add $0 = -ab_n + ab_n$ in the second equality, just in a funny way. This idea of adding "0" to expressions will also be useful to remember. Now the triangle inequality gives

$$|a_n b_n - ab| \le |a_n - a||b_n| + |a||b_n - b|,$$

and we have bounded $|a_n b_n - ab|$ in terms of the absolute values we know how to control. We'll leave this unfinished here, but check the second claim on `http://math.berkeley.edu/~scanez/courses/math104/summer10/Home/Home_files/examples.pdf` to see how to finish it off. This proof is different than the one given in the book, and I think is more enlightening and better at showing how to work with inequalities than the book's version.

Remark. Check the book for other well-known limit laws we didn't explicitly mention.

Lecture 6: Monotone Sequences

Today we spoke about monotone sequences, and the fact that any bounded monotone sequence converges. The point is that, for the first time so far, this gives us a way to show certain sequences converge without knowing ahead of time what their limit will be, as would be required if trying to show they converged using the definition of convergence.

Warm-Up. Suppose that $a_n \to a$ and $a_n \le b$ for all n. Then $a \le b$. This is actually just a special case of the "comparison law" given in the book (the special case where the larger sequence is a constant sequence), but it's useful to look at the proof of this simpler fact. Arguing by contradiction, the idea is that if $b < a$, then we would couldn't have the terms in our sequence a_n bounded by b and at the same time getting closer and closer to a. To make this clear, pick an interval around a which doesn't include b: the definition of convergence says that eventually all terms in our sequence would be in this interval, and such terms would have to be larger than b. Here is the proof.

Proof. By way of contradiction, suppose that $b < a$. Then $a - b > 0$ so we can apply the definition of sequence convergence to this choice of "ϵ". This means there exists $N \in \mathbb{N}$ such that

$$|a - a_n| < a - b \text{ for } n \ge N.$$

Since $a_n \le b < a$, $|a - a_n| = a - a_n$ so this becomes

$$a - a_n < a - b, \text{ or } b < a_n \text{ for } n \ge N.$$

This contradicts the assumption that $a_n \le b$ for all n, so we conclude that $a \le b$. □

Closed intervals are "closed" in $\mathbb{R}$. The Warm-Up (and a similar result when $b \le a_n$ for all n) says the following. Take any sequence inside a closed interval $[a, b]$. If this sequence converges, the thing which it converges to must also be in $[a, b]$. Subsets of $\mathbb{R}$ with this property are said to be *closed* in $\mathbb{R}$. Closedness is a notion you will learn about more in Math 320-2, but I will also use this idea from time to time.

As a contrast, open intervals (a, b) are not closed in $\mathbb{R}$ since a sequence in (a, b) converging to one of the endpoints will not have its limit inside of (a, b) itself. Also, $\mathbb{Q}$ is not closed in $\mathbb{R}$: any sequence of rationals converging to an irrational is an example of a sequence in $\mathbb{Q}$ whose limit is not in $\mathbb{Q}$.

Definition. A sequence (x_n) is said to be *increasing* if

$$x_1 \le x_2 \le x_3 \le \dots$$

and *strictly increasing* if

$$x_1 < x_2 < x_3 < \dots.$$

Note that according to these definitions, a constant sequence is increasing although it is not strictly increasing. Similarly, (x_n) is *decreasing* if

$$x_1 \geq x_2 \geq x_3 \geq \dots$$

and is *stricly decreasing* if

$$x_1 > x_2 > x_3 > \dots.$$

We say that (x_n) is *monotone* if it is either increasing or decreasing.

Example 1. Define the sequence (x_n) recursively by setting

$$x_1 = \sqrt{2},\ x_{n+1} = \sqrt{2 + x_n} \text{ for } n \geq 1.$$

Computing a few terms of this sequence:

$$\sqrt{2},\ \sqrt{2+\sqrt{2}},\ \sqrt{2+\sqrt{2+\sqrt{2}}}, \dots$$

suggests that it is increasing, which can prove using induction; that is, we show that $x_n \leq x_{n+1}$ for all n.

First, $x_1 = \sqrt{2} \leq \sqrt{2+\sqrt{2}} = x_2$ so our claim is true for $n = 1$. Suppose that $x_k \leq x_{k+1}$ for some k. Then

$$x_{k+1} = \sqrt{2 + x_k} \leq \sqrt{2 + x_{k+1}} = x_{k+2}$$

so $x_k \leq x_{k+1}$ implies $x_{k+1} \leq x_{k+2}$. We conclude by induction that $x_n \leq x_{n+1}$ for all n as claimed, so (x_n) is increasing and hence monotone.

Theorem. Any bounded monotone sequence is convergent. The point of this Theorem is that it gives us our first example of a way to show a sequence converges without knowing ahead of time what it converges to.

Proof. We suppose that (x_n) is increasing and bounded above. The proof in the case that (x_n) is decreasing and bounded below will be very similar.

The set $\{x_n \mid n \in \mathbb{N}\}$ containing the terms of our sequence is nonempty and bounded above, so by the completeness property of $\mathbb{R}$ it has a supremum, call it b. We claim that $x_n \to b$. Indeed, let $\epsilon > 0$. By the alternate characterization of supremums there exists $N \in \mathbb{N}$ such that

$$b - \epsilon < x_N \leq b.$$

Since (x_n) is increasing, we know that $x_N \leq x_n$ for $n \geq N$, and thus

$$b - \epsilon < x_n \leq b \text{ for } n \geq N.$$

Hence if $n \geq N$, $|x_n - b| < \epsilon$ so we conclude that (x_n) converges to b. □

Back to Example 1. Going back to the same sequence (x_n) from Example 1, we can now show that it converges and determine its limit. We previously showed it was increasing, so according to the above Theorem if we can also show it is bounded above it will converge. We claim that it is bounded above by 2 and use induction to show this.

First, $x_1 = \sqrt{2} \leq 2$ so our claim holds for $n = 1$. Suppose that $x_k \leq 2$ for some k. Then

$$x_{k+1} = \sqrt{2 + x_n} \leq \sqrt{2 + 2} = 2.$$

Thus $x_k \leq 2$ implies $x_{k+1} \leq 2$ so by induction we conclude that $x_n \leq 2$ for all n. Since (x_n) is bounded and monotone, it converges—let x denote its limit.

To determine x we proceed as follows. First, from the recursive definition of x_n we get

$$x_{n+1}^2 = 2 + x_n.$$

The sequence (x_{n+1}) is the subsequence of (x_n) consisting of all terms except for the first, so (x_{n+1}) also converges to x since (x_n) does. Thus using some limit laws,

$$x_{n+1}^2 \to x \text{ and } 2 + x_n \to 2 + x.$$

However, $x_{n+1}^2 = 2 + x_n$ so since limits of a sequence are unique we must have $x^2 = 2 + x$. Solving for x gives $x = -1, 2$. We can't have the limit of (x_n) equal -1 since all terms x_n are positive, so we must have $x = 2$. Thus (x_n) converges to 2, a fact which is pretty much impossible to show without making use of the fact that x_n is monotone and bounded.

Lecture 7: Bolzano-Weierstrass Theorem and Cauchy Sequences

Today we proved the Bolzano-Weierstrass Theorem and started talking about Cauchy sequences. The Bolzano-Weierstrass will be a pretty important result for us; even though it does not seem all that interesting at first glance, it truly is a cornerstone of analysis.

Warm-Up. Let $x_1 \in \mathbb{R}$ and set $x_{n+1} = \frac{1+x_n}{2}$ for $n \geq 1$. We show that (x_n) converges. The interesting part is that the behavior of (x_n) changes depending on x_1: for $x_1 = 1$ this is a constant sequence, for $x_1 < 1$ this sequence is increasing, and for $x_1 > 1$ it is decreasing. Thus is clearly converges for $x_1 = 1$, and for $x_1 < 1$ or $x_1 > 1$ we will show that (x_n) is monotone bounded. Let as consider only the case $x_1 < 1$ since the case $x_1 > 1$ is similar.

Thus, suppose $x_1 < 1$. First we show that x_1 is bounded above by 1. Indeed, $x_1 \leq 1$ so the base case holds. If $x_k \leq 1$ we have

$$x_{k+1} = \frac{1 + x_k}{2} \leq \frac{1 + 1}{2} = 1.$$

Thus by induction we conclude $x_n \leq 1$ for all n. Now we show (x_n) is increasing; that is, $x_n \leq x_{n+1}$ for all n. For $n = 1$ we have

$$2x_2 = 1 + x_1, \text{ so } x_2 - x_1 = 1 - x_2.$$

Since we already know that $x_2 \leq 1$, this shows that $x_2 - x_1 \geq 0$ so $x_1 \leq x_2$ and our base case holds. If $x_k \leq x_{k+1}$, then

$$x_{k+1} = \frac{1 + x_k}{2} \leq \frac{1 + x_{k+1}}{2} = x_{k+2}.$$

Hence $x_n \leq x_{n+1}$ for all n by induction. Since (x_n) is bounded above and increasing, it converges as claimed.

Proposition. Any sequence has a monotone subsequence.

Proof. Consider the collection of all indices n such that x_n is $\geq$ everything coming after it. There are two possibilities, either there are infinitely many such indices or finitely many.

If there are infinitely many such indices we can list them in increasing order:

$$n_1 < n_2 < n_3 < \dots.$$

Then by the property these indices satisfy we have

$$x_{n_1} \geq x_{n_2} \geq x_{n_3} \geq \dots,$$

giving a decreasing subsequence of (x_n) in this case.

If there are finitely many such indices pick $m_1 \in \mathbb{N}$ larger than them all. Then x_{m_1} is not $\geq$ everything coming after it so there is some x_{m_2} with $m_1 < m_2$ such that $x_{m_1} < x_{m_2}$. Similarly, m_2 is not among the indices considered above so x_{m_2} is not $\geq$ everything coming after it. Hence there is some $m_2 < m_3$ such that $x_{m_2} < x_{m_3}$. Continuing in this manner gives a sequence of indices $m_1 < m_2 < m_3 < \dots$ such that

$$x_{m_1} < x_{m_2} < x_{m_3} < \dots,$$

giving an increasing sequence of (x_n) in this case. Thus either way, (x_n) has a monotone subsequence. □

Bolzano-Weierstrass Theorem. Any bounded sequence has a convergent subsequence.

Proof. Suppose that (x_n) is a bounded sequence. According to the previous proposition, this has a monotone subsequence (x_{n_k}). Since (x_n) is bounded the subsequence (x_{n_k}) is also bounded. Thus (x_{n_k}) is bounded and monotone, so it converges and is thus a convergent subsequence of (x_n). □

Remark. Compare this (relatively) simple proof of the Bolzano-Weierstrass Theorem as opposed to the one in the book. There a certain property of "nested intervals" is used, which is useful to know but seems to complicates matters. The power of the Bolzano-Weierstrass Theorem lies in the fact that it is simple to state and to prove, and yet has far reaching consequences. In particular, it will lead to many nice properties of continuous function which are essential for proving things in calculus. For now, we will use this to prove an important property of what are called "Cauchy sequences" in the coming days.

Closed intervals are "compact". Consider a sequence (x_n) inside a closed interval $[a, b]$. Since $[a, b]$ is bounded so is the sequence (x_n) and hence by the Bolzano-Weierstrass Theorem it has a convergent subsequence (x_{n_k}). Since all terms in this subsequence are in $[a, b]$, we have previously seen that its limit must also be in $[a, b]$. Thus we come to the conclusion that any sequence whatsoever in $[a, b]$ always has a subsequence which converges to something in $[a, b]$. This is what it means to say that $[a, b]$ is a *compact* subset of $\mathbb{R}$.

Compactness is something the book doesn't talk about until later when looking at higher-dimensional spaces, and is something you would learn more about in Math 320-2. However, the concept is useful enough that I will use this notion from time to time. I'll be sure to remind you of what it means when we need it.

Things that aren't compact. The sequence $x_n = \frac{1}{n+1}$ of terms in $(0, 1)$ has no convergent subsequence in $(0, 1)$ (meaning that the limit should also be in this same interval), so $(0, 1)$ is not compact in the above sense. Indeed, no open interval is compact. The set of rationals is also not compact: any sequence of rationals which converges to an irrational will have no convergent subsequence *in* $\mathbb{Q}$.

Definition of a Cauchy sequence. A sequence (x_n) is a *Cauchy* sequence if for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$|x_m - x_n| < \epsilon \text{ for } n, m \geq N.$$

This looks similar to the definition of convergence, but there are two important differences: here there is no mention of a "limit", and the absolute value involved uses *two* terms from the sequence. Intuitively, this definition says that terms in our sequence are getting closer and closer to *each other*, as opposed to getting closer and closer to a limit. Indeed, given any positive ϵ no matter how small, the above condition guarantees that *all* terms of (x_n) past some index are within ϵ apart from each other.

Since the terms of a Cauchy sequence are getting closer and closer to each other, it "feels" like they should indeed be approaching a fixed number; i.e. it makes some intuitive sense to believe that Cauchy sequences do in fact converge. This is true(!), and we'll prove it next time.

Proposition. Convergent sequences are always Cauchy. (This should make sense intuitively: if the terms of a sequence are getting closer and closer to some limit, they should certainly be getting closer and closer to each other.)

Proof. ("$\frac{\epsilon}{2}$-trick") Suppose that (x_n) converges to x and let $\epsilon > 0$. Then there exists $N \in \mathbb{N}$ such that

$$|x_n - x| < \frac{\epsilon}{2} \text{ for } n \geq N.$$

Hence if $m, n \geq N$ we have

$$|x_m - x_n| = |(x_m - x) + (x - x_n)| \leq |x_m - x| + |x - x_n| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus (x_n) is Cauchy as was to be shown. □

Proposition. Cauchy sequences are bounded. (The proof is similar to the proof that convergent sequences are bounded, but since we didn't give that proof in class we'll give this one instead.)

Proof. Suppose that (x_n) is Cauchy. Then for $\epsilon = 1$ there exists $N \in \mathbb{N}$ such that

$$|x_n - x_m| < 1 \text{ for } m, n \geq N.$$

In particular, $|x_n - x_N| < 1$ for $n \geq N$, so $|x_n| < 1 + |x_N|$ for $n \geq N$. Thus all terms in our sequence beyond the Nth one are bounded by $1 + |x_N|$. As for the terms before, there are only finitely many so they are bounded by the largest one among them. Hence all terms in our sequence are bounded by

$$\max\{|x_1|, |x_2|, \ldots, |x_{N-1}|, 1 + |x_N|\},$$

so (x_n) is bounded. □

Lecture 8: Cauchy Sequences and Series

Today we continued talking about Cauchy sequences and proved that any Cauchy sequence in $\mathbb{R}$ converges in $\mathbb{R}$. We also looked at examples of sequences coming from series where the notion of being Cauchy is perhaps the only way to check convergence.

Warm-Up. Suppose that (x_n) is a sequence with property that

$$|x_{n+1} - x_n| < \frac{1}{2^n} \text{ for all } n.$$

We show that (x_n) is Cauchy. The point is that we are only assuming something about the distance between *successive* terms in the sequence, and yet from this we'll be able to say something about the distance between *any* two terms.

For a given $\epsilon > 0$, we want to make $|x_m - x_n| < \epsilon$ past some index. Assume that $m \geq n$ so that x_m is further along in the sequence than x_n. To be sure, say that $m = n + k$ for some $k \geq 0$. Then we can rewrite the expression $x_m - x_n$ by adding and subtracting all intermediate terms:

$$x_{n+k} - x_n = (x_{n+k} - x_{n+k-1}) + (x_{n+k-1} - x_{n+k_2}) + \cdots + (x_{n+2} - x_{n+1}) + (x_{n+1} - x_n).$$

Again, we have just added "0" in an elaborate way. The triangle inequality then gives

$$|x_{n+k} - x_n| \leq |x_{n+k} - x_{n+k-1}| + |x_{n+k-1} - x_{n+k_2}| + \cdots + |x_{n+2} - x_{n+1}| + |x_{n+1} - x_n|,$$

which is good since these remaining absolute values are precisely the ones we know something about. Using the bounds given in the setup, we get

$$|x_{n+k} - x_n| < \frac{1}{2^{n+k-1}} + \frac{1}{2^{n+k-2}} + \cdots + \frac{1}{2^{n+1}} + \frac{1}{2^n}.$$

Now, this sum of powers of $\frac{1}{2}$ is actually $\leq \frac{1}{2^{n-1}}$; this can be proved using induction, but the idea is clear is you consider a special case such as:

$$\begin{aligned}
\frac{1}{2^6} + \frac{1}{2^5} + \frac{1}{2^4} + \frac{1}{2^3} &\leq \frac{1}{2^5} + \frac{1}{2^5} + \frac{1}{2^4} + \frac{1}{2^3} \\
&\leq \frac{1}{2^4} + \frac{1}{2^4} + \frac{1}{2^3} \\
&\leq \frac{1}{2^3} + \frac{1}{2^3} \\
&= \frac{1}{2^2}.
\end{aligned}$$

Thus we get $|x_m - x_n| < \frac{1}{2^{n-1}}$, and by picking a large enough index we can make $\frac{1}{2^{n-1}} < \epsilon$. The full proof is given in the book as Example 2.30.

Remark. The same fact is true if we replace $\frac{1}{2}$ by any constant $0 < c < 1$. That is, if $0 < c < 1$ and a sequence (x_n) satisfies

$$|x_{n+1} - x_n| < c^n \text{ for all } n,$$

then (x_n) is Cauchy. However, note that knowing only that successive terms get closer and closer does NOT necessarily imply the sequence is Cauchy—what matters is *how* quickly successive terms are getting closer and closer. For instance, it is not true that a sequence satisfying

$$|x_{n+1} - x_n| < \frac{1}{n} \text{ for all } n$$

will have to be Cauchy. An example of this is the sequence where

$$x_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

This satisfies $|x_{n+1} - x_n| < \frac{1}{n}$ and yet (x_n) is not Cauchy. (This particular sequence is the *sequence of partial sums* of the *harmonic* series: $\sum_{k=1}^{\infty} \frac{1}{k}$. You might remember from calculus that this series does not converge, which is why (x_n) is not Cauchy.)

Proposition. Suppose that (x_n) is a Cauchy sequence which has a convergent subsequence (x_{n_k}) with limit x. Then (x_n) converges to x as well. Note that in general a sequence can have many convergent subsequences and yet not be convergent itself; the point here is that if we know the overall sequence is already Cauchy, the existence of at least one convergent subsequence is enough to tell us the entire sequence does converge.

Proof. Let $\epsilon > 0$. Since (x_n) is Cauchy there exists $N \in \mathbb{N}$ such that

$$|x_m - x_n| < \frac{\epsilon}{2} \text{ for } m, n \geq N.$$

Since the subsequence (x_{n_k}) converges to x there exists $N_1 \in \mathbb{N}$ such that

$$|x_{n_k} - x| < \frac{\epsilon}{2} \text{ for } k \geq N_1.$$

(Note that k is the number keeping track of the index in the subsequence $x_{n_1}, x_{n_2}, x_{n_3}$, etc. and n_k is the corresponding index in the original sequence.)

Pick some term x_{n_K} in the subsequence which is beyond x_N and $x_{n_{N_1}}$; in the other words, x_{n_K} is a term in the subsequence far enough along in it and in the original sequence to guarantee that both inequalities above hold. Then if $n \geq N$ we have

$$|x_n - x| = |x_n - x_{n_K} + x_{n_K} - x| \leq |x_n - x_{n_K}| + |x_{n_K} - x| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

so (x_n) converges to x as claimed. □

Remark. We emphasize the idea of the proof. We want to bound $|x_n - x|$ by things we know how to control, and the only things we know something about are $|x_{n_k} - x|$ and $|x_m - x_n|$. By going out far enough along in the original sequence and given subsequence both of these can be made smaller than whatever we'd like.

Theorem. Any Cauchy sequence in $\mathbb{R}$ converges in $\mathbb{R}$. After the above proposition and previous facts we've seen, this has a simple proof which is only a few lines long. So hopefully now we see how things we've done are starting to come together to build up more and more theory.

Remark. This Theorem is important: it gives us a new way to show that a sequence converges without knowing ahead of time what the limit will be. Indeed, in most cases where this Theorem is useful the actual limit of the sequence is pretty much impossible to compute directly, and yet we know that it must exist.

This Theorem gives us another sense of what it means to say that $\mathbb{R}$ is "complete": in other contexts, the fact that Cauchy sequences converge is taken as the definition of what it means for a space to be complete.

Proof of Theorem. Suppose that (x_n) is Cauchy. Then (x_n) is bounded, so the Bolzano-Weierstrass Theorem implies that it has a convergent subsequence (x_{n_k}) with some limit x. The previous proposition then says that (x_n) must itself converge to x, so we are done. $\square$

Contrast with $\mathbb{Q}$. The rationals are not complete in the sense that Cauchy sequences in $\mathbb{Q}$ do not always converge in $\mathbb{Q}$. Indeed, take any sequence of rationals which converges to an irrational. Such a sequence will be Cauchy (since it converges in the larger space $\mathbb{R}$), but will not converge "in" $\mathbb{Q}$ since its limit is not in $\mathbb{Q}$.

Example. Let (x_n) be the sequence defined by

$$x_n = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{(-1)^{n+1}}{n}.$$

We claim that (x_n) converges, and will show this by showing that (x_n) is Cauchy. This sequence is actually related to the so-called *alternating harmonic series*:

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

Indeed, recall what it means for a series to converge: $\sum_{k=1}^{\infty} a_k$ converges if the sequence (s_n) defined by

$$s_n = \sum_{k=1}^{n} = a_1 + a_2 + \cdots + a_n$$

converges. This sequence is called the *sequence of partial sums* of the series $\sum_{k=1}^{\infty} a_k$, and the upshot is that questions about series convergence are just questions about convergence of a special type of sequence. The sequence (x_n) in this example is just the sequence of partial sums of the alternating harmonic series. (Series are covered in Chapter 6 of our book and will be covered in detail next quarter; for our purposes, the only way in which we might see series come up is via their partials sums, giving us examples of sequences on which we can further test our techniques.)

To show (x_n) is Cauchy we must show that we can make $|x_m - x_n|$ smaller than ϵ past some index. Suppose that $m \geq n$. Then the expression for x_m contains the expression for x_n plus some more terms. Indeed, we have

$$x_m - x_n = \frac{(-1)^{n+2}}{n+1} + \frac{(-1)^{n+3}}{n+2} + \cdots + \frac{(-1)^m}{m-1} + \frac{(-1)^{m+1}}{m}.$$

For now, assume that $m - n$ is even so that there are an even number of terms above and group them in pairs:

$$x_m - x_n = \left(\frac{(-1)^{n+2}}{n+1} + \frac{(-1)^{n+3}}{n+2} \right) + \cdots + \left(\frac{(-1)^m}{m-1} + \frac{(-1)^{m+1}}{m} \right). \tag{1}$$

Focus on an individual pair, say the first one:

$$\frac{(-1)^{n+2}}{n+1} + \frac{(-1)^{n+3}}{n+2}.$$

Depending on whether n is even or odd, either the first term here will be negative or the second one will be, but either way one of these is negative and the other positive. Thus the absolute value of this expression is precisely

$$\frac{1}{n+1} - \frac{1}{n+2},$$

which is indeed positive since the first fraction is larger than the second. That is, either

$$\frac{(-1)^{n+2}}{n+1} + \frac{(-1)^{n+3}}{n+2} = \frac{1}{n+1} - \frac{1}{n+2} \text{ or } -\frac{1}{n+1} + \frac{1}{n+2},$$

but either way the absolute value is as we claim. This is true for any of the pairs grouped above so the triangle inequality gives

$$\begin{aligned} |x_m - x_n| &= \left| \left(\frac{(-1)^{n+2}}{n+1} + \frac{(-1)^{n+3}}{n+2} \right) + \cdots + \left(\frac{(-1)^m}{m-1} + \frac{(-1)^{m+1}}{m} \right) \right| \\ &\leq \left| \frac{(-1)^{n+2}}{n+1} + \frac{(-1)^{n+3}}{n+2} \right| + \cdots + \left| \frac{(-1)^m}{m-1} + \frac{(-1)^{m+1}}{m} \right| \\ &= \left(\frac{1}{n+1} - \frac{1}{n+2} \right) + \cdots + \left(\frac{1}{m-1} - \frac{1}{m} \right). \end{aligned}$$

Now regroup these terms and rewrite this expression as

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

Each of the terms in parentheses is positive, so this entire expression is a positive number, namely $\frac{1}{n+1}$, minus a bunch of other positive numbers. Thus this entire expression is $\leq \frac{1}{n+1}$, so we finally end up with the bound

$$|x_m - x_n| \leq \frac{1}{n+1} \text{ when } m-n \text{ is even.}$$

When $m - n$ is odd we can't do the grouping into pairs we did in equation (1) since there will be one term left over, so instead we regroup pairs *after* the first term:

$$x_m - x_n = \frac{(-1)^{n+2}}{n+1} + \left(\frac{(-1)^{n+3}}{n+2} + \frac{(-1)^{n+4}}{n+3} \right) + \cdots + \left(\frac{(-1)^m}{m-1} + \frac{(-1)^{m+1}}{m} \right).$$

Taking the absolute value of all of this and using the triangle inequality gives us

$$|x_m - x_n| \leq \frac{1}{n+1} + \left(\frac{1}{n+2} - \frac{1}{n+3} \right) + \cdots + \left(\frac{1}{m-1} - \frac{1}{m} \right)$$

in a similar manner as before. Regrouping things again as before will imply that

$$|x_m - x_n| \leq \frac{1}{n+1} + \frac{1}{n+2} \text{ when } m-n \text{ is odd.}$$

Thus in either case ($m - n$ is even or odd), an application of the Archimedean Property now gives us a way to make the resulting expressions smaller than ϵ, and we end up with a proof (after written out formally) that (x_n) is indeed Cauchy as claimed. Thus, it converges, and trying to show this in any other way will be pretty much impossible. Nice example.

Lecture 9: Lim Sup and Lim Inf

Today we spoke about the notion of the "limit superior" and "limit inferior" of a sequence, which ends up giving us another way to check that sequences converge.

Warm-Up. Fix $x \in \mathbb{R}$. We show that the series $\sum_{k=1}^{\infty} \frac{\sin(kx)+\cos(kx)}{5^k}$ converges knowing that $\sum_{k=1}^{\infty} \frac{1}{5^k}$ converges. As usual for series, we look at the sequence of partial sums:

$$s_n = \sum_{k=1}^{n} \frac{\sin(kx)+\cos(kx)}{5^k}$$

and show that this sequence is Cauchy. For $m \geq n$ we have

$$s_m - s_n = \frac{\sin[(n+1)x] + \cos[(n+1)x]}{5^{n+1}} + \cdots + \frac{\sin(mx)+\cos(mx)}{5^m}.$$

In absolute value, the numerators of each these terms is bounded by 2 so the absolute value of the entire expression is bounded by

$$\frac{2}{5^{n+1}} + \cdots + \frac{2}{5^m} = 2\left(\frac{1}{5^{n+1}} + \cdots + \frac{1}{5^m}\right).$$

This final sum is precisely $y_m - y_n$ where (y_n) is the sequence of partial sums of the series we already know converges, so we are in business. We can make $|y_m - y_n|$ smaller than anything we'd like since we know (y_n) is Cauchy, and doing this in the right way makes $|s_m - s_n|$ smaller than ϵ. Here's our proof.

Proof. Let $\epsilon > 0$. Since $\sum_{k=1}^{\infty} \frac{1}{5^k}$ converges, its sequence of partial sums (y_n) is Cauchy so there exists $N \in \mathbb{N}$ such that

$$|y_m - y_n| < \frac{\epsilon}{2} \text{ for } m, n \geq N.$$

Denoting the sequence of partial sums of our series by (s_n), we then have for $m \geq n \geq N$:

$$\begin{aligned}
|s_m - s_n| &= \left| \frac{\sin[(n+1)x] + \cos[(n+1)x]}{5^{n+1}} + \cdots + \frac{\sin(mx)+\cos(mx)}{5^m} \right| \\
&\leq \left| \frac{\sin[(n+1)x] + \cos[(n+1)x]}{5^{n+1}} \right| + \cdots + \left| \frac{\sin(mx)+\cos(mx)}{5^m} \right| \\
&\leq \frac{2}{5^{n+1}} + \cdots + \frac{2}{5^m} \\
&= 2\left(\frac{1}{5^{n+1}} + \cdots + \frac{1}{5^m}\right) = 2|y_m - y_n| < 2\left(\frac{\epsilon}{2}\right) = \epsilon.
\end{aligned}$$

Thus (s_n) is Cauchy, so $\sum_{k=1}^{\infty} \frac{\sin(kx)+\cos(kx)}{5^k}$ converges as claimed. □

Definitions. Given a (bounded) sequence (x_n), we define its *limit superior*, denoted $\limsup_{n\to\infty} x_n$, to be

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \left(\sup_{k\geq n} x_k \right)$$

and its *limit inferior*, denoted $\liminf_{n\to\infty} x_n$, to be

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \left(\inf_{k\geq n} x_k \right).$$

Some clarifications are in order. The notation $\sup_{k\geq n} x_n$ means we are taking the supremum of all terms in our sequence starting at x_n and beyond. The idea is the following: define b_1 to be the supremum of all terms in our sequence, b_2 to be the supremum of all terms starting at x_2, b_3 the supremum of all terms starting at x_3, and so on:

$$b_n = \sup\{x_n, x_{n+1}, x_{n+2}, \ldots\}.$$

The lim sup of x_n is the limit of this sequence b_n: $\limsup_{n\to\infty} x_n = \lim_{n\to\infty} b_n$, so the lim sup is a "limit of supremums". The lim inf is thought of in a similar way only using infimums instead of supremums.

Example 1. Define $x_n = 3 + (-1)^n + \frac{1}{n}$. We compute $\limsup x_n$ and $\liminf x_n$. Our sequence looks like:

$$2 + \frac{1}{1},\ 4 + \frac{1}{2}, 2 + \frac{1}{3}, 4 + \frac{1}{4}, 2 + \frac{1}{5}, 4 + \frac{1}{5}, \ldots.$$

Using the notation b_n from above, we have

$$\begin{aligned}
b_1 &= \text{supremum of all terms } = 4 + \frac{1}{2} \\
b_2 &= \text{supremum of all terms starting at the second } = 4 + \frac{1}{2} \\
b_3 &= \text{supremum of all terms starting at the third } = 4 + \frac{1}{4} \\
b_4 &= \text{supremum of all terms starting at the fourth } = 4 + \frac{1}{4},
\end{aligned}$$

and in general

$$b_n = \begin{cases} 4 + \frac{1}{n} & \text{if } n \text{ is even} \\ 4 + \frac{1}{n+1} & \text{if } n \text{ is odd} \end{cases}.$$

This sequence (b_n) converges to 4, so $\limsup x_n = 4$. Using c_n to denote the infimum of all terms starting at x_n, we have

$$\begin{aligned}
c_1 &= \text{infimum of all terms } = 2 \\
c_2 &= \text{infimum of all terms starting at the second } = 2 \\
c_3 &= \text{infimum of all terms starting at the third } = 2
\end{aligned}$$

and in general $c_n = 2$ for all n. Thus $c_n \to 2$, so $\liminf x_n = 2$.

Example 2. Define $y_n = \frac{(-1)^n}{n}$, so our sequence is

$$-1, \frac{1}{2}, -\frac{1}{3}, \frac{1}{4}, -\frac{1}{5}, \frac{1}{6}, -\frac{1}{7}, \ldots.$$

Then using the same notation as before,

$$b_1 = \frac{1}{2},\ b_2 = \frac{1}{2},\ b_3 = \frac{1}{4},\ b_4 = \frac{1}{4},$$

and in general

$$b_n = \begin{cases} \frac{1}{n} & \text{if } n \text{ is even} \\ \frac{1}{n+1} & \text{if } n \text{ is odd} \end{cases}.$$

Thus $b_n \to 0$ so $\limsup y_n = 0$. Also,

$$c_1 = -1,\ c_2 = -\frac{1}{3},\ c_3 = -\frac{1}{3},\ c_4 = -\frac{1}{5},$$

and in general

$$c_n = \begin{cases} -\frac{1}{n} & \text{if } n \text{ is odd} \\ -\frac{1}{n+1} & \text{if } n \text{ is even} \end{cases}.$$

Thus $c_n \to 0$, so $\liminf y_n = 0$.

Remark. Note that the sequence in our first example had $\limsup \neq \liminf$ while in the second example we had $\limsup = \liminf$. This is no accident, and is a reflection of the fact that the sequence in Example 1 does not converge while the one in Example 2 does.

Existence of Lim Sup and Lim Inf. Say that (x_n) is bounded. Then for sure all the supremums and infimums used in the definitions of $\limsup x_n$ and $\liminf x_n$ actually exist. Note also that

$$b_1 \geq b_2 \geq b_3 \geq \dots$$

since at each step we take the supremum of a set with one fewer element, while

$$c_1 \leq c_2 \leq c_3 \leq \dots$$

for a similar reason. Thus (b_n) is a decreasing and bounded sequence, and so always converges, while (c_n) is increasing and bounded, so it converges as well. Hence lim sup and lim inf of a bounded sequence *always* exist! This is one of the most useful things about these concepts: they exist even for a sequence which does not converge.

When (x_n) is not bounded above we say that $\limsup x_n = \infty$ and when (x_n) is not bounded below we say $\liminf x_n = -\infty$. We won't really have the need to consider such things in our class. It is true that a sequence (x_n) is bounded if and only if $\limsup x_n$ and $\liminf x_n$ both exist and are finite.

Theorem. Suppose that (x_n) is a bounded sequence. Then there exists a subsequence (x_{n_k}) converging to $\limsup x_n$, and there exists a subsequence converging to $\liminf x_n$.

The proof of lim sup case is in the book. Originally I outlined the idea behind the proof in class, but in reality the book's proof is not any more complicated than what I outlined. The key idea is, for each k, to pick a term x_{n_k} in our sequence such that

$$b_k - \frac{1}{k} < x_{n_k} \leq b_k,$$

which can do using the alternate characterization of supremums. The tricky part is to do this in a way which guarantees that what we get: $x_{n_1}, x_{n_2}, \dots$ is indeed a *subsequence* of (x_n), meaning we want x_{n_1} to come before the term x_{n_2} in the original sequence, which should come before x_{n_3} in the original sequence, and so on. This is done by, after having chosen x_{n_1}, picking the next term x_{n_2}, where $n_2 \geq n_1 + 1$ in order to guarantee we're further along in the original sequence than x_{n_1}, to satisfy

$$b_{n_1+1} - \frac{1}{2} < x_{n_2} \leq b_{n_1+1},$$

and then continuing in a similar manner for the rest of the terms.

Alternate Characterization of Lim Sup and Lim Inf. Given a sequence (x_n), consider the set S of all possible limits of convergent subsequences of (x_n):

$$S = \{x \in \mathbb{R} \mid \text{there exists a subsequence } (x_{n_k}) \text{ of } (x_n) \text{ converging to } x\}.$$

So, take every possible convergent subsequence of (x_n) and throw its limit into S. For instance, the Theorem above says that $\limsup x_n$ and $\liminf x_n$ both belong to S. It is true that $\limsup x_n$ is actually the largest thing in S and $\liminf x_n$ the smallest thing:

$$\limsup x_n = \sup S \text{ and } \liminf x_n = \inf S.$$

For example, if (x_n) converges, say to x, then any subsequence of (x_n) will also converge to x. Thus in this case S contains only a single element: $S = \{x\}$. Hence $\sup S = x = \inf S$, so the lim sup and lim inf of a convergent sequence both equal the limit of the original sequence.

Theorem. A sequence (x_n) converges if and only if $\limsup x_n = \liminf x_n$.

Proof. The forward direction, that (x_n) converges implies its lim sup and lim inf are the same, is justified in the second paragraph in the alternate characterization above. So we need only prove the other direction. To this end, suppose that

$$\limsup x_n = \liminf x_n.$$

Note that for any n,

$$\inf_{k \geq n} x_k \leq x_n \leq \sup_{k \geq n}$$

since x_n is itself one of the numbers we are taking the infimum and supremum of. The left side converges to $\liminf x_n$ and the right side to $\limsup x_n$, so the squeeze theorem tells us that x_n converges to the common value $\liminf x_n = \limsup x_n$ as well. ☐

Remark. Note the simplicity of the above proof as compared to the one in the book which, although not really too complex, isn't as straightforward.

Lecture 10: Limits of Functions

Today we started talking about limits of functions, both in terms of sequences and in terms of so-called "ϵ-δ" arguments.

Warm-Up 1. Suppose that (x_n) is a sequence such that $\limsup x_n < r$ for some $r \in \mathbb{R}$. We claim that then $x_n < r$ for $n >> 0$. (This notation means that $x_n < r$ for *large enough* n; i.e. there exists $N \in \mathbb{R}$ so that this inequality is true for $n \geq N$.) Indeed, recall that $\limsup x_n$ is the limit of the sequence whose n-th term is

$$b_n := \sup_{k \geq n} x_k.$$

By what it means for this sequence to converge, for $\epsilon = r - \limsup x_n > 0$ there exists $N \in N$ such that

$$b_n \in (\limsup x_n - \epsilon, \limsup x_n + \epsilon) \text{ for } n \geq N.$$

In particular, since $\limsup x_n + \epsilon = r$, this means that

$$x_n \leq b_n < r \text{ for } n \geq N,$$

which is what we want.

Warm-Up 2. Suppose that (x_n) and (y_n) are bounded sequences. Then

$$\liminf x_n + \liminf y_n \leq \liminf(x_n + y_n).$$

Indeed, fix n. Then for any $\ell \geq n$ we have

$$\inf_{k\geq n} x_k + \inf_{k\geq n} y_k \leq x_\ell + y_\ell$$

since the first infimum on the left is $\leq x_\ell$ and the second is $\leq y_\ell$. Thus the sum on the left is a lower bound for the set of all $x_\ell + y_\ell$ for $\ell \geq n$, so

$$\inf_{k\geq n} x_k + \inf_{k\geq n} y_k \leq \inf_{\ell\geq n}(x_n + y_n).$$

Now, the sequences on the left converge to $\liminf x_n$ and $\liminf y_n$ respectively, so

$$\lim_{n\to\infty}(\inf_{k\geq n} x_k + \inf_{k\geq n} y_k) = \lim_{n\to\infty}\inf_{k\geq n} x_k + \lim_{n\to\infty}\inf_{k\geq n} y_k = \liminf x_n + \liminf y_n.$$

The sequence on the right converges to $\liminf(x_n + y_n)$, so after taking limits of both sides of the inequality above we find that

$$\liminf x_n + \liminf y_n \leq \liminf(x_n + y_n)$$

as claimed.

Note that the above inequality is not necessarily an equality. For instance, taking $x_n = (-1)^n$ and $y_n = (-1)^{n+1}$ gives an example where the inequality is strict: the left hand side is -2 and the right side is 0.

Definition of Limit, ϵ-δ version. Take a function $f : E \to \mathbb{R}$ and some $a \in \mathbb{R}$. (Here, E is some subset of $\mathbb{R}$.) We say that the *limit* of f as x approaches a is L, denoted $\lim_{x\to a} f(x) = L$, if for any $\epsilon > 0$ there exists $\delta > 0$ such that

$$\text{if } 0 < |x - a| < \delta, \text{ then } |f(x) - L| < \epsilon.$$

This definition says the following: given any measure ϵ of how close we want to end up to L, there is some measure δ which guarantees that if $x \neq a$ is within δ from a, then $f(x)$ is indeed within ϵ from L. Another way of saying it: given any interval around L no matter how small, there is some small interval around a such that anything apart from a in this latter interval is sent into the former interval.

Note also that according to this definition, f need not even be defined at a since the given condition only requires something of x satisfying $0 < |x - a|$, meaning a is not among the x we consider.

Definition of Limit, sequence version. We say that $\lim_{x\to a} f(x) = L$ if for any sequence (x_n) of terms $x_n \neq a$ which converges to a, the sequence $f(x_n)$ converges to L.

Remarks. These two characterizations of limits are equivalent, which we will soon see. The first is more appealing graphically, especially when thinking of it in terms of intervals, and the second better matches up with our intuition as to what "limit" should mean.

Example. We show that $\lim_{x\to 1}(x^2 - 9) = -8$. This is easy to justify in terms of sequences: if $x_n \to 1$ with $x_n \neq 1$, then $x_n^2 \to 1^2 = 1$ and $x_n^2 - 9 \to 1 - 9 = -8$.

Now, justifying this in terms of ϵ's and δ's is more involved but important to go through since ϵ-δ arguments will show up all over the place from now on, and the more practice we get with them the better off we'll be. Let $\epsilon > 0$. We want $\delta > 0$ such that

$$0 < |x - 1| < \delta \text{ implies } |(x^2 - 9) - (-8)| = |x^2 - 1| < \epsilon.$$

As usual, we start with the inequality we want to end up with and use various bounds to determine what δ should be. We first have

$$|x^2 - 1| = |x - 1||x + 1|,$$

which is good since we know something about bounding $|x - 1|$; indeed, the x's we will eventually consider have the property that $|x - 1| < \delta$, so using this we get

$$|x^2 - 1| < \delta|x + 1|.$$

Now, here is our first (bad) attempt: pick $\delta = \frac{\epsilon}{|x+1|}$. then

$$|x^2 - 1| < \delta|x + 1| = \epsilon,$$

which is we want, correct? NO!!! The problem is that δ should NOT depend on x since we don't know what x's to consider until *after* we've chosen δ. The definition says that for any $\epsilon > 0$ there is some $\delta > 0$ such that for any x satisfying some inequality something happens, so the values of x which we need to test these inequalities on depend on the δ we're saying will work. Thus, we can't have δ depending on x since x actually depends on δ! In this case, our choice of δ can actually only depend on ϵ.

So, back to the drawing board. We need to find a way to bound $|x + 1|$ in a manner which doesn't depend on x anymore, and then we can find the δ we need. Here is one thing we can do: $|x - 1| < \delta$ implies $|x| < \delta + 1$, so we can bound $|x + 1|$ by

$$|x + 1| \leq |x| + 1 < \delta + 2.$$

(To see that $|x - 1| < \delta$ implies $|x| < \delta + 1$, use the *reverse triangle inequality*: $|a| - |b| \leq |a - b|$.) Using this, we now have

$$|x^2 - 1| = |x - 1||x + 1| < \delta(\delta + 2).$$

Setting this equal to ϵ and solving for δ gives a δ which will work... almost: we would still have to verify that the δ we find is indeed positive. There are two values of δ satisfying $\delta(\delta + 2) = \epsilon$, and one of them is positive so we're okay in this case, but notice that if we had ended up with a more complicated expression we might be in trouble: here, we know there is a δ satisfying $\delta(\delta + 2) = \epsilon$ since we can use the quadratic formula to solve for δ, and then we can take the resulting expression for δ and check directly that $\delta > 0$. But, this wouldn't work if we had ended up with δ wanting to satisfy something like

$$\delta^3 - 3\delta^2 + \delta - 2 = \epsilon.$$

In this case, it is not clear that there is a δ satisfying this equation, let alone that $\delta > 0$.

The point is that while what we did above works in this case, the technique does not generalize so easily. Instead, we do the following. The whole reason why we ended up with a complicated expression for δ in the end is that we bounded $|x + 1|$ by $\delta + 2$—to avoid this, we use a simpler bound. Indeed, suppose that our x actually satisfied something like

$$|x - 1| < 1.$$

Of course, we don't know whether this is the case or not yet, but for now let's ignore this detail. For such an x we would have $|x| < 2$ so $|x + 1| < 3$, and we get the simpler bound

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

Now we can take $\delta = \frac{\epsilon}{3}$ and we avoid the issues we had earlier. All good right? Almost, except for the fact that we made the additional assumption $|x-1| < 1$ in the course of coming up with this choice of δ. However, depending on ϵ it may be that

$$|x-1| < \delta = \frac{\epsilon}{3}$$

is not enough to guarantee that x also satisfies $|x-1| < 1$ as we need. But, there is an easy way out: we reconsider our choice of δ and choose δ to be whichever of

$$1 \text{ or } \frac{\epsilon}{3}$$

is smaller! Indeed, setting $\delta = \min\{1, \frac{\epsilon}{3}\}$ we have $\delta > 0$ and for any x satisfying $0 < |x-1| < \delta$, x also satisfies $|x-1| < 1$ so the bound $|x+1| < 3$ we used above holds and all is right with the world. This is a technique you see often in ϵ-δ arguments, so know it well. Here is what our final proof would look like (finally!).

Proof of Claimed Limit. Let $\epsilon > 0$ and set $\delta = \min\{1, \frac{\epsilon}{3}\} > 0$. Suppose that $0 < |x-1| < \delta$. Since $\delta \leq 1$, we also have

$$|x-1| < 1, \text{ so } |x| < 2 \text{ and thus } |x+1| \leq |x| + 1 < 3.$$

Thus

$$|(x^2 - 9) - (-8)| = |x^2 - 1| = |x+1||x-1| < 3\delta \leq \epsilon$$

since $\delta \leq \frac{\epsilon}{3}$. We conclude that $\lim_{x \to 1}(x^2 - 9) = -8$ as claimed. □

Remark. I think my explanation in the Example above is pretty epic, so make sure you follow it well. Arguments involving ϵ's and δ's are tough, and the only way we get comfortable with them is through practice, practice, and more practice.

Lecture 11: More on Limits of Functions

Today we continued talking about limits of functions, looking at more examples and showing why the sequence definition and ϵ-δ definition are equivalent.

Warm-Up. We show that for any $a \in \mathbb{R}$, $\lim_{x \to a} x^3 = a^3$. Let $\epsilon > 0$. We want $\delta > 0$ such that

$$\text{if } 0 < |x-a| < \delta, \text{ then } |x^3 - a^3| < \epsilon.$$

First, we need to somehow express or bound the expression we want to make $< \epsilon$ in terms of $|x-a|$ since this is the only absolute value we know something about. In this case, we have

$$|x^3 - a^3| = |x-a||x^2 + ax + a^2|.$$

The first part is $< \delta$. To find a simple bound for the second part, we use the idea we've used previously and assume for the time being that $|x-a| < 1$. (Later we will shrink δ if necessary in order to guarantee that this holds.) Then

$$|x| < 1 + |a|, \text{ so } |x^2 + ax + a^2| \leq |x|^2 + |a||x| + |a|^2 < (1+|a|)^2 + |a|(1+|a|) + |a|^2.$$

Thus we get

$$|x^3 - a^3| < \delta[(1+|a|)^2 + |a|(1+|a|) + |a|^2],$$

so picking $\delta \le \epsilon/[(1+|a|)^2 + |a|(1+|a|) + |a|^2]$ gives us what we want. (Note that this expression is defined since the denominator is nonzero and that it is positive.) Finally, in order to guarantee that $|x-a| < 1$, we pick our final δ to be whichever of 1 or this expression involving ϵ is smaller. Here is our final proof.

Proof of Claimed Limit. Let $\epsilon > 0$ and set

$$\delta = \min\left\{1, \frac{\epsilon}{(1+|a|)^2 + |a|(1+|a|) + |a|^2}\right\},$$

which is positive. Suppose that $0 < |x-a| < \delta$. Since $\delta \le 1$ we have $|x-a| < 1$ so

$$|x| < 1 + |a|, \text{ and thus } |x^2 + ax + a^2| \le |x|^2 + |a||x| + |a|^2 < (1+|a|)^2 + |a|(1+|a|) + |a|^2.$$

Hence

$$|x^3 - a^2| = |x-a||x^2 + ax + a^2| < \delta[(1+|a|)^2 + |a|(1+|a|) + |a|^2].$$

By the choice of δ, this expression is smaller than or equal to ϵ, so $|x^3 - a^3| < \epsilon$ and we conclude that $\lim_{x\to a} x^3 = a^3$ as claimed. □

Example 1. We claim that $\lim_{x\to 0} \frac{1}{x}$ does not exist. Indeed, letting $f(x) = \frac{1}{x}$, the sequence $x_n = \frac{1}{n}$ converges to 0 but $f(x_n) = n$ does not converge, so the sequence definition of limit cannot be satisfied. In general, to show that $\lim_{x\to a} f(x) \ne L$ using sequences requires an example of a sequence (x_n) with $x_n \ne a$ and converging to a such that $f(x_n)$ does not converge to L.

We can also see that this limit does not exist using ϵ's and δ's. To be clear, let $L \in \mathbb{R}$; we show that $\lim_{x\to 0} \frac{1}{x} \ne L$, so no real number can be the limit of this function as $x \to 0$. Negating the ϵ-δ definition of limit in this case requires us to show that

$$\exists\, \epsilon > 0 \text{ such that } \forall\, \delta > 0\ \exists\, x \text{ satisfying } 0 < x < \delta \text{ but } \left|\frac{1}{x} - L\right| \ge \epsilon.$$

We claim that $\epsilon = 1$ works. Graphically this is clear: no matter what interval we take around 0, no matter how small, we can always find some x in this interval such that $\frac{1}{x}$ is ≥ 1 away from L. To show this precisely, for any $\delta > 0$ the number

$$x = \min\left\{\frac{\delta}{2}, \frac{1}{1+L}\right\}$$

satisfies $0 < |x| < \delta$ and $\left|\frac{1}{x} - L\right| \ge (1+L) - L = 1$. Thus no L is the limit of $f(x) = \frac{1}{x}$ as $x \to 0$, so this limit does not exist.

Example 2. For any $a > 0$, $\lim_{x\to a} \frac{1}{x} = \frac{1}{a}$. For simplicity in our proof, we will actually only show this for $a > 1$. (We'll see why this simplifies things later.) So, let $\epsilon > 0$. We want $\delta > 0$ so that for any x satisfying $0 < |x-a| < \delta$ we will have

$$\left|\frac{1}{x} - \frac{1}{a}\right| < \epsilon.$$

Note that

$$\left|\frac{1}{x} - \frac{1}{a}\right| = \frac{|x-a|}{|x||a|}.$$

The numerator we bound by δ and we want to bound (from below) the denominator by something which doesn't involve δ nor $|x|$. Now, suppose that $|x - a| < 1$. Then

$$-1 < x - a, \text{ so } a - 1 < x.$$

Since we are only considering $a > 1$, $a - 1 > 0$ so we can bound x from below by the positive number $a - 1$. Then

$$\frac{|x - a|}{|x||a|} < \frac{\delta}{(a-1)a}.$$

(Note that here is the reason we wanted $a > 1$: if $0 < a < 1$, then $a - 1$ is no longer positive and using this to bound $|x|$ might mess with the inequalities we're using since inequalities flip when multiplying by negative numbers. So, in order to avoid this we're assuming that $a > 1$. A modification of this proof will work even when $0 < a < 1$, but it would just take a little more effort to figure it out. Try to do it on your own!)

Now we see that we can take $\delta \leq (a-1)a\epsilon$, and to guarantee that the bound $|x - a| < 1$ we used in the course of the above argument also works we use

$$\delta = \min\{1, (a-1)a\epsilon\}.$$

Again, note that our simplification $a > 1$ guarantees that $\delta > 0$.

Theorem. The sequence definition and ϵ-δ definition of limits are equivalent.

Proof. Suppose that $\lim_{x\to a} f(x) = L$ according to the ϵ-δ definition. Let (x_n) be a sequence converging to a with $x_n \neq a$ and let $\epsilon > 0$. We want to show that $f(x_n)$ converges to L. By the ϵ-δ definition of limit, there exists $\delta > 0$ such that

$$0 < |x - a| < \delta \text{ implies } |f(x) - L| < \epsilon.$$

Since $x_n \to a$ there exists $N \in \mathbb{N}$ such that

$$0 < |x_n - a| < \delta \text{ for } n \geq N.$$

Thus for $n \geq N$, this implies that $|f(x_n) - L| < \epsilon$, so $f(x_n) \to L$ as claimed and the ϵ-δ definition hence implies the sequence definition.

To show that the sequence definition implies the ϵ-δ definition, we prove the contrapositive. To this end, suppose that $\lim_{x\to a} f(x) \neq L$ according to the ϵ-δ definition. Then there exists $\epsilon > 0$ such that for any $\delta > 0$ we can find x satisfying

$$0 < |x - a| < \delta \text{ but } |f(x) - L| \geq \epsilon.$$

In particular, applying this condition to δ of the form $\delta = \frac{1}{n}$ gives for each $n \in \mathbb{N}$ a number x_n such that

$$0 < |x_n - a| < \frac{1}{n} \text{ and } |f(x_n) - L) \geq \epsilon.$$

Then (x_n) is a sequence converging to a with $x_n \neq a$ such that $f(x_n)$ does not converge to L since each $f(x_n)$ is bounded away from L by at least ϵ. Thus the sequence definition is not satisfied either. $\square$

Remark. It would be extremely difficult to show that the sequence definition implies the ϵ-δ definition directly for the following reason: the sequence definition only tells us something about sequences $x_n \to a$ while the ϵ-δ definition requires something about any x satisfying $0 < |x - a| < \delta$. In other words, it is hard to make the jump from knowing something about terms in sequences approaching a to knowing something about *all* numbers within δ away from a.

Lecture 12: Continuous Functions

Today we started talking about continuous functions, focusing on examples. We will come to crucial properties of continuous functions next time.

Warm-Up. We show that $\lim_{x\to\infty} \frac{5x^2+3x-2}{3x^2-2x+1} = \frac{5}{3}$. Now, we haven't spoken about limits as $x \to \infty$ yet, but the definition is simple enough: $\lim_{x\to\infty} f(x) = L$ if for any $\epsilon > 0$ there exists $M \in \mathbb{R}$ such that

$$|f(x) - L| < \epsilon \text{ for } x > M.$$

In fact, this is somewhat similar in spirit to the definition of sequence convergence, and indeed the argument we give is similar to sequence arguments we've seen. Check the book for more about limits as $x \to \pm\infty$, and for "one-sided limits" where x approaches a from the right or from the left.

Let $\epsilon > 0$. We want M where for $x > M$ we have

$$\left| \frac{5x^2+3x-2}{3x^2-2x+1} - \frac{5}{3} \right| < \epsilon.$$

We can rewrite this absolute value as

$$\left| \frac{5x^2+3x-2}{3x^2-2x+1} - \frac{5}{3} \right| = \left| \frac{19x-11}{3(3x^2-2x+1)} \right|.$$

Now, the numerator we can bound by $19x$ as long as x is such that $19x > 11$. Since we just want our inequalities to hold for large enough positive x, this won't be an issue. So we have

$$\left| \frac{19x-11}{3(3x^2-2x+1)} \right| \leq \frac{19x}{|9x^2-6x+3|}.$$

The denominator here is larger than $3x^2$ once x is large enough (say for $x > 10000$), so we have

$$\frac{19x}{|9x^2-6x+3|} \leq \frac{19x}{3x^2} = \frac{19}{3x}.$$

Note that we can't replace the denominator simply with $9x^2$ since $9x^2$ is not necessarily smaller than $9x^2 - 6x + 3$. Our values of x will satisfy $x > M$, so we finally get

$$\left| \frac{5x^2+3x-2}{3x^2-2x+1} - \frac{5}{3} \right| \leq \frac{19}{3x} \leq \frac{19}{3M}.$$

Picking M large enough to make this smaller than ϵ and to make $x > M$ satisfy $x > 10000$ gives us what we want.

Definitions of continuous. Let $f : E \to \mathbb{R}$ be a function defined on some subset E of $\mathbb{R}$ and let $a \in E$. We say that f is *continuous* at a if any of the following three equivalent conditions hold:

- (ϵ-δ characterization) For any $\epsilon > 0$ there exists $\delta > 0$ such that

$$\text{if } |x - a| < \delta, \text{ then } |f(x) - f(a)| < \epsilon.$$

- (sequence characterization) For any sequence (x_n) converging to a, the sequence $f(x_n)$ converges to $f(a)$.

- (limit characterization) The limit of $f(x)$ as x approaches a is $f(a)$.

We say that f is *continuous* on E if it is continuous at all $a \in E$.

Remark. As with limits, there are various ways of thinking about continuity. The sequence characterization is perhaps most intuitive: for a sequence x_n getting closer and closer to a, the values $f(x_n)$ get closer and closer to $f(a)$. However, most examples we look at will involve the ϵ-δ characterization since this is the one which illustrates important techniques. The ϵ-δ characterization says: given any interval around $f(a)$, there is an interval around a so that points in this latter interval are sent into the former.

Well-known continuous functions. From now on, you can assume that well-known functions such as $f(x) = x^n$ or other polynomial functions, e^x, $\sin x$ and $\cos x$, and so on are continuous unless otherwise stated. In particular, some limit examples we previously looked at can now be interpreted in terms of continuity: our proof that $\lim_{x\to a} x^3 = a^3$ shows that $f(x) = x^3$ is continuous at any $x \in \mathbb{R}$, and our proof that $\lim_{x\to a} \frac{1}{x} = \frac{1}{a}$ for $a > 0$ shows that $g(x) = \frac{1}{x}$ is continuous on $(0, \infty)$.

Theorem. If $f, g : E \to \mathbb{R}$ are both continuous at $a \in E$, then $f + g : E \to \mathbb{R}$ is continuous at a.

Proof. ($\frac{\epsilon}{2}$-trick) Let $\epsilon > 0$. Since f is continuous at a there exists $\delta_1 > 0$ such that

$$|x - a| < \delta_1 \text{ implies } |f(x) - f(a)| < \frac{\epsilon}{2},$$

and since g is continuous at a there exists $\delta_2 > 0$ such that

$$|x - a| < \delta_2 \text{ implies } |g(x) - g(a)| < \frac{\epsilon}{2}.$$

Set $\delta = \min\{\delta_1, \delta_2\} > 0$. Then if $|x - a| < \delta$, we have $|x - a| < \delta_1$ and $|x - a| < \delta_2$ so x satisfies $|f(x) - f(a)| < \frac{\epsilon}{2}$ and $|g(x) - g(a)| < \frac{\epsilon}{2}$. Thus

$$\begin{aligned} |(f+g)(x) - (f+g)(a)| &= |(f(x) + g(x)) - (f(a) + g(a))| \\ &= |(f(x) - f(a)) + (g(x) - g(a))| \\ &\leq |f(x) - f(a) + |g(x) - g(a)| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \end{aligned}$$

so $f + g$ is continuous at a. ☐

Remark. The point above is that the choices of δ which come from what it means for f and g to be continuous at a might be different, so we need to take their minimum to guarantee that all our inequalities work. This "minimum choice of δ's" is common in ϵ-δ arguments.

Theorem. If $f, g : \mathbb{R} \to \mathbb{R}$ are both continuous, then the composition $g \circ f : \mathbb{R} \to \mathbb{R}$ is continuous.
The key to the proof is the following. We want

$$|g(f(x)) - g(f(a))| < \epsilon.$$

Reading this as $|g(\text{something}) - g(\text{something else})| < \epsilon$, we see that such an inequality comes from g being continuous; so if $|(\text{something}) - (\text{something else})| < \delta$, the above inequality will hold. Now, the "somethings" we are considering are of the form $f(x)$ and $f(a)$, so to guarantee that these "somethings" satisfy

$$|(\text{something}) - (\text{something else})| < \delta$$

we use the assumption that f is continuous. Here is the proof.

Proof. Let $a \in \mathbb{R}$ and let $\epsilon > 0$. Since g is continuous at $f(a)$, there exists $\delta_1 > 0$ such that

$$|y - f(a)| < \delta_1 \text{ implies } |g(y) - g(f(a))| < \epsilon.$$

Since f is continuous at a there exists $\delta > 0$ such that

$$|x - a| < \delta \text{ implies } |f(x) - f(a)| < \delta_1.$$

Then for $|x - a| < \delta$ we have for $y = f(x)$ that $|y - f(a)| < \delta_1$ so

$$|(g \circ f)(x) - (g \circ f)(a)| = |g(f(x)) - g(f(a))| = |g(y) - g(f(a))| < \epsilon.$$

Hence $g \circ f$ is continuos at a. Since a was arbitrary, $g \circ f$ is continuous on all of $\mathbb{R}$. $\square$

Remark. Check the book for other basic properties of continuous functions, such as that the product of continuous functions is continuous and that the quotient of continuous functions is continuous as long as the denominator is defined.

Example. The function $f : [0, 1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$$

is nowhere continuous. Indeed, let $r \in \mathbb{Q}$. We have seen previously (see Week 3 Lecture Notes) that there exists a sequence of irrationals y_n converging to r. Then $f(y_n) = 0$ does not converge to $f(r) = 1$, so f is not continuous at $r \in \mathbb{Q}$. If $y \notin \mathbb{Q}$, taking a sequence of rationals r_n converging to y gives $f(r_n) = 1$ not converging to $f(y) = 0$, so f is not continuous at $y \notin \mathbb{Q}$ either.

My favorite function. This is my all-time favorite function and example, both because of the following properties this function has and because later it will give an interesting example dealing with "integrability". Enjoy.

Define $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} 0 & \text{if } x \notin \mathbb{Q} \\ 1 & \text{if } x = 0 \\ \frac{1}{q} & \text{if } x = \frac{p}{q} \text{ where } p, q \in \mathbb{N} \text{ have no common factors} \end{cases}.$$

For instance,

$$f\left(\frac{2}{3}\right) = \frac{1}{3}, \ f(1) = 1, \ f\left(\frac{1}{\pi}\right) = 0, \text{ and } f\left(\frac{2}{4}\right) = \frac{1}{2}.$$

The claim is that this function is discontinuous at each rational $r \in [0, 1]$, and yet (amazingly) continuous at each irrational $y \in [0, 1]$! Understanding why this is really forces you to understand what continuous really means and how to work with the definition correctly. The book has a proof of this fact using sequences, but on the homework I'll have you show this using ϵ's and δ's, which I think is more enlightening.

Remark. My favorite example shows that issues of continuity can be subtle, and strange looking things can happen. The graph of this function looks really far away from being "continuous" anywhere since it consists of a bunch of "dots" jumping around, and yet it turns out to be continuous at many, many points. Hooray!

Lecture 13: More on Continuous Functions

Today we continued talking about continuous functions, looking at examples and more properties. We proved the Extreme Value Theorem, which says that any continuous function on a closed interval is bounded and has a maximum and a minimum. This is an absolutely crucial fact, and is the source of many applications in other fields. For us, it will eventually lead us directly to the Fundamental Theorem of Calculus.

Warm-Up 1. We show that the function $f : (-1, \infty) \to \mathbb{R}$ defined by $f(x) = \sqrt{x+1}$ is continuous. Take $a \in (-1, \infty)$ and let $\epsilon > 0$. We want $\delta > 0$ such that

$$|x - a| < \delta \text{ implies } |\sqrt{x+1} - \sqrt{a+1}| < \epsilon.$$

There are two ways in which we can bound the absolute value on the right:

$$|\sqrt{x+1} - \sqrt{a+1}| = \frac{|x-a|}{\sqrt{x+1} + \sqrt{a+1}} \leq \frac{|x-a|}{\sqrt{a+1}}$$

and

$$|\sqrt{x+1} - \sqrt{a+1}| \leq \sqrt{|x-a|}.$$

In the first we are using:

$$|\sqrt{p} - \sqrt{q}| = \frac{|p-q|}{\sqrt{p} + \sqrt{q}}$$

and in the second

$$|\sqrt{p} - \sqrt{q}| \leq \sqrt{|p-q|}.$$

Thus in the first case $\delta = \epsilon\sqrt{a+1}$ works, and in the second $\delta = \epsilon^2$ works. Since $a \in (-1, \infty)$ was arbitrary, f is continuous on all of $(-1, \infty)$.

Note that in the first bound we used δ ended up depending on the point a we were checking continuity at, while in the second bound it did not. The fact that it is possible to choose δ in a way which does not depend on the point we're checking continuity at says that this function is actually *uniformly continuous*, a concept we will come back to next time.

Warm-Up 2. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuous and that for some $a \in \mathbb{R}$ we have $f(a) > 0$. We claim that there is an entire interval $(a - \delta, a + \delta)$ around a on which f only has positive values. Indeed, since $f(a) > 0$ the definition of continuous gives a $\delta > 0$ such that

$$|x - a| < \delta \text{ implies } |f(x) - f(a)| < f(a).$$

The first inequality says that $x \in (a - \delta, a + \delta)$ and the second says that

$$f(x) \in (f(a) - f(a), f(a) + f(a)) = (0, 2f(a)),$$

so $f(x) > 0$ for any $x \in (a - \delta, a + \delta)$ as desired.

Remark. The previous Warm-Up says that a continuous function which is nonzero at a point is bounded away from zero near that point, and the fact that f is continuous is important. For instance, the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$$

is positive at 1 but any interval around 1 contains an irrational, at which f has the value 0. So there is no interval around 1 (or around any rational) on which f has only positive values.

Warm-Up 3. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a continuous function such that $f(r) = 0$ for any $r \in \mathbb{Q}$. We claim that in fact it must be the case that $f(x) = 0$ for all $x \in \mathbb{R}$. Indeed, let $y \in \mathbb{R}$. By the denseness of $\mathbb{Q}$ in $\mathbb{R}$ there exists a sequence (r_n) of rationals converging to y:

$$r_n \to y.$$

(We used this fact last time, and as a reminder is proved in the Week 3 Lecture Notes.) Since f is continuous, the sequence $f(r_n)$ converges to $f(y)$:

$$f(r_n) \to f(y).$$

(This is the sequence characterization of continuous.) But $f(r_n) = 0$ for all n since each r_n is rational, and since the constant zero sequence converges to 0, we must have $f(y) = 0$. Thus $f(x) = 0$ for all $x \in \mathbb{R}$.

The point here is that a continuous function is fully determined by its values at rational numbers: if f is continuous and all we know is $f(r)$ for each rational r, we can completely determine what $f(x)$ is for any x. Note how peculiar this seems: $\mathbb{Q}$ is countable, and yet we can determine an uncountable amount of information (i.e. the values of $f(x)$ for any $x \in \mathbb{R}$) using only the countably many data points $f(r)$ for each $r \in \mathbb{Q}$.

Cardinality of the set of continuous functions (Optional). We didn't do this in class, and it's purely optional material, but let's see how we can use the above fact to determine the cardinality of the set $C(\mathbb{R})$ of continuous functions from $\mathbb{R}$ to $\mathbb{R}$. ($C(\mathbb{R})$ is standard notation for the space of continuous functions from $\mathbb{R}$ to $\mathbb{R}$.)

You might have seen in Math 300 that the set of *all* functions from $\mathbb{R}$ to $\mathbb{R}$ has cardinality equal to $2^{|\mathbb{R}|}$, which is the cardinality of the power set of $\mathbb{R}$. (If you didn't see this, you can find a proof in some old notes of mine at `http://math.northwestern.edu/~scanez/courses/math300/spring13/handouts/examples.pdf`.) We claim that $C(\mathbb{R})$ only has cardinality equal to that of $\mathbb{R}$, so that there are way more functions from $\mathbb{R}$ to $\mathbb{R}$ than there are continuous ones.

Since $\mathbb{Q}$ is countable, we can list its elements in an infinite list as

$$r_1,\ r_2,\ r_3,\ r_4,\ \ldots.$$

The remark at the end up Warm-Up 3 says that a continuous function $f : \mathbb{R} \to \mathbb{R}$ is completely determined by the numbers

$$f(r_1),\ f(r_2),\ f(r_3),\ f(r_4), \ldots.$$

Such a list of numbers gives an element of $\mathbb{R}^\infty$, which is Cartesian product of countably infinite many copies of $\mathbb{R}$, and is also the space of sequences of real numbers. Thus, $C(\mathbb{R})$ is in one-to-one correspondence with $\mathbb{R}^\infty$. Since $\mathbb{R}^\infty$ has the same cardinality as $\mathbb{R}$, so does $C(\mathbb{R})$. Again, the point is that continuity gives us a way to determine an uncountable amount of information using only a countable amount of data. (See `http://math.northwestern.edu/~scanez/courses/math300/spring13/handouts/final-practice-solns.pdf` for a proof that $\mathbb{R}^\infty$ has the same cardinality as $\mathbb{R}$.) We now come back from this tangent and return to more standard material.

Extreme Value Theorem. If $f : [a, b] \to \mathbb{R}$ is continuous, then f is bounded and attains a maximum and a minimum. The proof of this is in the book. In class we gave the same proof that f is bounded as the book, but a different proof that f has a maximum and a minimum, so let me reproduce only the new proof here.

Proof that f has a maximum and a minimum. Since f is bounded, the set $\{f(x) \mid x \in [a,b]\}$ of values of f has a supremum and an infimum. Call the supremum M; we prove that there exists $y \in [a,b]$ such that $f(y) = M$, which is what it means to say that f has a maximum. The proof that f attains its minimum value is similar.

By a previous property of supremums, there exists a sequence

$$f(x_1), f(x_2), f(x_3), \ldots \text{ in } \{f(x) \mid x \in [a,b]\}$$

which converges to M. Consider the corresponding sequence (x_n) in $[a,b]$. By the Bolzano-Weierstrass Theorem, (x_n) has a convergent subsequence (x_{n_k}) converging to some $y \in [a,b]$. Since $x_{n_k} \to y$, continuity of f gives

$$f(x_{n_k}) \to f(y).$$

But $f(x_{n_k})$ also converges to M, so we must have $f(y) = M$ and f attains a maximum value as claimed. □

Remark. Note that both the assumption that f is continuous and that it is defined on a *closed* interval are important in the Extreme Value Theorem. For instance, the function $f : (0,\infty) \to \mathbb{R}$ defined by $f(x) = \frac{1}{x}$ is unbounded even though it is continuous, and the function $f : [0,2] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \frac{1}{1-x} & \text{if } x \neq 1 \\ 100 & \text{if } x = 1 \end{cases}$$

is unbounded even though the interval $[0,2]$ is closed.

Generalization. The property of $[a,b]$ which makes the above work is that $[a,b]$ is *compact.* (A notion I defined in previous Lecture Notes.) In more general settings, we can talk about compact spaces in higher dimensions, and it is still true that any continuous function $f : X \to \mathbb{R}$ from a compact space X to $\mathbb{R}$ has a maximum and a minimum. It is this general version which has numerous applications in other fields. Essentially, anytime you have a result saying that a certain expression has a maximum value or a minimum value, there's a version of the Extreme Value Theorem and a compact space hiding in the background.

Importance. Note that the Extreme Value Theorem depended in a crucial way on the Bolzano-Weierstrass Theorem, which I previously said was going to be an extremely important result. Indeed, here is the chain of logic which will lead us from Bolzano-Weierstrass to the Fundamental Theorem of Calculus:

$$\text{Bolzano-Weierstrass Theorem} \implies \text{Extreme Value Theorem}$$

$$\implies \text{Mean Value Theorem} \implies \text{Fundamental Theorem of Calculus.}$$

We will talk about the Mean Value Theorem next week. Essentially, in the end much of calculus traces back to the Bolzano-Weierstrass Theorem and the fact that closed intervals are compact. Good stuff.

Intermediate Value Theorem. If f is a continuous function and a, b, y are such that

$$f(a) < y < f(b),$$

then there exists x such that $f(x) = y$. In other words, the "intermediate" value of y is actually attained. The proof of this is in the book, and I will leave it up to you to look it up. Instead, let's look at an application, which is where the real power of this theorem lies.

Example. We claim that there is a real number y satisfying the equation

$$e^y = y^3.$$

The continuous function we will apply the Intermediate Value Theorem to is

$$f(x) = e^x - x^3.$$

The number y we are looking for should satisfy $f(y) = 0$, and we can show that there is such a y by showing that 0 is an "intermediate value" of our function. We have:

$$f(0) = e^0 - 0^3 = 1 > 0 > f(3) = e^3 - e^3,$$

so the Intermediate Value Theorem implies there exists $0 < y < 3$ such that $f(y) = 0$ as desired.

Lecture 14: Uniformly Continuous Functions

Today we spoke about the notion of "uniform continuity", which is a strong form of continuity in the sense that uniformly continuous functions are always continuous but not vice-versa, and uniformly continuous functions have many nice properties that continuous functions in general don't necessarily have. Essentially, uniformly continuous functions are continuous functions which don't change "too rapidly".

Remark. The Warm-Ups today are a little out there, in that they give non-standard applications of higher-dimensional analogues of the Intermediate Value Theorem and the Extreme Value Theorem. This is not something we will come back to this quarter, but is something you may see if you continue on next quarter. Also, it is these higher-dimensional versions which have nice applications in other fields. The point is that although we will certainly use these two theorems in the weeks to come, they really shine once you move to higher dimensions. So, in the name of presenting more interesting-then-usual examples, here you go ;)

Warm-Up 1. We claim that at any instant in time, there are a pair of antipodal points on the surface of the Earth with the same exact temperature. (Antipodal points are points which are directly opposite each other, such as the north and south poles.) The reason is an application of the Intermediate Value Theorem.

Let X denote the set of all pairs of antipodal points on a sphere which we use to model the surface of the Earth:

$$X := \{(p, q) \mid p \text{ and } q \text{ are antipodal points on a sphere}\}.$$

This set X sits inside of $\mathbb{R}^6$ since each antipodal point is itself a point in $\mathbb{R}^3$. Now, consider the function $f : X \to \mathbb{R}$ defined by

$$f(p, q) = (\text{temperature at } p) - (\text{temperature at } q).$$

I claim this function is continuous, but justifying this requires that we understand what continuity means in higher dimensions. For us, this is just a reflection of the fact that a small change in either

p or q causes a small change in f: i.e. moving away from a point by a small amount does not create a large change in temperature. Being a continuous function, the higher-dimensional version of the Intermediate Value Theorem applies. (The key property of X which allows this generalization to apply is that it is *connected*, a concept you would learn about next quarter.)

Now, take some antipodal points a and b. If the temperature at a and b is the same we are done; otherwise, the temperature at one is larger than that at the other. Thus, one of $f(a,b)$ or $f(b,a)$ is positive and the other negative; say $f(a,b)$ is the negative one. Then

$$f(a,b) < 0 < f(b,a),$$

so the Intermediate Value Theorem says there exists $(p,q) \in X$ such that $f(p,q) = 0$. These p and q are then antipodal points on the surface of the sphere (i.e. Earth) at which the temperature is the same.

Warm-Up 2. Take any two bounded subsets A and B of $\mathbb{R}^2$ which contain their boundary. (This guarantees that A and B are *compact* subsets of $\mathbb{R}^2$. Recall that in our class, to say that a closed interval $[a,b]$ is compact is to say that any sequence in it has a convergent subsequence; this same condition defines what "compact" means in higher dimensions, once you have said what it means for a sequence in higher dimensions to converge. This is all straightforward to do, and will be covered in Math 320-2.) We claim that there are points $a \in A$ and $b \in B$ such that the distance between a and b *is* the minimum possible distance between a point of A and a point of B. Note that the set of all distances from a point of A to a point of B is bounded below by 0, so it has an infimum: we are saying that this infimum is actually realized as the distance between two specific points.

This is an application of the Extreme Value Theorem in higher dimensions. Define a function $f : A \times B \to \mathbb{R}$ by setting

$$f(a,b) = \text{the distance from } a \text{ and } b.$$

This function is continuous since the formula for this distance is something of the form

$$\sqrt{(a_1 - b_1)^2 + (a_2, b_2)^2},$$

which is made up of continuous expressions. Also, $A \times B$ is compact a subset of $\mathbb{R}^4$ since A and B are each compact (again a fact you'll see next quarter), so the Extreme Value Theorem says that f has a minimum; that is, there are points $a \in A$ and $b \in B$ such that $f(a,b)$ is the minimum distance between points of A and points of B, as desired.

Remark. Again, these two Warm-Ups are only meant to illustrate the use of the Intermediate Value and Extreme Value Theorems in higher dimensions. We will not be looking at such things this quarter, and only the simpler versions of these theorems we gave last lecture will show up in our course.

Examples to motivate uniform continuity. Consider the functions $f(x) = x^2$ and $g(x) = 3x$, both defined on all of $\mathbb{R}$. Both of these are continuous at any $a \in \mathbb{R}$, and so for a fixed $\epsilon > 0$ there exists $\delta > 0$ satisfying the requirements in the ϵ-δ definition of continuous at a. In particular, if you work this out in each case using techniques we've seen and by looking at similar examples worked out in previous Lecture Notes, you'll find that for f the value

$$\delta = \frac{\epsilon}{2 + |a|}$$

works while for g the value

$$\delta = \frac{\epsilon}{3}$$

works.

Here is the key observation: for f the value of δ we find depends on a, the point we're checking continuity at, while for g it does not. As a gets larger and larger, the δ for f gets smaller and smaller, while the δ for g remains the same. In fact, because $\delta \to 0$ as $a \to \infty$ in the case of $f(x) = x^2$, it is not possible to find a "smallest" possible δ which will work for *all* $a \in \mathbb{R}$ at once since we want $\delta > 0$. This distinction is what tells us that f is not uniformly continuous on $\mathbb{R}$ but that g is uniformly continuous on $\mathbb{R}$.

Definition of uniformly continuous. A function $f : E \to \mathbb{R}$ defined on some subset E of $\mathbb{R}$ is *uniformly continuous* if for any $\epsilon > 0$ there exists $\delta > 0$ such that

$$\text{if } |x - y| < \delta, \text{ then } |f(x) - f(y)| < \epsilon.$$

This looks very similar to the usual definition of continuous, except that there we fixed a point $y \in E$ we were checking continuity at while here y is not fixed; there, x was allowed to vary, while here both x and y are allowed to vary.

Practically, this means that in usual continuity δ can depend on ϵ and the point you're checking continuity at, while in uniform continuity δ can only depend on ϵ. Since the same δ works for all $y \in E$, f is continuous in a "uniform" way across all of E. Geometrically, a continuous function fails to be uniformly continuous when it changes "too rapidly", such as when its graph gets steeper and steeper. This is what happens in the $f(x) = x^2$ on $\mathbb{R}$ case, but does not happen for $g(x) = 3x$ on $\mathbb{R}$. We will talk more about the relation between "uniformly continuous" and "steepness" when we talk about derivatives.

The domain matters. Consider the function $f(x) = x^2$ on the interval $[0, 1]$. In this case, for $|x - y| < \delta$ we can bound $|f(x) - f(y)|$ as follows:

$$|x^2 - y^2| = |x - y||x + y| < \delta|x + y| \le 2\delta$$

since $|x + y| \le 2$ for $x, y \in [0, 1]$. Thus for $\epsilon > 0$, $\delta = \frac{\epsilon}{2}$ satisfies the ϵ-δ definition of continuous, so f is uniformly continuous on $[0, 1]$.

The point is that when asking whether a function is uniformly continuous or not, the domain of the function matters: $f(x) = x^2$ is not uniformly continuous on all of $\mathbb{R}$, but it is uniformly continuous on $[0, 1]$. Geometrically, when restricting the domain to be $[0, 1]$ the graph of $f(x) = x^2$ does not get arbitrarily steep.

Example. The function $f(x) = \frac{1}{x}$ is not uniformly continuous on $(0, 1)$. When going through a proof that f is continuous at $a \in (0, 1)$, for $\epsilon > 0$ you find that

$$\delta = \min\left\{\frac{a}{2}, \frac{a^2\epsilon}{2}\right\}$$

satisfies the required definition, since

$$\left|\frac{1}{x} - \frac{1}{a}\right| = \frac{|x - a|}{xa} \le \frac{|x - a|}{a^2/2}$$

for $|x - a| < \frac{a}{2}$. However, note that $\delta \to 0$ as $a \to 0$, so there will not be a single positive δ which satisfies the required definition for all $a \in (0, 1)$ at once. This suggests that f is not uniformly continuous; the book has a more precise proof of this. Again, geometrically, note that the graph of f gets steeper and steeper as $a \to 0$.

Properties of uniformly continuous functions. Here are two basic properties of uniformly continuous functions, both of which are proved in the book:

- If $f : E \to \mathbb{R}$ is uniformly continuous and (x_n) is a Cauchy sequence in E, then $(f(x_n))$ is Cauchy as well. Thus, uniformly continuous functions send Cauchy sequences to Cauchy sequences.
- If $f : (a, b) \to \mathbb{R}$ is uniformly continuous, then f can be "extended" to a continuous function $f : [a, b] \to \mathbb{R}$. Thus, uniformly continuous functions on open intervals can be defined at the endpoints so as to still remain continuous.

Note that for the Cauchy sequence $\frac{1}{n+1}$ in $(0, 1)$, the function from the previous example has $f(\frac{1}{n+1}) = n + 1$, which is not Cauchy. Thus this function does not satisfy the first property above, which gives a proof that $f(x) = \frac{1}{x}$ is not uniformly continuous on $(0, 1)$.

Theorem. Any continuous function $f : [a, b] \to \mathbb{R}$ is uniformly continuous. So, continuous functions on closed intervals are automatically uniformly continuous. You should view this as being one of the many properties (together with the Bolzano-Weierstrass Theorem and the Extreme Value Theorem) which make closed intervals special.

The book has a proof of this, and here I will present a sketch of another "proof". I write this in quotation marks since what follows is not a precise proof because certain parts will be a bit hand-wavy, but I feel that this sketch better captures the role that the fact we have a closed interval plays. (In fact, we will use the Extreme Value Theorem.) There is a way to make what I'm going to outline precise and rigorous, but doing so fully probably isn't worth all the work involved. So, use this only to get some intuition for why this Theorem is true: the point is that we want to find a single $\delta > 0$ which works for all points in our domain at once.

"Proof". Let $\epsilon > 0$. Then for any $y \in [a, b]$, f is continuous at y so there exists $\delta_y > 0$ such that

$$|x - y| < \delta_y \text{ implies } |f(x) - f(y)| < \epsilon.$$

(We're using δ_y to emphasize the δ depends on y, and different y's might require different δ's.) Now, view the assignment $y \mapsto \delta_y$ as defining a function $g : [a, b] \to \mathbb{R}$:

$$g(y) = \delta_y.$$

We claim (and this is the hand-wavy part) that g is continuous: intuitively, changing y by a small amount should only change δ_y by a small amount, and indeed in the examples we've seen where δ depends on a this has been the case. (There is another issue, in that δ_y isn't uniquely defined yet since there could be different δ's which satisfy the definition of continuity for the same y. This is easier to deal with: we can define δ_y to be the supremum among all δ's which work.) So, taking it for granted that there is a way to make $g(y) = \delta_y$ continuous, we push onward.

Since $g : [a, b] \to \mathbb{R}$ is continuous, it has a minimum value; call it δ. Note that $\delta > 0$ since it is the minimum of positive numbers. We claim that this one δ satisfies the definition of continuity at any $y \in [a, b]$. Indeed, suppose that $|x - y| < \delta$. Since $\delta \leq \delta_y$, we then also have $|x - y| < \delta_y$, so by the choice of δ_y we get

$$|f(x) - f(y)| < \epsilon.$$

Thus $|x - y| < \delta \implies |f(x) - f(y)| < \epsilon$, so f is uniformly continuous as claimed. □

Lecture 15: Differentiable Functions

Today we started talking about differentiable functions. As opposed to a calculus course, our focus will be on understanding what differentiability and its consequences mean rather than on techniques for computing derivatives.

Warm-Up 1. We show that the function $f(x) = \sqrt{x}$ is uniformly continuous on $[0, \infty)$, giving two proofs using the two ways we've seen so far of dealing with an expression like $\sqrt{x} - \sqrt{y}$. The first is straightforward, while the second uses the fact that continuous functions on closed intervals are automatically uniformly continuous.

Proof 1. Let $\epsilon > 0$ and set $\delta = \epsilon^2$. Then $\delta > 0$ and for any $x, y \in [0, \infty)$ such $|x - y| < \delta$ we have

$$|f(x) - f(y)| = |\sqrt{x} - \sqrt{y}| \leq \sqrt{|x - y|} < \sqrt{\delta} = \epsilon.$$

Thus f is uniformly continuous on $[0, \infty)$. $\square$

Proof 2. First we show that f is uniformly continuous on $[1, \infty)$. Let $\epsilon > 0$ and set $\delta = \epsilon$. Then if $|x - y| < \delta$ we have

$$|f(x) - f(y)| = |\sqrt{x} - \sqrt{y}| = \frac{|x - y|}{|\sqrt{x} + \sqrt{y}|} \leq \frac{|x - y|}{\sqrt{y}} \leq |x - y| < \delta = \epsilon,$$

where we use the fact that $\sqrt{y} \geq 1$ since $y \geq 1$ in our domain $[1, \infty)$. Thus f is uniformly continuous on $[1, \infty)$.

Now, since f is continuous on $[0, 2]$, it is uniformly continuous on $[0, 2]$ also. Hence f is uniformly continuous on $[0, 2]$ and on $[1, \infty)$, and we claim that this implies f is uniformly continuous on their union $[0, \infty)$. Indeed, let $\epsilon > 0$. Since f is uniformly continuous on $[0, 2]$ there exists $\delta_1 > 0$ such that

$$\text{if } x, y \in [0, 2] \text{ and } |x - y| < \delta_1, \text{ then } |f(x) - f(y)| < \epsilon,$$

and since f is uniformly continuous on $[1, \infty)$ there exists $\delta_2 > 0$ such that

$$\text{if } x, y \in [1, \infty) \text{ and } |x - y| < \delta_2, \text{ then } |f(x) - f(y)| < \epsilon.$$

Thus for $\delta = \min\{\delta_1, \delta_2, 1\}$ we have

$$|f(x) - f(y)| < \epsilon \text{ whenever } |x - y| < \delta$$

since either $x, y \in [0, 2]$ in which case $|x - y| < \delta \leq \delta_1$ or $x, y \in [1, \infty)$ in which case $|x - y| < \delta \leq \delta_2$. (Making δ also smaller than 1 guarantees that we rule out the possibility that, say, $x < 1$ but $y > 2$ since in that case we don't have $x, y \in [0, 2]$ or $x, y \in [1, \infty)$.) $\square$

Remark. Note why the argument in the second proof above doesn't work if we don't restrict ourselves to $[1, \infty)$: without this restriction we cannot bound

$$\frac{|x - y|}{\sqrt{y}} \text{ by } |x - y|.$$

The end result is that in this case we would take $\delta = \epsilon\sqrt{y}$, which works fine when showing that f is continuous at each $y \in (0, \infty)$ but does not work for showing uniform continuity since here δ depends on y. Restricting y to be in $[1, \infty)$ gives a way to come up with a δ which only depends on

ϵ. Then, we deal with the $[0,1]$ part of the domain using the general fact that a continuous function on a closed interval is uniformly continuous. We actually use $[0,2]$ instead of just $[0,1]$ so that our two domains have more of an overlap, which we need for the last part of the proof. Hopefully you will agree that the first proof is much simpler.

Sequential Characterization of Uniform Continuity. A function $f:\mathbb{R}\to\mathbb{R}$ is uniformly continuous if and only if whenever x_n and y_n are two sequences with $|x_n-y_n|\to 0$, then $|f(x_n)-f(y_n)|\to 0$. In other words, uniformly continuous functions have the property that sequences which approach each other are sent to sequences which still approach each other, and functions with this property must be uniformly continuous.

Remark. Before giving the proof, note that continuous functions don't necessarily have this property. Indeed, $f:(0,1)\to\infty$ defined by $f(x)=\frac{1}{x}$ is continuous but for $x_n=\frac{1}{n+1}$ and $y_n=\frac{1}{2(n+1)}$ we have

$$|x_n-y_n|=\frac{1}{2(n+1)}\to 0 \text{ but } |f(x_n)-f(y_n)|=|(n+1)-2(n+1)|=n+1\not\to 0.$$

That is, x_n and y_n approach each other but $\frac{1}{x_n}$ and $\frac{1}{y_n}$ do not. The issue is that when looking at something like $|f(x_n)-f(y_n)|$, in order to say this converges to 0 we need to vary both terms in this absolute value, but continuity only allows us to vary one term at a time; i.e. in the condition

$$|x-a|<\delta \text{ implies } |f(x)-f(a)|<\epsilon$$

in the definition of continuous, a and $f(a)$ are fixed and only x and $f(x)$ vary.

Proof of sequential characterization. Suppose f is uniformly continuous and let x_n and y_n be sequences such that $|x_n-y_n|\to 0$. Let $\epsilon>0$. Since f is uniformly continuous there exists $\delta>0$ such that

$$|x-y|<\delta \text{ implies } |f(x)-f(y)|<\epsilon.$$

Since $|x_n-y_n|\to 0$ there exists $N\in\mathbb{N}$ such that

$$|x_n-y_n|<\delta \text{ for } n\geq N.$$

But then this implies that

$$|f(x_n)-f(y_n)|<\epsilon \text{ for } n\geq N,$$

so $|f(x_n)-f(y_n)|\to 0$ as claimed.

To establish the converse, we instead prove its contrapositive. To this end, suppose that f is not uniformly continuous. Then there exists $\epsilon>0$ such that for any $\delta>0$ there exist $x,y\in\mathbb{R}$ such that

$$|x-y|<\delta \text{ but } |f(x)-f(y)|\geq\epsilon.$$

In particular, for each $n\in\mathbb{N}$ there exist $x_n,y_n\in\mathbb{R}$ such that

$$|x_n-y_n|<\frac{1}{n} \text{ but } |f(x_n)-f(y_n)|\geq\epsilon.$$

The sequences x_n and y_n constructed in this way satisfy $|x_n-y_n|\to 0$ but $|f(x_n)-f(y_n)|\not\to 0$ since $f(x_n)$ and $f(y_n)$ are always a distance of at least $\epsilon>0$ apart from each other. This proves the contrapositive of the claim: if f satisfies the condition that $|x_n-y_n|\to 0$ implies $|f(x_n)-f(y_n)|\to 0$, then f is uniformly continuous. $\square$

Definition of differentiable. A function $f : (a, b) \to \mathbb{R}$ is said to be *differentiable* at $y \in (a, b)$ if

$$\lim_{x \to y} \frac{f(x) - f(y)}{x - y}$$

exists, in which case we call the value of this limit the *derivative of f at y* and denote it by $f'(y)$. This limit can equivalently be written as

$$\lim_{h \to 0} \frac{f(y+h) - f(y)}{h}$$

after making the substitution $x = y + h$ and noting that then saying $x \to y$ and $h \to 0$ mean the same thing. We say that f is differentiable on (a, b) if it is differentiable at each $y \in (a, b)$. The same definition works for domains other than just open intervals.

Remark. The above definition is no doubt one you've seen in a previous calculus course. The fraction we are taking the limit of is the slope of secant line passing through $(x, f(x))$ and $(y, f(y))$ on the graph of f in the first version and through $(y, f(y))$ and $(y+h, f(y+h))$ in the second; thus the limit, when it exists, indeed gives us the slope of the tangent line at y itself.

You can from now on safely assume that whatever functions you computed derivatives of in calculus are indeed differentiable: polynomial functions, trig functions, exponentials and logarithms, etc. You can also from now on use whatever differentiation rules you want: product rule, chain rule, etc. These are proved in Section 4.2 of our book; I encourage you to look at those proofs but we probably won't say much about them in class.

Example 1. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

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

We claim that this function is not differentiable at 0. Indeed, we have:

$$\lim_{x \to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x \to 0} \frac{x \sin \frac{1}{x}}{x} = \lim_{x \to 0} \sin \frac{1}{x},$$

and this limit does not exist due to the oscillatory behavior of $\sin \frac{1}{x}$ as $x \to 0$. (Note that when we are considering the limit as $x \to 0$, we are looking at values approaching 0 but never equal to zero itself, which is why we were able to substitute $f(x) = x \sin \frac{1}{x}$ for such x.

Now consider the function $g : \mathbb{R} \to \mathbb{R}$ defined by the same formula as f only using $x^2 \sin \frac{1}{x}$ instead of $x \sin \frac{1}{x}$. In this case we end up with:

$$\lim_{x \to 0} \frac{g(x) - g(0)}{x - 0} = \lim_{x \to 0} \frac{x^2 \sin \frac{1}{x}}{x} = \lim_{x \to 0} x \sin \frac{1}{x} = 0,$$

so g is differentiable at 0 and $g'(0) = 0$. The function g is differentiable for $x \neq 0$ since near such values g is the same as the function

$$x^2 \sin \frac{1}{x}$$

which is differentiable at $x \neq 0$ as a consequence of some differentiantion rules, in particular the product and chain rules. Hence g is differentiable on all of $\mathbb{R}$. The value of $g'(x)$ for $x \neq 0$ is obtained by differentiating $x^2 \sin \frac{1}{x}$ for $x \neq 0$, and thus we find that the derivative of $g : \mathbb{R} \to \mathbb{R}$ is

$$g'(x) = \begin{cases} 2x \sin \frac{1}{x} - \cos \frac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Note that this derivative g', however, is not continuous at 0 since $\lim_{x\to 0} g'(x)$ does not exist. (It should equal $g'(0)$ in order for g' to be continuous at 0.) This is due to the $\cos\frac{1}{x}$ term, which has no limit as $x \to 0$. We say that even though g is differentiable, it is not *continuously differentiable*.

The function where we use the same formula only replacing $x^2 \sin\frac{1}{x}$ now by $x^3 \sin\frac{1}{x}$ is in fact continuously differentiable, as you will verify on the homework.

Definitions. A function $f : \mathbb{R} \to \mathbb{R}$ is *continuously differentiable* if it is differentiable and its derivative f' is continuous. A function $f : \mathbb{R} \to \mathbb{R}$ is *twice differentiable* if it is differentiable and its derivative f' is itself differentiable. Similarly, we have notions of f being three times, four times, or k times differentiable. We say that f is *infinitely differentiable*, or *smooth*, if it is k times differentiable for all $k \in \mathbb{N}$. (So f is differentiable and all higher-order derivatives exist.)

Example 2. We claim that the function $g : \mathbb{R} \to \mathbb{R}$ defined by

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

is smooth. Near any $x > 0$, g agrees with the function $e^{-1/x}$ which is infinitely differentiable by repeated application of the chain rule. Near any $x < 0$, g agrees with the constant zero function, which again is infinitely differentiable. So, we really only need to look at what happens as 0. The key thing which makes this work is that $e^{-1/x} \to 0$ much faster than any polynomial in x as $x \to 0$. There is a way to make this precise, but we will take it for granted in what follows.

First, we have

$$\lim_{x\to 0} \frac{g(x) - g(0)}{x - 0} = \lim_{x\to 0} \frac{e^{-1/x}}{x} = 0$$

since the numerator goes to 0 much faster than the denominator. Thus g is differentiable at 0 and we compute:

$$g'(x) = \begin{cases} \frac{e^{-1/x}}{x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Next, we have

$$\lim_{x\to 0} \frac{g'(x) - g'(0)}{x - 0} = \lim_{x\to 0} \frac{e^{-1/x}}{x^3} = 0,$$

since again the numerator goes to 0 much faster than the denominator. Thus g is twice differentiable and

$$g''(x) = \begin{cases} \frac{e^{-1/x}}{x^4} - 2\frac{e^{-1/x}}{x^3} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

In general, say by using induction, the expression for the kth derivative of g at $x \neq 0$ will be something of the form

$$\left(\frac{\text{polynomial in } x}{\text{another polynomial in } x}\right) e^{-1/x},$$

and the limit of such things as $x \to 0$ will always be 0. We find that $g^{(k)}(0) = 0$ for all k, and that g is smooth as claimed.

Remark. Smooth functions are awesome. Indeed, they show up all over the place in applications to physics and other fields. I'm a bit biased since my research is essentially all about the use of smooth functions in geometry, but, really, they are great.

Proposition. If $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$, then f is continuous at a.

Remark. This should again be a well-known fact from calculus. The basic idea is that in order for the limit

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

to exist, the numerator had better approach 0 since the denominator does. (Otherwise with the denominator approaching 0 but not the numerator we would end up with a fraction whose limit did not exist.) Thus for this limit to exist we need $f(x) - f(a) \to 0$ as $x \to a$, so $\lim_{x \to a} f(x) = f(a)$ and thus f is continuous at a.

To make this rigorous we would have to fully justify that if the above limits exists the numerator does approach 0. This is more trouble than its worth since the book has a better proof of this Proposition, so I defer to that instead.

Proposition. If $f : \mathbb{R} \to \mathbb{R}$ is differentiable and has a local maximum or minimum at a, then $f'(a) = 0$. Note that to say f has a *local maximum* at a means that there exists an interval $(a - \delta, a + \delta)$ around a on which $f(a)$ is a maximum, and similarly for a local minimum.

This is a very well-known fact from calculus and forms the basis of various optimization techniques. This fact is more-or-less clear when considering the graph of f, since the tangent line at a local max or min is horizontal, but of course not every possible differentiable function will have an easy-to-draw graph, so that geometric intuition isn't enough to constitute a full justification.

Proof. Suppose that f has a local maximum at a; the proof for the local minimum case is very similar. So there exists $\delta > 0$ such that $f(x) \geq f(a)$ for all $x \in (a - \delta, a + \delta)$. We know that

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

exists. In particular, this means that the limit as x approaches a from the left and right both exist and equal $f'(a)$. For $x \in (a - \delta, a)$, we have that $x - a < 0$ and $f(x) - f(a) \leq 0$, so the fraction in the above limit is positive for x to the left of a. Thus

$$\lim_{x \to a^-} \frac{f(x) - f(a)}{x - a} = f'(a) \geq 0$$

where $x \to a^-$ means the limit as we approach a from the left. For $x \in (a, a + \delta)$, $x - a > 0$ and $f(x) - f(a) \leq 0$ so the fraction in the above limit is negative. Thus

$$\lim_{x \to a^+} \frac{f(x) - f(a)}{x - a} = f'(a) \leq 0.$$

Since $f'(a) \geq 0$ and $f'(a) \leq 0$, we must have $f'(a) = 0$ as claimed. $\square$

Lecture 16: More on Differentiable Functions

Today we continued talking about differentiable functions, looking at some more basic examples and properties.

Warm-Up 1. Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by

$$f(x) = \begin{cases} x^2 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases}$$

We claim that f is only differentiable at 0. Indeed, to show that f is not differentiable at any $y \neq 0$, we show that it is not even continuous at any such y; this rules out f being differentiable at such y since differentiable implies continuous.

If y is a nonzero rational number, take any sequence of irrationals (y_n) converging to y. (As we've seen before, such a sequence exists since the irrationals are dense in $\mathbb{R}$.) Then $f(y_n) = 0$ for all n so $f(y_n) \to 0$. Thus we have

$$y_n \to y \text{ but } f(y_n) \not\to f(y) = y^2 > 0,$$

so f is not continuous at y.

If y is an irrational number, take a sequence of rationals (y_n) converging to y. Then $f(y_n) = y_n^2$, which converges to $y^2 > 0$ according to some limit laws. Thus

$$y_n \to y \text{ but } f(y_n) \not\to f(y) = 0,$$

so f is not continuous at y. We conclude that f is not continuous at any $y \neq 0$, and thus it is not differentiable at ay $y \neq 0$.

Now, to determine differentiability at 0 we consider the limit

$$\lim_{x \to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x \to 0} \frac{f(x)}{x}.$$

As $x \to 0$, $f(x)$ is either 0 (when x is irrational) or x^2 (when x is rational); in the first case $\frac{f(x)}{x} = 0$ and in the second $\frac{f(x)}{x} = x$. Since $x \to 0$ as $x \to 0$ (duh), no matter when x is rational or irrational we have $f(x)/x \to 0$ so

$$\lim_{x \to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x \to 0} \frac{f(x)}{x} = 0.$$

Thus f is differentiable at 0 and $f'(0) = 0$.

Important Remark. At first glance, you might be tempted to say that since x^2 is differentiable at all $x \in \mathbb{Q}$ and since the constant 0 is differentiable as well, f is differentiable everywhere and

$$f'(x) = \begin{cases} 2x & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$$

by computing the derivative of each term separately. However, this is total nonsense as our argument in the Warm-Up shows. The problem is that since the derivative is defined as a limit, it depends on values *near* the point we're approaching and not the value at that point itself. In other words, just knowing the value of f at a point x is not enough to determine the derivative at x; we need to know how f behaves "close" to x.

In this case, any interval around $x \in \mathbb{Q}$ will contain an irrational y, and at such points f will have the value 0 and *not* y^2. So f does not look like the function x^2 everywhere near $x \in \mathbb{Q}$, so we cannot just simplify use this expression itself to determine differentiability. Similarly, at an irrational f has the value 0 but it does not have the value 0 everywhere near an irrational since any interval around an irrational will contain a rational r, will f has the value r^2.

Comparing with a previous example where we had a function with the value $x^2 \sin \frac{1}{x}$ for $x \neq 0$ and 0 for $x = 0$, in that case at any $x \neq 0$ there is an interval consisting of only nonzero numbers y, and the value of f at those points is still given by $y^2 \sin \frac{1}{y}$. That is, in that case everywhere "near" some $x \neq 0$ the function in question was the same as the function $x^2 \sin \frac{1}{x}$ so we can use what we know about this function to say something about differentiability; that doesn't happen in the function in the first Warm-Up.

Warm-Up 2. We claim that the function $g(x) = x|x|$ is differentiable on all of $\mathbb{R}$. Near $x > 0$ this function is the same as x^2, which is differentiable at any $x > 0$, while near $x < 0$ this function is the same as $-x^2$, which is differentiable at any $x < 0$. Hence g is definitely differentiable at any $x \neq 0$, and to check differentiability at 0 we consider the limit:

$$\lim_{x\to 0} \frac{g(x) - g(0)}{x - 0} = \lim_{x\to 0} \frac{x|x|}{x} = \lim_{x\to 0} |x| = 0,$$

so g is differentiable at 0 and $g'(0) = 0$. The derivative of g is thus given by

$$g'(x) = \begin{cases} 2x & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -2x & \text{if } x < 0, \end{cases}$$

which can be succinctly captured by the formula $g'(x) = 2|x|$.

Remark. Recall my favorite function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & \text{if } x \notin \mathbb{Q} \\ 1 & \text{if } x = 0 \\ \frac{1}{q} & \text{if } x = \frac{p}{q} \text{ where } 0 \neq p \in \mathbb{Z} \text{ and } q \in \mathbb{N} \text{ have no common factors.} \end{cases}$$

(Previously defined this only on $[0, 1]$, but this above modification defines f on all of $\mathbb{R}$.) Since this function is not continuous at $x \in \mathbb{Q}$ (still true even when defined on $\mathbb{R}$), it is not differentiable at any $x \in \mathbb{Q}$. This function is continuous at any $x \notin \mathbb{Q}$ (again, still true when defined on $\mathbb{R}$), but it turns out that it is not differentiable at any $x \notin \mathbb{Q}$ either. So, f is nowhere differentiable. I'll leave it to you to try to show that the limit which defines differentiability at $x \notin \mathbb{Q}$ does not exist, or you can search the interwebs for information on "Thomae's function" to see how it's done.

Proposition. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is differentiable at a and that $f'(a) > 0$. Then there is an interval $(a - \delta, a + \delta)$ around a such that for any $x, y \in (a - \delta, a + \delta)$ with $x < a < y$, we have $f(x) < f(a) < f(y)$.

This is also a well-known from calculus: if the derivative at some point a is positive, then $x < a$ is sent to $f(x) < f(a)$ and $y > a$ is sent to $f(y) > f(a)$, so that f looks like it is increasing "at" a. However, there is a subtlety in interpreting $f'(a) > 0$ as a statement about f being increasing, as we'll point out after the proof.

Proof. Since $f'(a)$ exists and

$$f'(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a} > 0,$$

the ϵ-δ definition of limits implies that there exists $\delta > 0$ such that

$$\frac{f(x) - f(a)}{x - a} > 0 \text{ for any } a \neq x \in (a - \delta, a + \delta).$$

(This comes from applying the definition of limit to the positive number $\epsilon = f'(a)$. The justification is very similar to the one for the second Warm-Up from October 28th, which you can find on the Week 6 Lecture Notes. To be clear, for $f'(a) > 0$ there exists $\delta > 0$ such that

$$\text{if } 0 < |x-a| < \delta, \text{ then } \left|\frac{f(x)-f(a)}{x-a} - f'(a)\right| < f'(a),$$

and the claim follows by rewriting this last inequality.) This implies that

$$f(x) - f(a) \text{ and } x - a$$

have the same sign for $x \in (a-\delta, a+\delta)$. Hence if $x < a < y$, so $x - a < 0$ and $y - a > 0$, we have $f(x) - f(a) < 0$ and $f(y) - f(a) > 0$, so $f(x) < f(a) < f(y)$ as claimed. □

Careful. Based on intuition from calculus, you might guess that $f'(a) > 0$ means that f is increasing near a, so that the graph of f has positive slope at any point close to a. However, this is *not* necessarily true, and might seem kind of strange. The trouble is that all of our calculus intuitions are based on what we know about "simple" functions (polynomials, trig functions, exponentials and logarithms) and might not hold for differentiable functions in general.

To say that f is increasing near a we would need to know that $f'(x) > 0$ for *all* x near a; knowing that the derivative is positive only at a itself is not enough. All we can say in this case, as the above proposition shows, is that any x "close" and to the left of a is sent to something smaller than $f(a)$ and any y "close" and to the right of a is sent to something larger than $f(a)$. For example, we could have a function with positive derivative at a but which oscillates rapidly enough to the left and to the right of a in a way which prevents f from being strictly increasing to the left and to the right of a.

Other well-known properties. Recall other well-known properties of derivatives:

- (Linearity) $(f+g)' = f' + g'$ and $(cf)' = cf'$ for any scalar c,
- (Product Rule) $(fg)' = f'g + fg'$,
- (Quotient Rule) $\left(\frac{f}{g}\right)' = \frac{gf' - fg'}{g^2}$.

The proofs of these are fairly straightforward, and are in the book. From now on feel free to use them whenever you'd like.

Chain Rule. Another well known derivative rule is the chain rule, which says that if f is differentiable at a and g is differentiable at $f(a)$, then the composition $g \circ f$ is differentiable at a and

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

The proof of this is also in the book, and this is the only one of the basic derivative rules which is not so straightforward, and requires some thought.

The idea is actually pretty simple: consider the limit

$$\lim_{x\to a} \frac{g(f(x)) - g(f(a))}{x-a}$$

defining $(g \circ f)'(a)$. We can rewrite the fraction as

$$\frac{g(f(x)) - g(f(a))}{x-a} = \frac{g(f(x)) - g(f(a))}{f(x)-f(a)} \frac{f(x)-f(a)}{x-a}.$$

Now, the limit of the final fraction is precisely $f'(a)$ and the limit of the middle fraction is $g'(f(a))$, which you can see using the fact that $f(x) - f(a) \to 0$ as $x \to a$ since f is continuous at a. Thus

$$\lim_{x\to a} \frac{g(f(x)) - g(f(a))}{x - a} = \lim_{x\to a} \frac{g(f(x)) - g(f(a))}{f(x) - f(a)} \frac{f(x) - f(a)}{x - a} = g'(f(a))f'(a),$$

as the chain rule claims, almost. The problem with this approach is that the fraction

$$\frac{g(f(x)) - g(f(a))}{f(x) - f(a)}$$

is not defined when $f(x) = f(a)$, and it is certainly possible for $f(x)$ to equal $f(a)$ for x approaching a. If this happens, our above limit manipulations are not valid, and we need to do something else.

The real proof, as given in the book, avoids the above complication by not phrasing things in terms of the problematic fraction. It takes a little more effort to make it work, but the end result is the same. So, check the book for the full proof, but the basic point is to take the above idea and figure out how to have it make sense in all cases.

Second Derivative Test. In class we started proving another well-known fact from calculus, the so-called "Second Derivative Test" for classifying local maximums and minimums of a function. We stopped at the point where we needed the Mean Value Theorem to continue, and I said we would finish the proof next time. Really, I should have just saved that entire discussion until after we talk about the Mean Value Theorem on Wednesday to avoid having to stop midway through the proof. So, I'll save the Second Derivative Test for Wednesday's lecture notes where it will fit in much better.

Lecture 17: The Mean Value Theorem

Today we spoke about the Mean Value Theorem and some of its consequences. Many well-known facts from calculus depend on this result, and eventually it will play a crucial role in the Fundamental Theorem of Calculus.

Warm-Up. Suppose that f is a differentiable function. We claim that f' has the following intermediate value property: whenever $f'(a) < y < f'(b)$, there exists c between a and b such that $f'(c) = y$. In other words, derivatives always attain any "intermediate" values. If f' is continuous this is just a consequence of the Intermediate Value Theorem, but the important part is that this is even when f' is *not* continuous. The book has a proof of this, but let I'll reproduce it here in order to fill in some details which the book glosses over.

Suppose without loss of generality that $a < b$ and let F be the function defined by $F(x) = f(x) - yx$. Then F is differentiable on $[a, b]$ since it is the difference of differentiable functions, and hence it is continuous on $[a, b]$. Thus by the Extreme Value Theorem F has a minimum at some point in $[a, b]$. Now, we claim that this minimum does not occur at a nor at b. Indeed, since

$$F'(a) = f'(a) - y < 0,$$

we have that

$$\frac{F(x) - F(a)}{x - a} < 0 \text{ for } x \text{ in some interval } (a - \delta, a + \delta) \text{ around } a.$$

(This is a fact we've seen before: the limit of the fraction above defines $F'(a)$, and if this limit is negative then the expression we take the limit of must be negative near the point a we're approaching.) In particular, for $x \in (a, a+\delta)$ we have that $x - a > 0$ so $F(x) - F(a) < 0$. Hence $F(x) < F(a)$ so the minimum of F does not occur at a. Similarly, since

$$F'(b) = f'(b) - y > 0,$$

we have

$$\frac{F(x) - F(b)}{x - b} > 0 \text{ for } x \text{ in some interval } (b - \delta, b + d\delta) \text{ around } b.$$

In particular for $x \in (b - \delta, b)$, $x - b < 0$ so $F(x) - F(b) < 0$ and again $F(x) < F(b)$ meaning that minimum of F does not occur at b. Thus the minimum of F must occur at some $c \in (a, b)$. At a minimum the derivative must be zero, so $F'(c) = f'(c) - y = 0$, meaning that $f'(c) = y$ and c is the desired point.

Derivatives do not have jump discontinuities. The fact that derivatives have the intermediate value property says that certain functions can never arise as the derivatives of other functions; in particular, any function with a "jump" discontinuity is not the derivative of anything else. For instance, consider the function

$$g(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}.$$

If there was a differentiable function G such that $G' = g$, we would have

$$G'(-1) = g(-1) = -1 < 0 < G'(1) = g(1) = 1,$$

so the Warm-Up would say that there should exist $c \in (-1, 1)$ such that $G'(c) = g(c) = 0$, which is nonsense. Thus there can be no such G; in other words, g has no *antiderivative*.

The problem is that g has a jump discontinuity at 0. Derivatives of course can have discontinuities, but the Warm-Up places restrictions on just what types of discontinuities those can be: the only way in which a derivative might fail to be continuous is because of some weird oscillation behavior, such as what happens with the derivative of the function which is $x^2 \sin \frac{1}{x}$ for $x \neq 0$ and 0 at $x = 0$.

Rolle's Theorem. If f is differentiable at $f(a) = f(b)$, then there exists c between a and b such that $f'(c) = 0$. Actually, the proof of this was essentially included in the Warm-Up: since f is differentiable, f is continuous on $[a, b]$ so it has a maximum and a minimum on $[a, b]$; at least one of these (as long as f is non-constant) must occur in (a, b), and at that maximum or minimum the derivative will be zero. The proof is also in the book.

Mean Value Theorem. If f is differentiable on (a, b) and continuous on $[a, b]$, then there exists $c \in (a, b)$ such that $f(b) - f(a) = f'(c)(b - a)$. The proof of this is in the book, and boils down to applying Rolle's Theorem to the function

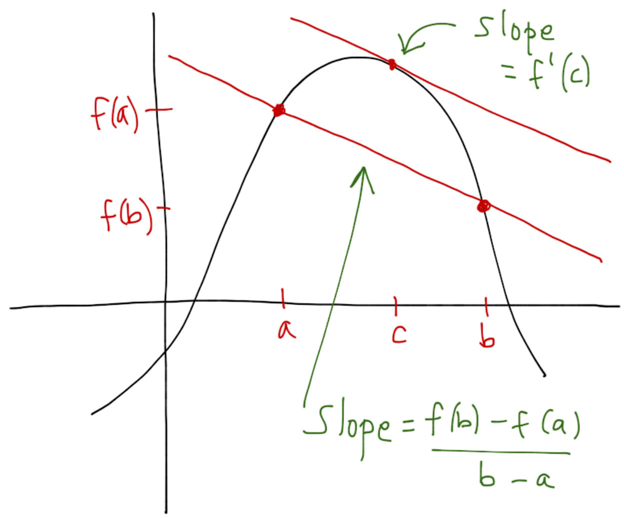
$$F(x) = f(x)(b - a) - x(f(b) - f(a)).$$

Two comments are in order. First, why use the function F defined above? Essentially this function takes f and "shifts" the left and right endpoints by the appropriate amounts in order to make $F(a) = F(b)$. Think of F as a way of "linearly deforming" f to give something to which Rolle's Theorem applies.

Second, the Mean Value Theorem is completely obvious if you draw a picture of what it says. Take the graph of a differentiable function f and draw the points $(a, f(a))$ and $(b, f(b))$ on the graph. The line passing through these two points has slope

$$\frac{f(b) - f(a)}{b - a},$$

and from the picture:



it looks as though there should be some point $c \in (a, b)$ at which the slope of the tangent line has same slope as this line above:

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

This is precisely what the Mean Value Theorem says, after rewriting this equation. The Mean Value Theorem is the most important tool we'll have for comparing differences of the form $f(b) - f(a)$ to those of the form $b - a$.

Consequences. The Mean Value Theorem is how we can justify many well-known facts from calculus. For instance, the fact that zero derivative on an interval implies constant function on that interval is an application of the Mean Value Theorem, as is the fact that positive (respectively negative) derivative on an interval implies increasing (respectively decreasing) function on that interval. (To clarify this last property, we saw earlier that having $f'(a) > 0$ at a single point a was not enough to say that f was increasing near a; it is crucial that f have positive derivative not just at a but everywhere near a in order for this to be true.)

All of these facts are pretty much impossible to prove without the Mean Value Theorem. For instance, knowing that f has zero derivative on an interval means that for any a in that interval

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = 0.$$

But, this doesn't say much about the fraction we're taking the limit of since you can definitely have a nonzero expression which gives a limit of zero. So, we can't directly conclude that the numerator must be zero; we need some way of comparing $f(x) - f(a)$ to $x - a$, which is precisely what the Mean Value Theorem gives us.

Second Derivative Test. Suppose that f is differentiable, that $f'(a) = 0$ and that $f''(a) > 0$. We claim that f then has a *local minimum* at a, meaning there exists an interval $(a - \delta, a + \delta)$ around a such that $f(a) \leq f(x)$ for all $x \in (a - \delta, a + \delta)$. This is also well-known from calculus (as is the corresponding test for local maximums, which is on the homework), and says that a critical point at which the graph of f is concave up is a local minimum.

To prove this, start with

$$\lim_{x \to a} \frac{f'(x) - f'(a)}{x - a} = f''(a) > 0.$$

Since this limit is positive, there exists $\delta > 0$ such that

$$\frac{f'(x) - f'(a)}{x - a} > 0 \text{ for } x \in (a - \delta, a + \delta).$$

Recall that $f'(a) = 0$, so this inequality means that

$$f'(x) \text{ and } x - a \text{ have the same sign for } x \in (a - \delta, a + \delta).$$

Now, take any $x \in (a - \delta, a + \delta)$. By the Mean Value Theorem there exists c between x and a such that

$$f(x) - f(a) = f'(c)(x - a).$$

Now, either $x < c < a$ or $a < c < x$, and either way $x - a$ and $c - a$ have the same sign, and thus $f'(c)$ and $x - a$ have the same sign. This means that

$$f(x) - f(a) = f'(c)(x - a) \geq 0, \text{ so } f(x) \geq f(a) \text{ for } x \in (a - \delta, a + \delta).$$

Hence f has a local minimum at a as claimed.

One More Mean Value Application. Suppose that f is continuous on $\mathbb{R}$, differentiable at all $x \neq a$, and that $\lim_{x \to a} f'(x) = L$ exists. Then f is differentiable at a as well and $f'(a) = L$. Before looking at the proof, note again that this places a restriction on how badly derivatives can actually behave. In particular, if f' exists everywhere near a point a and the limit of the derivative exists as you approach a, this fact says that f' is actually continuous at a! As in the Warm-Up, this rules out the possibility that a function with a jump discontinuity can be the derivative of another function.

To prove the claim, consider the limit defining $f'(a)$:

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

For any fixed x, f is differentiable on the open interval between x and a so the Mean Value Theorem says that there exists some c_x in this interval such that

$$f(x) - f(a) = f'(c_x)(x - a).$$

Substituting this above gives

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{x \to a} \frac{f'(c_x)(x - a)}{x - a} = \lim_{x \to a} f'(c_x).$$

Since c_x is sandwiched between x and a, as $x \to a$ we also have $c_x \to a$. Thus the above limit is the same as

$$\lim_{c_x \to a} f'(c_x),$$

which exists and equals L by our assumption on f. Hence $f'(a)$ exists and $f'(a) = L$.

Lecture 18: Taylor's Theorem

Today we spoke about Taylor's Theorem, which you should view as a generalization of the Mean Value Theorem to higher-order derivatives. In practice, this is one of the most useful tools available when it comes to developing techniques for approximating functions.

Warm-Up. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is differentiable and that its derivative is bounded everywhere. We claim that f is then uniformly continuous. Say that M is a bound for f', so $|f'(x)| \leq M$ for all $x \in \mathbb{R}$. For any x and y, by the Mean Value Theorem there exists $c \in \mathbb{R}$ such that

$$|f(x) - f(y)| = |f'(c)||x - y| \leq M|x - y|.$$

This says that f is a Lipschitz function, and the second problem on Homework 6 says that f is thus uniformly continuous.

Remark. Even though this result is good for visualizing what the graph of a uniformly continuous function might look like, it does not describe all possible uniformly continuous functions. For one thing, not every uniformly continuous function is differentiable, and this fact only applies to differentiable functions. More importantly, the converse of this fact is not true: if f is differentiable and uniformly continuous, it is not necessarily true that f' must be bounded. For instance, the function $f(x) = \sqrt{x}$ is uniformly continuous on $(0, \infty)$ but its derivative is unbounded there.

So, be careful when applying the fact from the Warm-Up, but it does give a pretty good intuitive idea as to what uniformly continuous functions look like, in that they should not change "too rapidly".

The Mean Value Theorem restated. The equality $f(x) - f(x_0) = f'(c)(x - x_0)$ obtained from the Mean Value Theorem can be written as

$$f(x) = f(x_0) + f'(c)(x - x_0).$$

You might recognize this right-hand side as the type of thing which starts off a "Taylor expansion" of f, except with $f'(c)$ instead of $f'(x_0)$. From this point of view, by expanding the right-side above to include higher-order terms, the Mean Value Theorem leads to a certain generalization known as *Taylor's Theorem.* The book contains one proof of Taylor's Theorem, but I'll give a different one which better emphasizes the role which the Mean Value Theorem plays; indeed, Taylor's Theorem will be obtained by repeated applications of the Mean Value Theorem.

Taylor polynomials and remainders. Suppose f is n-times differentiable. The *n-th order Taylor polynomial of f centered at x_0* is the polynomial $P_n(x)$ given by

$$P_n(x) = f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n.$$

Note that the derivatives (up to the n-th one) of this polynomial evaluated at x_0 agree with those of f, which you can see from a direct computation:

$$f^{(k)}(x_0) = P_n^{(k)}(x_0) \text{ for } k = 0, 1, \ldots, n.$$

The corresponding *Taylor remainder* $R(x)$ is the difference

$$R(x) = f(x) - P_n(x).$$

The point is that the Taylor polynomials $P_n(x)$ give a way to approximate $f(x)$, and the remainder $R(x)$ is the "error" in this approximation; i.e. how far off it is. Note that since f and P_n have the same k-th derivatives at x_0 for $k = 0, 1, \ldots, n$, the remainder has

$$R^{(k)}(x_0) = 0 \text{ for } k = 0, 1, \ldots, n.$$

Since $P_n(x)$ is a polynomial of order n, its $(n+1)$-st derivative is zero so for f which is $(n+1)$-times differentiable, we have

$$R^{(n+1)}(x) = f^{(n+1)}(x).$$

Taylor's Theorem is all about giving an explicit formula for this remainder.

Special case of Taylor's Theorem. Before looking at Taylor's Theorem in its full generality, let's consider the special case of Taylor's Theorem when $n = 1$ to get a feel for how the proof works. In this case, the claim is that if f is twice differentiable, for any x and x_0 there is some c between x and x_0 satisfying

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(c)}{2!}(x - x_0)^2.$$

Consider the function h of a variable y defined by

$$h(y) = f(y) - f(x_0) - f'(x_0)(y - x_0) - \frac{R(x)}{(x - x_0)^2}(y - x_0)^2,$$

where $R(x) = f(x) - f(x_0) - f'(x_0)(x - x_0)$ is the first order Taylor remainder of f centered at x_0. The function h is twice differentiable with respect to y and its derivatives with respect to y are:

$$\begin{aligned} h'(y) &= f'(y) - f'(x_0) - 2\frac{R(x)}{(x - x_0)^2}(y - x_0), \text{ and} \\ h''(y) &= f''(y) - 2\frac{R(x)}{(x - x_0)^2}. \end{aligned}$$

The function h satisfies:

$$\begin{aligned} h(x_0) &= f(x_0) - f(x_0) - f'(x_0)(x_0 - x_0) - \frac{R(x)}{(x - x_0)^2}(x_0 - x_0)^2 = 0, \\ h'(x_0) &= f'(x_0) - f'(x_0) - 2\frac{R(x)}{(x - x_0)^2}(x_0 - x_0) = 0, \text{ and} \\ h(x) &= f(x) - f(x_0) - f'(x_0)(x - x_0) - \frac{R(x)}{(x - x_0)^2}(x - x_0)^2 \\ &= f(x) - f(x_0) - f'(x_0)(x - x_0) - R(x) = 0. \end{aligned}$$

By the Mean Value Theorem there exists x_1 between x and x_0 such that

$$0 - 0 = h(x) - h(x_0) = h'(x_1)(x - x_0), \text{ so } h'(x_1) = 0.$$

Now applying the Mean Value Theorem to h' says that there exists c between x_1 and x_0 such that

$$0 - 0 = h'(x_1) - h'(x_0) = h''(c)(x_1 - x_0), \text{ so } h''(c) = 0.$$

But $h''(c) = 0$ is

$$f''(c) - 2\frac{R(x)}{(x - x_0)^2} = 0, \text{ so } R(x) = \frac{f''(c)}{2}(x - x_0)^2.$$

Recalling that $R(x) = f(x) - f(x_0) - f'(x_0)(x - x_0)$ gives the desired claim. Since c is between x_1 and x_0, and x_1 is between x and x_0, c is indeed between x and x_0.

Taylor's Theorem. Suppose that f is $(n + 1)$-times differentiable. Then for any x and x_0 there exists c between x and x_0 such that

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(c)}{(n + 1)!}(x - x_0)^{n+1}.$$

Letting $P_n(x)$ denote the n-th order Taylor polynomial of f centered at x_0 and $R(x) = f(x) - P_n(x)$ the corresponding Taylor Remainder, the point is that we get the following explicit expression for this remainder:

$$R(x) = \frac{f^{(n+1)}(c)}{(n + 1)!}(x - x_0)^{n+1}$$

for some c between x and x_0. We'll use Taylor remainders in the proof to keep the notation clean; recall that these remainders satisfy

$$R^{(k)}(x_0) = 0 \text{ for } k = 0, 1, \dots, n \text{ and } R^{(n+1)}(x_0) = f^{(n+1)}(x_0).$$

Proof. For any $x \neq x_0$, define the function $h(y)$ by

$$h(y) = R(y) - \frac{R(x)}{(x - x_0)^{n+1}}(y - x_0)^{n+1}$$

where $R(y)$ is the n-th order Taylor polynomial of f centered at x_0. Since $f(y)$ is $(n + 1)$-times differentiable and $(y - x_0)^{n+1}$ is $(n+1)$-times differentiable, $h(y)$ is $(n+1)$-times differentiable with respect to y and a direct computation gives

$$h^{(k)}(y) = R^{(k)}(y) - \frac{R(x)}{(x - x_0)^{n+1}}(n + 1)n(n - 1) \cdots (n + 2 - k)(y - x_0)^{n+1-k}$$

for $k = 1, 2, \dots, n + 1$. In particular, $h^{(k)}(x_0) = 0$ for $k = 1, \dots, n$.

Now, $h(x) = 0$ and $h(x_0) = R(x_0) = 0$, so by the Mean Value Theorem there exists x_1 between x and x_0 such that

$$0 - 0 = h(x) - h(x_0) = h'(x_1)(x - x_0), \text{ and hence } h'(x_1) = 0.$$

Now, $h'(x_1) = 0$ and $h'(x_0) = R'(x_0) = 0$, so again by the Mean Value Theorem there exists x_2 between x_1 and x_0 such that

$$0 - 0 = h'(x_1) - h'(x_0) = h''(x_2)(x_1 - x_0), \text{ and hence } h''(x_2) = 0.$$

Continuing in the same manner and using the fact that $h^{(k)}(x_0) = 0$ for $k = 1, \dots, n$ produces terms $x_1, x_2, x_3, \dots, x_n$ such that x_k is between x_{k-1} and x_0 and $h^{(k)}(x_k) = 0$; to be clear, these terms come from applying the Mean Value Theorem at each step to $0 - 0 = h^{(k-1)}(x_{k-1}) - h^{(k-1)}(x_0)$. At the end, by the Mean Value Theorem we get c between x_n and x_0 such that

$$0 - 0 = h^{(n)}(x_n) - h^{(n)}(x_0) = h^{(n+1)}(c)(x_n - x_0), \text{ so } h^{(n+1)}(c) = 0.$$

Recalling the computation of the derivatives of h above, this gives

$$R^{(n+1)}(c) - \frac{R(x)}{(x - x_0)^{n+1}}(n + 1)! = 0,$$

and since $R^{(n+1)}(c) = f^{(n+1)}(c)$, this equation becomes

$$R(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x - x_0)^{n+1}$$

as desired. If $x < x_0$, the terms constructed satisfy

$$x < x_1 < x_2 < \cdots < x_n < c < x_0,$$

while if $x > x_0$ the opposite inequalties hold; either way, c is indeed between x and x_0. □

Remark. When $n = 0$, Taylor's Theorem is precisely the statement of the Mean Value Theorem, so not only does the Mean Value Theorem imply Taylor's Theorem as above, the Mean Value Theorem is also a special case.

Example 1. Say we want to approximate the value of $\sin x$ for some x. The $(2n+1)$-st order Taylor polynomial for $f(x) = \sin x$ centered at 0 looks like

$$P_{2n+1}(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + \frac{(-1)^n}{(2n+1)!}x^{2n+1}.$$

Taylor's Theorem says that for any x there exists c between x and 0 such that

$$|\sin x - P_{2n+1}(x)| = \left| \frac{f^{(2n+2)}(c)}{(2n+2)!}x^{2n+2} \right|.$$

Since all derivatives of $f(x) = \sin x$ are bounded by 1, this gives

$$|\sin x - P_{2n+1}(x)| \leq \frac{1}{(2n+2)!}|x|^{2n+2}.$$

This then gives us a way to approximate $\sin x$ to any degree of accuracy we want. For instance, to approximate $\sin x$ to four decimal places, all we need to do is pick n such that

$$\frac{|x|^{2n+2}}{(2n+2)!} < \frac{1}{10000}.$$

For such an n we will have

$$|\sin x - P_{2n+1}(x)| < \frac{1}{10000},$$

so $P_{2n+1}(x)$ and $\sin x$ will agree to four decimal places.

Example 2. We claim that for any $x > 0$ and any $n \in \mathbb{N}$,

$$1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} < e^x.$$

Indeed, Taylor's Theorem gives

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \frac{e^c}{(n+1)!}x^{n+1}$$

for some c between x and 0. This the last term is positive, so this gives the desired inequality. (This inequality is used in showing that the Taylor series for e^x actually converges to e^x.)

L'Hopital's Rule and the Inverse Function Theorem. The rest of Chapter 4 in the book covers L'Hopital's Rule and the Inverse Function Theorem. These are both important results, but we won't cover them explicitly in class. In particular, L'Hopital's rule is no doubt something you saw in a calculus course before, and the Inverse Function Theorem, although still interesting in the single variable case, doesn't really shine until you look at its higher-dimensional formulation. So, I'll leave it to you to look over the statements of these; feel free to use either whenever it might be applicable, but they won't form an important part of any future material.

Lecture 19: Integration

Today we started talking about integration, which will be our final and probably most important topic. We all of course know that integrals give areas under graphs of functions and we've all spent countless hours computing integrals before, and yet not many of us have probably thought much about what "integral" really means. Before we can possibly hope to apply integrals to concrete problems, we need a precise definition of "integral" in order to justify various properties we all know and love. That will be our focus over these final weeks.

Warm-Up. Say we want to approximate $\log x$ for $x \in [1, 2]$ to three decimal places. We can hope to do so using a Taylor polynomial for $f(x) = \log x$. First let's compute the derivatives of $\log x$:

$$\begin{aligned} f'(x) &= \frac{1}{x} \\ f''(x) &= -\frac{1}{x^2} \\ f'''(x) &= \frac{2}{x^3} \\ f^{(4)}(x) &= -\frac{2 \cdot 3}{x^4}, \end{aligned}$$

and in general

$$f^{(k)}(x) = (-1)^{k-1}\frac{(k-1)!}{x^k} \text{ for } k \geq 1.$$

Thus the n-th order Taylor polynomial of f centered at 1 is

$$\begin{aligned} P_n(x) &= f(1) + f'(1)(x-1) + \cdots + \frac{f^{(n)}(1)}{n!}(x-1)^n \\ &= (x-1) - \frac{(x-1)^2}{2} + \cdots + \frac{(-1)^{n-1}(x-1)^n}{n}. \end{aligned}$$

For a fixed $x \in [1, 2]$, by Taylor's Theorem there exists c between 1 and x such that

$$\log x - P_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-1)^{n+1} = \frac{(-1)^n(x-1)^{n+1}}{c^k(n+1)}.$$

Since $c > 1$, this gives the bound

$$|\log x - P_n(x)| = \frac{(x-1)^{n+1}}{c^k(n+1)} \leq \frac{(x-1)^{n+1}}{n+1}.$$

Thus, if we pick n large enough so that

$$\frac{(x-1)^{n+1}}{n+1} < \frac{1}{1000},$$

for this n the Taylor polynomial $P_n(x)$ satisfies

$$|\log x - P_n(x)| < \frac{1}{1000},$$

and thus the first three decimal places of $P_n(x)$ and $\log x$ will agree, giving us the desired approximation.

Inverse function example. I mentioned again today that we won't talk about the Inverse Function Theorem in class, and that you should read up about it on your own. But, for good measure, here is one example, which we didn't do in class. The Inverse Function Theorem states the following: if f is invertible and $f'(a) \neq 0$, then the inverse f^{-1} is differentiable at $b = f(a)$ and

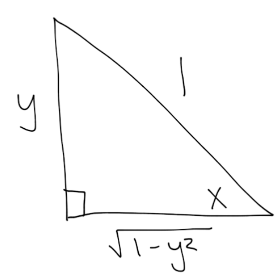
$$(f^{-1})'(b) = \frac{1}{f'(a)}.$$

The important part is that f^{-1} is differentiable at b; the formula for the derivative of the inverse comes from applying the chain rule to $f^{-1}(f(a)) = a$.

Consider the function $f(x) = \sin x$ on $(-\frac{\pi}{2}, \frac{\pi}{2})$. Everywhere on this interval $f'(x) = \cos x$ is nonzero, so the Inverse Function Theorem implies that $f^{-1}(y) = \arcsin y$ is differentiable. For $y = \sin x$, we have

$$(f^{-1})'(y) = \frac{1}{f'(x)} = \frac{1}{\cos x}.$$

Now, consider a right triangle with hypotenuse 1 and a non-right angle x:

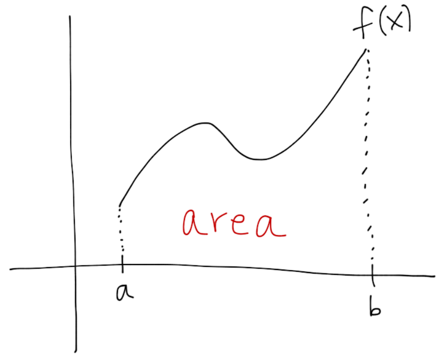


The side opposite this angle has length $y = \sin x$, and the adjacent side has length $\cos x$. By the Pythagorean Theorem, this adjacent side also has length $\sqrt{1-y^2}$, so $\cos x = \sqrt{1-y^2}$ and we get that

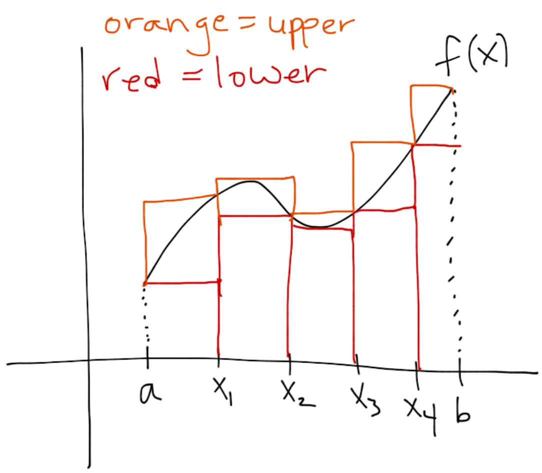
$$\text{derivative of } \arcsin y = \frac{1}{\sqrt{1-y^2}},$$

a well-known formula from calculus.

Motivation for integration. An integral should give us the area of the region under the graph of a function:



To compute this area, we approximate it using areas of rectangles as follows. First, we divide the interval $[a, b]$ into smaller pieces. Over each of these, we take a rectangle of height equal to the infimum of f over that piece, and a rectangle of height equal to the supremum of f over that piece:



The actual area we want is sandwiched between the sum of the areas of the "lower" rectangles and the sum of the areas of the "upper" rectangles. The idea is that by considering all possible such sums corresponding to all possible ways of breaking up $[a, b]$ into smaller pieces, we can get better and better approximations to the area we want. The goal is now to make this all precise.

Definitions. A *partition* P of $[a, b]$ is a collection of points $x_0, \ldots, x_n$ such that

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

The practical point is that this "breaks" the interval $[a, b]$ up into the smaller intervals

$$[x_0, x_1],\ [x_1, x_2], \ldots, [x_{n-2}, x_{n-1}], \text{ and } [x_{n-1}, x_n].$$

For a bounded function f on $[a, b]$ and a partition P of $[a, b]$, the *lower Riemann sum* $L(f, P)$ is (letting I_k denote the k-th subinterval $[x_{k-1}, x_k]$ determined by the partition):

$$L(f, P) = \sum_{I_k} (\inf f \text{ over } I_k)(\text{length of } I_k)$$

and the *upper Riemann sum* $U(f,P)$ is

$$U(f,P) = \sum_{I_k} (\sup f \text{ over } I_k)(\text{length of } I_k).$$

Graphically, $L(f,P)$ is the sum of the areas of the "lower" rectangles in the above picture and $U(f,P)$ is the sum of the areas of the "upper" rectangles.

Remark. Note in the definition above we require that f be bounded in order to guarantee that the infimums and supremums used actually exist. However, this is not really much of a restriction on f: later we will see another approach to integration in terms of another type of "Riemann sum", and it will turn out that if a function is "integrable" with respect to this other approach then that function must in fact be bounded. So, we do not lose anything now assuming that our functions are bounded.

Example 1. Suppose that $f(x) = c$ is a constant function on $[a,b]$. Then for any partition P of $[a,b]$, the supremum of f over any smaller subinterval is always c, so

$$U(f,P) = \sum_{I_k} (\sup f \text{ over } I_k)(\text{length of } I_k) = \sum_{I_k} c(\text{length of } I_k) = c\sum_{I_k} (\text{length of } I_k).$$

But the intervals I_k together make up all of $[a,b]$, so adding together their lengths gives the length of $[a,b]$. Thus

$$U(f,P) = c(b-a), \text{ and similarly } L(f,P) = c(b-a)$$

since the infimum of f over any smaller interval is also always c.

Note that this makes sense graphically: the graph of a constant function is a horizontal line, and any "lower" or "upper" rectangles we use should cover the entire region under the graph, which has area $c(b-a)$.

Example 2. Consider the function f on $[0,1]$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases}$$

No matter what partition P of $[0,1]$ we take, the supremum of f over any subinterval is 1 since any subinterval contains a rational and the infimum of f over any subinterval is 0 since any subinterval contains an irrational. This means that

$$U(f,P) = 1 \text{ and } L(f,P) = 0$$

for any partition P of $[0,1]$.

Definition. The *upper Riemann integral* of f over $[a,b]$ is

$$(U)\int_a^b f(x)\,dx = \inf\{U(f,P) \mid P \text{ is a partition of } [a,b]\}$$

and the *lower Riemann integral* of f over $[a,b]$ is

$$(L)\int_a^b f(x)\,dx = \sup\{L(f,P) \mid P \text{ is a partition of } [a,b]\}.$$

Again, the intuition is that all upper sums overestimate the area under the graph of f, so this area should be $\leq$ the infimum of all upper sums, and all lower sums underestimate the area under the graph of f, so this area should be $\geq$ the supremum of all lower sums.

We say that f is *integrable* over $[a, b]$ when the upper Riemann integral and lower Riemann integrals agree, in which case we call this common value the *integral* of f over $[a, b]$:

$$\int_a^b f(x)\,dx = (U)\int_a^b f(x)\,dx = (L)\int_a^b f(x)\,dx.$$

To say that the upper and lower integrals are the same means precisely the area under the graph of f is well-defined.

Remark. I'm approaching integration in a slightly different way than the book does. The book doesn't define "integrable" in terms of upper and lower integrals, but rather using the idea that the difference $U(f, P) - L(f, P)$ between upper and lower sums should get closer and closer to 0. Then later the book shows that this definition agrees with the one in terms of upper and lower integrals. I'm doing it the other way around: defining integration in terms of upper and lower integrals and then showing that this can be phrased in terms of differences $U(f, P) - L(f, P)$. Both approaches are pretty much the same; I'm doing it the way I am simply because every other book I've ever used has done it this way.

Back to Example 1. Back to the constant function example, since all upper sums equal $c(b - a)$ the upper integral is $c(b - a)$, and since all lower sums equal to $c(b - a)$ the lower integral is also $c(b - a)$. Thus a constant function is integrable and

$$\int_a^b c\,dx = c(b - a),$$

which is the expected area under the graph of $f(x) = c$.

Back to Example 2. The function f of Example 2 has all upper sums equal to 1 and all lower sums equal to 0, so the upper integral is 1 and the lower integral is 0. Thus

$$(U)\int_a^b f(x)\,dx \neq (L)\int_a^b f(x)\,dx,$$

so f is not integrable over $[0, 1]$. This means that the area under the graph of f is not well-defined.

Remark. Actually, this is a lie. On the last day of class I hope to briefly talk about what's called *Lebesgue integration*, which is a more general approach to integration than the Riemann integration we are looking at. Using this more general approach, the region under the graph of the function from Example 2 *does* have a well-defined area: it is zero. This illustrates that the Riemann integral, although having many important and useful properties, is not quite strong enough to capture all types of "areas" we would ever want to compute. Still, the Riemann integral was the first type of integral ever rigorously defined, and the discovery of more general types of integrals would not have been possible without having this starting point.

Lecture 20: More on Integration

Today we continued talking about integration, giving an equivalent characterization of "integrability" which in most instances is the one we should actually use when determining whether or not a function is integrable.

Warm-Up. We compute explicitly the upper and lower sums of $f(x) = x$ on the interval $[0, b]$ determined by the partition P_n given by the points $x_k = \frac{kb}{n}$:

$$0 < \frac{b}{n} < \frac{2b}{n} < \cdots < \frac{(n-1)b}{n} < b.$$

The point of this partition is that all the partition points are evenly spaced, so that all subintervals $I_k = [x_{k-1}, x_k]$ have the same length $\frac{b}{n}$; this will be important in finding explicit values for the upper and lower sums.

Since $f(x) = x$ is strictly increasing, its supremum on $I_k = [x_{k-1}, x_k]$ is $f(x_k) = x_k$ and its infimum on I_k is $f(x_{k-1}) = x_{k-1}$. Thus

$$\begin{aligned} U(f, P_n) &= \sum_{I_k} (\sup f \text{ over } I_k)(\text{length of } I_k) \\ &= \sum_{k=1}^{n} \left(\frac{kb}{n}\right)\left(\frac{b}{n}\right) = \frac{b^2}{n^2}\sum_{k=1}^{n} k = \frac{b^2 n(n+1)}{2n^2} = \frac{b^2(n+1)}{2n} \end{aligned}$$

and

$$\begin{aligned} L(f, P_n) &= \sum_{I_k} (\inf f \text{ over } I_k)(\text{length of } I_k) \\ &= \sum_{k=1}^{n} \left(\frac{(k-1)b}{n}\right)\left(\frac{b}{n}\right) = \frac{b^2}{n^2}\sum_{k=1}^{n} (k-1) = \frac{b^2(n-1)n}{2n^2} = \frac{b^2(n-1)}{2n}, \end{aligned}$$

where we have used the fact that $1 + 2 + \cdots + \ell = \frac{\ell(\ell+1)}{2}$. Again, note that we were only able to compute this explicitly due to the fact that all subintervals in P_n had the same length.

Remark. From the computations above we have

$$L(f, P_n) \le \frac{b^2}{2} \le U(f, P_n).$$

Since both $L(f, P_n)$ and $U(f, P_n)$ actually converge to $\frac{b^2}{2}$, this suggests the the lower integral and upper integral of $f(x) = x$ over $[0, b]$ should both equal $\frac{b^2}{2}$, meaning that f is integrable over $[0, b]$ with integral equal to:

$$\int_0^b x\,dx = \frac{b^2}{2}.$$

Now, of course we know from calculus that this is absolutely true, but this is not something we can fully conclude just yet.

The problem is that the upper integral is supposed to be the infimum of *all* possible upper sums and the lower integral the supremum of *all* possible lower sums, and so far we only know these sums for the special partitions P_n where all partition points are evenly spaced. Knowing that the infimum of the specific upper sums $U(f, P_n)$ is $\frac{b^2}{2}$ is not enough (yet) to say that the infimum of all possible upper sums is also $\frac{b^2}{2}$. (I claimed this was so in class, but actually my argument was not quite right; in particular, I claimed that $\frac{b^2}{2}$ was a lower bound for *all* possible upper sums, but in reality we don't actually know that just yet.) Similarly, knowing that the supremum of the specific lower sums $L(f, P_n)$ is $\frac{b^2}{2}$ is by itself not enough to conclude that the lower integral has this same value.

This illustrates a problem with using the upper and lower integrals to check for integrability: for most random partitions P, the values of $U(f, P)$ and $L(f, P)$ are simply impossible to compute directly, and hence it is not feasible that we can directly find the supremum of all lower sums and the infimum of all upper sums. We need another way to test for integrability which avoids having to check all possible partitions. Fortunately, there is such a method, but before stating it we give some other basic facts about upper and lower sums in general.

Definition and basic properties of refinements. Given a partition P of $[a, b]$, a *refinement* of P is a partition P' where we take P and throw in additional partition points. This has the practical effect of taking the subintervals determined by P and breaking them up even further.

We are interested in what happens to upper and lower sums when taking a refinement of a given partition; in other words, what's the relation between $U(f, P)$ and $U(f, P')$, and between $L(f, P)$ and $L(f, P')$? The book has this as Remark 5.7, but I think the argument it gives obscures the idea, which is actually really simple. Suppose we had a subinterval $[x_{k-1}, x_k]$ for P which was broken up into two pieces after adding one more partition point s to create P':

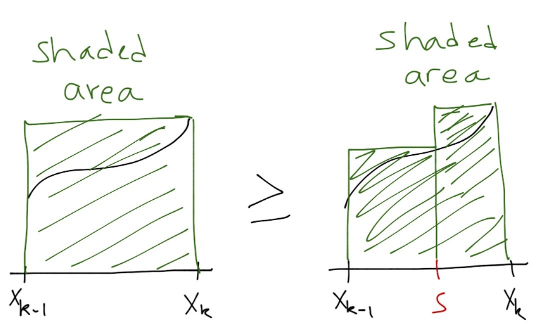
In the sum making up $U(f, P)$ we have a term which looks like

$$(\sup f \text{ over } [x_{k-1}, x_k])(\text{length of } [x_{k-1}, x_k])$$

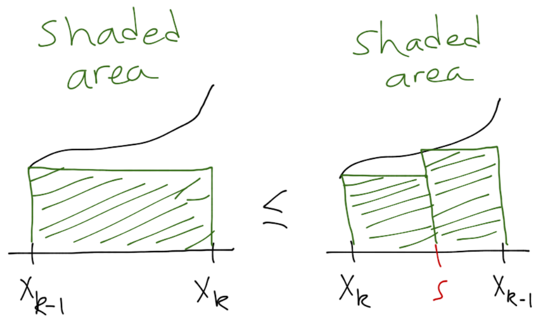
and corresponding to this in the sum making up $U(f, P')$ we have two terms which look like

$$(\sup f \text{ over } [x_{k-1}, s])(\text{length of } [x_{k-1}, s]) + (\sup f \text{ over } [s, x_k])(\text{length of } [s, x_k]).$$

But the supremum of f over all of $[x_{k-1}, x_k]$ is $\geq$ its supremum over either smaller interval $[x_{k-1}, s]$ or $[s, x_k]$, so the first expression above is $\geq$ the sum in the second expression; graphically we are saying that



For infimums the opposite is true: the infimum of f over all of $[x_{k-1}, x_k]$ is $\leq$ its infimum over either smaller interval, so



The same things are true for any subinterval of P which was broken up into smaller intervals in P', so we conclude that

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

In other words, adding more points to your partition either decreases an upper sum or keeps it the same, and either increases a lower sum or keeps it the same. Intuitively, upper sums get "smaller" under refinements and lower sums get "bigger".

Proposition. For any partitions P and Q, we have $L(f,P) \leq U(f,Q)$. That is, any lower sum whatsoever is $\leq$ any upper sum whatsoever.

Proof. Let $P \cup Q$ denote the partition formed by taking the points of P together with the points of Q; note that this is a refinement of both P and Q. By the property of refinements given above, we have

$$L(f,P) \leq L(f,P\cup Q) \leq U(f,P\cup Q) \leq U(f,Q),$$

where the middle inequality comes from the fact that for a single partition, the lower sum is always $\leq$ the upper sum since once uses $\inf f$ and the other $\sup f$. Thus $L(f,P) \leq U(f,Q)$ as claimed. □

Remark. This is actually an important fact! It says that a single lower sum is a lower bound for the set of all possible upper sums, and a single upper sum is an upper bound for the set of all possible lower sums. Thus, the set of all upper sums is bounded below, so its infimum exists, and the set of all lower sums is bounded above, so its supremum exist. Hence the lower and upper integrals of any bounded function always exist and

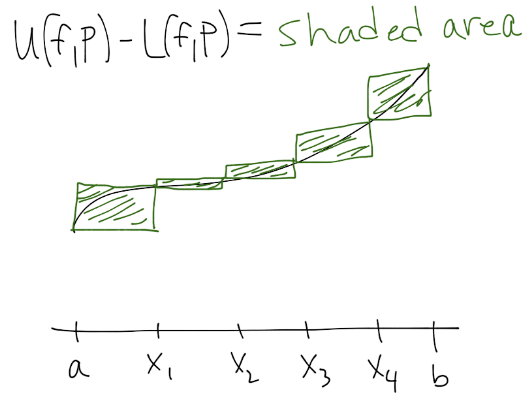
$$(L)\int_a^b f(x)\,dx \leq (U)\int_a^b f(x)\,dx$$

always holds. Note that this inequality should be intuitively clear based on the fact that upper sums overestimate the area under the graph of f and lower sums underestimate this area, but it takes some work to actually justify precisely!

Theorem. A bounded function f is integrable on $[a,b]$ if and only if for any $\epsilon > 0$ there exists a partition P of $[a,b]$ such that $U(f,P) - L(f,P) < \epsilon$.

Remark. As I alluded to earlier, this condition is actually the book's definition of "integrable", which the book then proves that this is equivalent to the definition we gave. Still, since I approached integration in a different order than the book, I'll include a full proof of this here.

The idea behind this theorem is very simple: integrable means that the infimum of the upper sums to equal the supremum of the lower sums, and intuitively this suggests that upper and lower sums should get "closer" to each other, which is what the condition in this theorem says. Graphically, $U(f,P) - L(f,P)$ is the sum of the areas of the small rectangles "between" the upper and lower sums:



and the condition in the theorem says that this sum of small areas can be made arbitrarily small.

Proof of Theorem. Suppose that f is integrable on $[a,b]$, so that the upper and lower integrals are the same:

$$(U)\int_a^b f(x)\,dx = (L)\int_a^b f(x)\,dx.$$

Call this common value I to make notation simpler. Let $\epsilon > 0$. By the alternate characterization of supremums and infimums (in terms of ϵ's) there exists a partition P of $[a,b]$ such that

$$I - \frac{\epsilon}{2} < L(f,P)$$

and there exists a partition Q of $[a,b]$ such that

$$U(f,Q) < I + \frac{\epsilon}{2}.$$

Thus for the partition $P \cup Q$, which is a refinement of both P and Q, we have

$$U(f, P\cup Q) - L(f, P\cup Q) \leq U(f,Q) - L(f,P) < \left(I + \frac{\epsilon}{2}\right) - \left(I - \frac{\epsilon}{2}\right) = \epsilon,$$

where the second inequality follows from the fact that we replaced $U(f,Q)$ by something larger and $L(f,P)$ by something smaller. Thus $P \cup Q$ satisfies the requirement in the statement of the theorem.

Conversely, suppose that for any $\epsilon > 0$ there exists a partition P of $[a,b]$ such that $U(f,P) - L(f,P) < \epsilon$. Then for such a partition we have

$$(U)\int_a^b f(x)\,dx - (L)\int_a^b f(x)\,dx \leq U(f,P) - L(f,P) < \epsilon$$

where again the first inequality follows from replacing the first term by something larger and the second by something smaller. This says that the nonnegative number

$$(U)\int_a^b f(x)\,dx - (L)\int_a^b f(x)\,dx$$

is smaller than any $\epsilon > 0$, and so must be zero. Hence the upper and lower integrals are the same so f is integrable on $[a,b]$. □

Remark. In practice, the condition given in this theorem is the one you should use when wanting to show a function is integrable. However, this definition says nothing about what the value of the integral should be! The way things usually work is we show integrability using this characterization, and then use either the upper or lower integral (or a different type of "Riemann sum" we will soon discuss) to actually find the value of the integral. Or, you appeal to the Fundamental Theorem of Calculus (still to come!) when it is applicable, but the point is that it is not always applicable.

Back to Warm-Up. Now we can justify that the integral from the Warm-Up is indeed $\frac{b^2}{2}$. For the partitions we used there, we have

$$U(f, P_n) - L(f, P_n) = \frac{b^2(n+1)}{2n} - \frac{b^2(n-1)}{2n} = \frac{b^2}{n}.$$

Thus, for $\epsilon > 0$, picking n such that $\frac{b^2}{n} < \epsilon$ gives a partition P_n satisfying the condition in the above theorem, so we conclude that $f(x) = x$ is integral over $[0, b]$. Again, note that we were able to show this using only the specific partitions P_n, so we didn't have to worry about what happens with other random partitions.

So, the upper and lower integrals of f over $[0, b]$ agree. This common value is sandwiched between the upper sums $U(f, P_n)$ and lower sums $L(f, P_n)$:

$$L(f, P_n) = \frac{b^2(n-1)}{2n} \leq \int_0^b x\,dx \leq U(f, P_n) = \frac{b^2(n+1)}{2n},$$

so since these outer numbers both converge to $\frac{b^2}{2}$, the squeeze theorem implies that

$$\int_0^b x\,dx = \frac{b^2}{2},$$

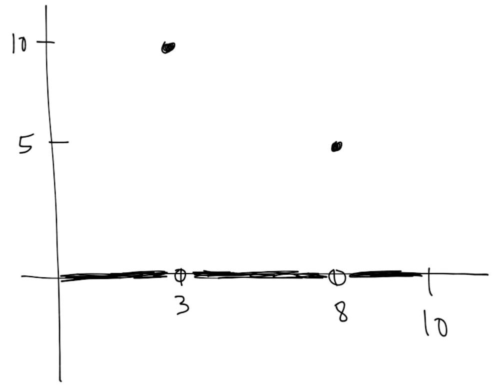
as expected. As mentioned before, finding this value explicitly by finding the exact value of the upper and lower integrals without using the alternate characterization of integrability given in the theorem would have been near impossible. (Essentially, any way in which this might work would end up reproving that theorem anyway.)

Example 1. Consider the function f on $[0, 1]$ which is 1 at each $x \in \mathbb{Q}$ and 0 at each $x \notin \mathbb{Q}$. We saw last time that this is not integrable, since all upper sums equal 1 and all lower sums equal 0. So, for any partition P of $[0, 1]$ we have

$$U(f, P) - L(f, P) = 1 - 0 = 1,$$

which cannot be made smaller than an arbitrary $\epsilon > 0$. Hence f fails the condition in the above theorem, which is another way to show that it is not integrable.

Example 2. Consider the function f on $[0, 10]$ which is zero everywhere, except at 3 and 8 where $f(3) = 10$ and $f(8) = 5$:



We claim that this function is integrable and that its integral over $[0, 10]$ is 0. This makes sense intuitively: the region "under" the graph of f consists of two vertical lines (above $x = 3$ and $x = 8$) and the "area" of these two vertical lines should indeed be 0.

Taking $\epsilon > 0$, we want to find a partition P of $[0, 10]$ such that

$$U(f, P) - L(f, P) < \epsilon.$$

This different in upper and lower sums looks like

$$U(f, P) - L(f, P) = \sum_{I_k} (\sup f - \inf f \text{ on } I_k)(\text{length of } I_k).$$

In this case, no matter what the subinterval I_k is, the infimum of f over it is zero, so the above becomes:

$$U(f, P) - L(f, P) = \sum_{I_k} (\sup f \text{ on } I_k)(\text{length of } I_k).$$

But on any I_k which does not contain 3 or 8, the supremum of f is also zero, so the above simplifies to just the sum over the intervals containing 3 and 8; say

$$U(f, P) - L(f, P) = (\sup f \text{ on } J)(\text{length of } J) + (\sup f \text{ on } K)(\text{length of } K)$$

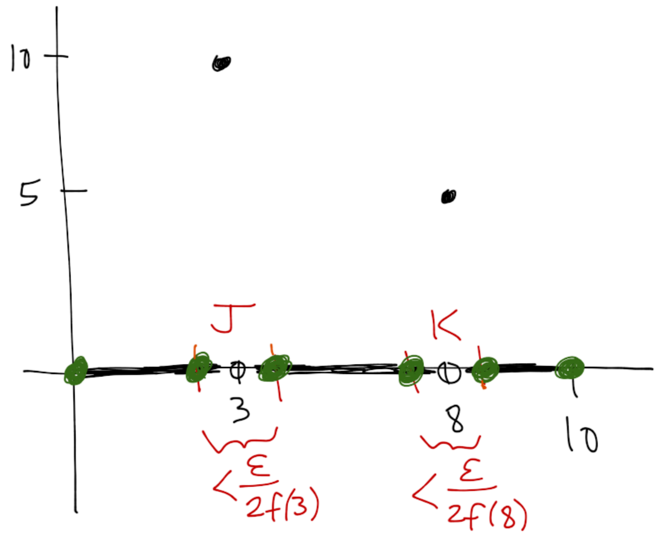
whee J is the subinterval containing 3 and K the subinterval containing 8. The supremum of f on J is $f(3) = 10$ and its supremum on K is $f(8) = 5$, so

$$U(f, P) - L(f, P) = 10(\text{length of } J) + 5(\text{length of } K).$$

This is the expression we want to make smaller than ϵ, and we can do so by constructing our partition P in such a way that the lengths of J and K are small enough that they balance out the values of $f(3)$ and $f(10)$! In particular, if the length of J was smaller than $\frac{\epsilon}{2 \cdot 10}$ and the length of K smaller than $\frac{\epsilon}{2 \cdot 5}$ (note the $\frac{\epsilon}{2}$-trick which is used here) then the above becomes

$$U(f, P) - L(f, P) = 10(\text{length of } J) + 5(\text{length of } K) < 10\left(\frac{\epsilon}{2 \cdot 10}\right) + 5\left(\frac{\epsilon}{2 \cdot 5}\right) = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

as required. Thus, given $\epsilon > 0$, picking an interval J around 3 of length smaller than $\frac{\epsilon}{20}$ and an interval K around 8 of length smaller than $\frac{\epsilon}{10}$:



gives a partition P (the points of which are the green points in the picture above, consisting of the endpoints of all subintervals used) such that $U(f,P) - L(f,P) < \epsilon$, so f is integrable on $[0,10]$. Since all lower sums are 0, the lower integral is 0, so

$$\int_0^{10} f(x)\,dx = 0$$

since, knowing already that f is integrable, the lower integral and upper integral must have the same value.

Remark. This idea of constructing a partition by making subintervals small enough to balance out values of the function is the crucial technique used in many integration problems and results; in particular, we'll use it next time to show that my favorite function is integrable and to show continuous functions are always integrable.

Lecture 21: Yet More on Integration

Today we continued with more examples dealing with integration, and in particular showed that my favorite function was integrable. (Yay!) We also proved that continuous functions are always integrable, which is why you never had to worry about integrability before in calculus: you pretty much only deal with continuous functions there, and so all integrals you ever had to compute automatically exist.

Warm-Up. Say that f is a function on $[a,b]$ such that $f(x) = 1$ for all but finitely many x; we claim that f is integrable. Denote the points where $f(x) \neq 1$ by $x_1, x_2, \ldots, x_n$. The idea is that given some partition, over any subintervals which contain none of the x_l we will have

$$\sup f - \inf = 0$$

since $f = 1$ is constant on those subintervals. Thus the expression for $U(f,P) - L(f,P)$ simplifies to one over only those subintervals which contain some x_k:

$$U(f,P) - L(f,P) = \sum_{I_k \text{ containing } x_k} (\sup f - \inf f)(\text{length } I_k).$$

On one of these subintervals I_k containing x_k, either

$$\sup f = 1 \text{ and } \inf f = f(x_k), \text{ or } \sup f = f(x_k) \text{ and } \inf f = 1$$

depending on whether $f(x_k) < 1$ or $f(x_k) > 1$. Regardless, either way on I_k containing x_k we have

$$(\sup f - \inf f) \leq |1 - f(x_k)|,$$

giving

$$U(f,P) - L(f,P) \leq \sum_k |1 - f(x_k)|(\text{length } I_k).$$

By constructing our partition so that these intervals I_k are small enough, we can make this above expression smaller than ϵ. Here is our proof.

Proof that f is integrable. Let $\epsilon > 0$ and denote the points where $f(x) \neq 1$ by $x_1, x_2, \dots, x_n$. For each $k = 1, \dots, n$, pick an interval I_k around x_k such that

$$\text{length } I_k < \frac{\epsilon}{n|1 - f(x_k)|}$$

and such that I_k contains no other point among $x_1, \dots, x_n$ apart from x_k. (The idea we outlined above assumes that each x_k is contained in only one subinterval, and we can guarantee this happens by making our intervals small enough.) Note that this denominator is nonzero since $f(x_k) \neq 1$.

Let P be the partition of $[a,b]$ consisting of a, b, and the endpoints of all intervals I_k constructed above. Over the subintervals determined by this partition which contain no x_k we have $\sup f = 1 = \inf f$, so $U(f,P) - L(f,P)$ reduces to

$$U(f,P) - L(f,P) = \sum_{k=1}^{n} (\sup f - \inf f \text{ on } I_k)(\text{length of } I_k).$$

But on I_k, $(\sup f - \inf f) \leq |1 - f(x_k)|$, so

$$\begin{aligned}
U(f,P) - L(f,P) &= \sum_{k=1}^{n} (\sup f - \inf f \text{ on } I_k)(\text{length of } I_k) \\
&\leq \sum_{k=1}^{n} |1 - f(x_k)|(\text{length of } I_k) \\
&< \sum_{k=1}^{n} |1 - f(x_k)| \frac{\epsilon}{n|1 - f(x_k)|} \\
&= \sum_{k=1}^{n} \frac{\epsilon}{n} \\
&= \epsilon.
\end{aligned}$$

Thus f is integrable on $[a,b]$ as claimed. □

Remark. Now that we know f is integrable, we can ask what the value of its integral actually is. Intuitively, the region under the graph of the constant function 1 over $[a,b]$ has area $b - a$, and the region under our f only differs from this by adding or subtracting a finite number of vertical line segments; since such line segments have "area" zero, this process should not change the area of the entire region, so we expect that

$$\int_a^b f(x)\,dx \text{ is still } b-a.$$

To show this precisely, we can argue as follows. First, no matter what partition P of $[a,b]$ we take, we have $\inf f \leq 1$ and $1 \leq \sup f$ on any subinterval, so

$$\sum_k (\inf f \text{ on } I_k)(\text{length } I_k) \leq \sum_k (\text{length } I_k) = b-a \leq \sum_k (\sup f \text{ on } I_k)(\text{length } I_k).$$

Thus $b-a$ is sandwiched between any lower sum and any upper sum, so $b-a$ is an upper bound for the set of all lower sums and a lower bound for the set of all upper sums. This implies that

$$(L)\int_a^b f(x)\,dx \leq b-a \leq (U)\int_a^b f(x)\,dx.$$

Since f is integrable, its lower and upper integrals agree, so we conclude that all of the above inequalities are actually equalities, so

$$\int_a^b f(x)\,dx = b-a$$

as claimed.

My favorite function. Recall that my favorite function of all time is the function $f:[0,1]\to\mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & \text{if } x \notin \mathbb{Q} \\ 1 & \text{if } x = 0 \\ \frac{1}{q} & \text{if } x = \frac{p}{q} \text{ where } p,q \in \mathbb{N} \text{ have no common factors} \end{cases}.$$

We show that f is integrable. This is tricky to prove, but once we know it is integrable the value of its integral is actually easy to compute since no matter what partition we take, any subinterval will contain an irrational, so $\inf f = 0$ on any subinterval and hence any lower sum is exactly zero; since the integral of f is equal to its lower integral, we will be able to conclude that

$$\int_0^1 f(x)\,dx = 0$$

once we've shown that f is actually integrable. (Try to think about whether or not it makes intuitive sense that the "area" under the graph of f should be zero; it does, if you think about it the right way.)

Note that given any partition P of $[0,1]$, $L(f,P)=0$ as stated above so

$$U(f,P)-L(f,P) = U(f,P) = \sum_k (\sup f \text{ on } I_k)(\text{length } I_k) \leq \sup_K (\text{length } I_k) = 1$$

since $\sup f \leq 1$ always; however, this is not going to help us if we want to make this expression smaller than ϵ. The idea we will use is the same one used to show that f is continuous at each irrational: given some $\epsilon > 0$, there are only finitely many rationals $r \in [0,1]$ satisfying $f(r) \geq \epsilon$. For all other rationals, we have $f(r) < \epsilon$, and this will give us a way to bound $\sup f$, at least over subintervals which contain none of the rationals where $f(r) \geq \epsilon$. Thus, we break up our entire sum into two pieces—the piece over the intervals J containing a rational where $f(r) \geq \epsilon$, and a piece over the intervals K containing no such rationals:

$$U(f, P) = \sum_J (\sup f)(\text{length}) + \sum_K (\sup f)(\text{length}).$$

Actually, based on this breaking up into two pieces, we guess that it might be helpful to use an $\frac{\epsilon}{2}$-trick, so we actually go back and replace the previous ϵ by $\frac{\epsilon}{2}$. That is, we consider the rationals where $f(r) \geq \frac{\epsilon}{2}$ and denote the intervals containing such a rational by J and the others by K. Over each K, $\sup f \leq \frac{\epsilon}{2}$ so the entire second piece above is bounded by

$$\sum_K (\sup f)(\text{length}) \leq \sum_K \frac{\epsilon}{2}(\text{length}) = \frac{\epsilon}{2} \sum_K (\text{length}) \leq \frac{\epsilon}{2}$$

since adding up all the lengths of the K intervals can't give more than the total length of $[0, 1]$. The goal is now to bound the first piece of $U(f, P)$ above also by $\frac{\epsilon}{2}$, giving us $U(f, P) < \epsilon$ in the end. But there are only finitely many rationals satisfying $f(r) \geq \frac{\epsilon}{2}$, so if we construct our partition to surround each of these rationals by a small enough interval, we can make the sum over the intervals K containing smaller than whatever we'd like. Here's our proof.

Proof that my favorite function is integrable. Let $\epsilon > 0$ and denote the finitely many rationals r such that $f(r) \geq \frac{\epsilon}{2}$ by $r_1, r_2, \ldots, r_n$. For each r_k, take an interval J_k around it such that

$$\text{length of } J_k < \frac{\epsilon}{2n},$$

and if need be make this intervas even smaller to guarantee that each J_k only contains one of the r_i's. (We implicitly assumed this in our scratch work above.) Take P to be the partition of $[0, 1]$ consisting of $0, 1$, and the endpoints of all the intervals J_k.

Then $L(f, P) = 0$ so

$$U(f, P) - L(f, P) = U(f, P) = \sum_{k=1}^{n} (\sup f \text{ on } J_k)(\text{length } J_k) + \sum_K (\sup f \text{ on } K)(\text{length } K)$$

where the second sum is over the subintervals K which contain none of $r_1, \ldots, r_n$. On each K, $\sup f \leq \frac{\epsilon}{2}$ while on each J_k, $\sup f \leq 1$. Thus

$$\begin{aligned} U(f, P) &\leq \sum_{k=1}^{n} (\text{length } J_k) + \sum_K \frac{\epsilon}{2}(\text{length } K) \\ &< \sum_{k=1}^{n} \frac{\epsilon}{2n} + \frac{\epsilon}{2} \sum_K (\text{length } K) \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon. \end{aligned}$$

Hence for this partition we have $U(f, P) - L(f, P) < \epsilon$, so we conclude that f is integrable on $[0, 1]$ as claimed. Since all lower sums are equal to 0, the value of its integral over $[0, 1]$ is zero. $\square$

Theorem. Any continuous function f on a closed interval $[a, b]$ is integrable. The book has a proof of this, which uses in an essential way the fact that any continuous function on $[a, b]$ is automatically uniformly continuous. This is perhaps the only "real" instance we will see where this fact really is crucial to justifying a fact which is taken for granted in calculus.

The book has a perfectly fine proof of this theorem, but I want to say a bit about the idea behind it. Given $\epsilon > 0$, we want a partition P of $[a, b]$ such that

$$U(f, P) - L(f, P) = \sum_k (\sup f - \inf f)(\text{length } I_k) < \epsilon.$$

Now, in order to bound $\sup f - \inf f$ over a subinterval I_k we want to able to bound expressions of the form

$$|f(x) - f(y)| \text{ for } x, y \in I_k.$$

Indeed, some bound M on this will give the bound $\sup f - \inf f \le M$. But luckily, such expressions are precisely the types of things which show up on the ϵ-δ definition of continuous, which says that $|f(x) - f(y)|$ can be bounded by whatever positive number we want as soon as x and y are close enough to each other... almost: actually, to get the bound we want we have to bound $|f(x) - f(y)|$ as *both* x and y vary, which is why we need to use the fact that f is uniformly continuous on $[a, b]$ and not just continuous.

Uniform continuity then says that for the positive number $\frac{\epsilon}{b-a}$ (why this? check the proof) there exists $\delta > 0$ such that

$$|x - y| < \delta \text{ implies } |f(x) - f(y)| < \frac{\epsilon}{b - a}.$$

Thus if we construct our partition so that each subinterval as length smaller than δ, then any points in this subinterval are close enough to guarantee that $|f(x) - f(y)| < \frac{\epsilon}{b-a}$ on that subinterval, giving

$$\sup f - \inf f < \frac{\epsilon}{b - a} \text{ on that subinterval.}$$

This then will imply that $U(f, P) - L(f, P) < \epsilon$. Again, check the book for the actual proof.

Lecture 22: Riemann Sums

Today we spoke about a different type of "Riemann sum" and how to phrase integrability in terms of these. Why do we do this? The point is that historically the definition of "integrability" in terms of these new sums was actually the first one discovered, and only later was this definition recast in terms of upper and lower sums. The definition in terms of these new Riemann sums is very hard to work with, so the development of upper and lower sums was pretty important; still, some properties of integrals are actually easier to justify using these new Riemann sums, so they're worth taking a look at.

Integral comparisons. Before the Warm-Up, we give one nice intuitive fact which we'll need. Suppose that f and g are both integrable with $g(x) \le f(x)$ for all $x \in [a, b]$. Then

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

This is completely obvious when interpreting these integrals in terms of areas: the graph of g lies below or at the same level as the graph of f, so the area under the graph of g should be $\le$ the area under the graph of f. This fact is also pretty easy to justify carefully: for any partition P,

$\sup g \leq \sup f$ on any subinterval so $U(g,P) \leq U(f,P)$. This implies that the upper integral of g is $\leq$ the upper integral of f, which is the stated claim.

Warm-Up. Say that f is a nonnegative function on $[a,b]$ such that

$$\int_a^b f(x)\,dx = 0.$$

Then at any point $x_0 \in [a,b]$ where f is continuous, we must have $f(x_0) = 0$. Note that it is certainly possible for a nonnegative and nonzero function to have zero integral—my favorite function is an example—but the claim is that such a function *must* be zero at any point where it is continuous. This is true for my favorite function, where f is continuous only at the irrationals and f is indeed zero at irrationals.

For a contradiction, suppose that $f(x_0) > 0$. The point is that then there is an interval around x_0 on which f is strictly positive, and the area under the graph of f over this interval will give a positive contribution to the integral of f overall. To be precise, since f is continuous at x_0, for the positive number $\frac{f(x_0)}{2}$ there exists $\delta > 0$ such that

$$|f(x) - f(x_0)| < \frac{f(x_0)}{2} \text{ for all } x \in (x_0 - \delta, x_0 + \delta),$$

which implies that $f(x) \geq \frac{f(x_0)}{2}$ on that interval. Define the piecewise constant function g by

$$g(x) = \begin{cases} \frac{f(x_0)}{2} & \text{if } x \in (x_0 - \delta, x_0 + \delta) \\ 0 & \text{otherwise.} \end{cases}$$

Then g is integrable (as all piecewise constant functions are), and $g(x) \leq f(x)$ for all $x \in [a,b]$. Thus

$$0 < 2\delta \frac{f(x_0)}{2} = \int_a^b g(x)\,dx \leq \int_a^b f(x)\,dx,$$

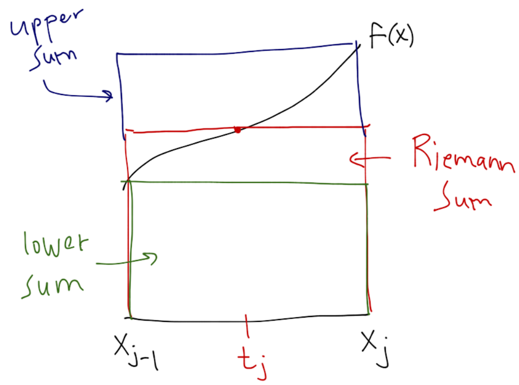
contradicting the fact that f should have integral equal to zero. (Hat tip to Charles Pugh, from whose book "Real Mathematical Analysis" this proof comes from.)

Riemann sums. Let f be a function on $[a,b]$, P be a partition of $[a,b]$, and $\{t_j\}$ be a collection of "sample points" where t_j is in the j-th subinterval determined by P. The *Riemann sum* associated to f, P, and t_j is

$$S(f,P,t_j) = \sum_{k=1}^{n} f(t_j)\Delta x_j,$$

where $\Delta x_j = x_j - x_{j-1}$ is the length of the j-th subinterval determined by P.

The product $f(t_j)\Delta x_j$ gives the area of a rectangle with base the j-th subinterval and height equal to the value $f(t_j)$. The point here is that, instead of taking a rectangle of height equal to $\inf f$ as we would for a lower sum or of height $\sup f$ as we would for an upper sum, we use the "sample points" t_j to determine the height:



Because of this, it is straightforward to see that lower, upper, and Riemann sums satisfy the inequalities:

$$L(f, P) \leq S(f, P, t_j) \leq U(f, P)$$

for any sample points t_j.

Riemann integrability. A function $f : [a, b] \to \mathbb{R}$ is said to be *Riemann integrable* with *integral* I if for any $\epsilon > 0$ there exists $\delta > 0$ such that for any partition P with $\|P\| < \delta$, we have

$$|S(f, P, t_j) - I| < \epsilon$$

for any sample points t_j. Here, $\|P\|$ is called the *norm* of P and is the length of the largest subinterval determined by P. Thus, this definition says that the difference between Riemann sums and the proposed value I of the integral can be made as small as we like once we take partitions made up by small enough subintervals. Suggestively, this definition can be phrased as the expression

$$\lim_{\|P\| \to 0} S(f, P, t_j) = I,$$

after we've made sense of what kind of limit we're taking here. (Take the definition we gave above as being the definition of what this limit means.)

Note that the book phrases this in a slightly different way. There the requirement is that for any $\epsilon > 0$ there exists a partition P_ϵ where the given inequality holds for all partitions $P \supseteq P_\epsilon$, that is for all partitions P which are refinements of P_ϵ. But this requirement is actually equivalent to the one I gave: if P_ϵ satisfies $\|P_\epsilon\| < \delta$, any refinement will still satisfy this inequality. My phrasing makes this definition more similar to other "ϵ-δ" definitions we've seen, which is why I like it better.

Theorem. A function is Riemann integrable in the above sense if and only if it is integrable in the previous sense we had in terms of upper and lower sums. The $\Leftarrow$ direction isn't so hard to justify, using the fact that the three types of sums we've seen satisfy

$$L(f, P) \leq S(f, P, t_j) \leq U(f, P).$$

However, the $\Rightarrow$ direction is not pretty, and requires coming up with a bunch of estimates and inequalities. Check the book for the full proof, but it's not something you'd be expected to know how to reproduce.

Remark. Compared with the earlier definition of integrability we had, this one is much harder to work with. For one thing, in order to prove that a function is integrable using this definition you would have to know what the value of its integral is going to end up being since this value shows up explicitly in the definition; this was not the case with the previous definition where all we needed was to find a partition satisfying $U(f,P) - L(f,P) < \epsilon$.

Another thing which makes this definition hard to work with is the fact that the inequality $|S(f,P,t_j) - I| < \epsilon$ should be true for *all* partitions with norm $< \delta$ and *all* sample points t_j: since there are infinitely many possible such partitions and infinitely many possible sample points, it seems unlikely you could say much about *all* of them.

The upshot is that this definition is not very useful for proving that a function is integrable, but it is in fact useful for saying something about the value of the integral once we know this integral exists. We'll see examples of this later; in particular, the Fundamental Theorem of Calculus is much easier to prove using this definition of Riemann integrability.

History lesson. A brief interlude. Now that we see both definitions of "integrable" are equivalent, it's fair to ask why we even introduce the second one at all, apart from some vague comments I made above that it will be useful for some purposes. The answer comes from going back and looking at how the notion of an integral historically came about.

"Integration" has been around a long time, and in fact ancient Greek mathematicians were already doing some types of computations which today we might refer to as integration. The big breakthrough of Newton and Leibniz in the 1600s was in putting together a single framework which encapsulated all such computations, and in proving the so-called Fundamental Theorem of Calculus which further described why such computations worked. However, mathematicians of that time did not think in completely rigorous terms, and indeed both Newton and Leibniz literally thought of dx in

$$\int_a^b f(x)\,dx$$

as an "infinitesimal length" and the above integral as "adding up" the areas of "infinitesimally small rectangles". This was all well and good for them, since they were able to do much with these vague notions, but it later (say 200 years) became apparent that there were problems with this approach. For one thing, there is no such thing physically as an "infinitesimal length", so why should we be allowed to work with such non-existent quantities? In addition, such vague and hand-wavy reasonings eventually led to certain paradoxes and conclusions which made no sense.

During the 1800s, mathematicians started to care more about rigor and giving precise definitions for all the mathematical objects they were using. In this vain, it was Riemann who first gave a precise definition of "integral", and his original definition was in fact the one given above in terms of this new type of Riemann sum! So, historically, this new definition actually came first. Riemann was able to put his definition to great use, reproving the Fundamental Theorem of Calculus and deducing all properties you would expect of integrals, now on a complete rigorous foundation.

But Riemann's definition was still hard for others to work with. Luckily, a few decades later a mathematician named *Darboux* came along and realized that you could simplify Riemann's definition by only considering sums were you use infimums and sums where you use supremums to give the heights of rectangles. In other words, the upper and lower sums we previously considered were actually Darboux's contribution to integration theory! He then showed that the definition of integrable using these types of sums was equivalent to Riemann's original definition. This was also important for the development of other types of "integrals" later on; for instance, the definition of the so-called *Lesbesgue integral* more closely mimics Darboux's definition than Riemann's.

In any other book I've ever seen, what our book calls the "lower Riemann sum" and "upper Riemann sum" of a function are referred to as the "lower Darboux sum" and the "upper Darboux sum". Then, the definition of "integrable" in terms of these is what it means to say that a function is "Darboux integrable", which in fact is equivalent to Riemann integrable. I think it's a shame that our book doesn't point out the contributions of Darboux, which really were crucial to the development of integration theory as a whole. But, c'est la vie. We now return to our regularly scheduled lecture.

(Incidentally, the only place our book mentions the name "Darboux" is in reference to Theorem 4.23, which says that derivatives have the intermediate value property. Indeed, functions in general which have the intermediate value property are usually referred to as *Darboux functions*; so, all continuous functions and all derivatives are Darboux in this sense.)

Integrable implies bounded. Note that in the definition of Riemann integrable given above, no reference is made to the function f being bounded. Indeed, this is not needed in order to form the Riemann sums we're looking at, since these sums now use the values of f at sample points in their definition; for upper and lower sums, we definitely needed f to be bounded in order for $\inf f$ and $\sup f$ to exist. But, we can now show that we did not lose anything previously by assuming our functions were bounded, since it is a fact that f being Riemann integrable in this new sense implies that f must be bounded. (This proof is also due to Charles Pugh.)

Proof. Suppose that f is Riemann integrable on $[a, b]$ with integral I, but that it is not bounded. Then there exists a partition P and sample points t_j such that

$$|S(f, P, t_j) - I| < 1,$$

simply by applying the definition of Riemann integrable to the positive number $\epsilon = 1$. Since f is unbounded on $[a, b]$, it must be unbounded on some subinterval $[x_{k-1}, x_k]$ determined by P. Hence there exists another sample point t'_k in this subinterval such that

$$|f(t'_k) - f(t_k)| > \frac{3}{|x_k - x_{k-1}|},$$

which comes from taking $f(t'_k)$ large enough. Define a new collection of sample points t'_j where we keep every sample point we originally had except for replacing t_k by t'_k.

According to the definition of Riemann integrable, this change in sample points should not change the fact that

$$|S(f, P, t'_j) - I| < 1.$$

However, comparing the Riemann sum we get for these new sample points to the old one we have

$$|S(f, P, t'_j) - S(f, P, t_j)| = |f(t'_k) - f(t_k)||x_k - x_{k-1}| > 3$$

since the only nonzero contribution to this difference comes from the subinterval $[x_{k-1}, x_k]$ because we kept all other sample points the same. This is a contradiction, since we cannot have $S(f, P, t'_j)$ and $S(f, P, t_j)$ both be within 1 away from I and yet be more than 3 away from each other. Thus f must have been bounded. □

Other properties of integrals. The book has tons of other properties of integrals listed in Section 5.2 which you should know, such as the fact that

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

or the fact that the product of integrable functions is integrable. We won't cover these explicitly in class, but you should definitely be familiar with them. We'll take them for granted from now on.

Lecture 23: The Fundamental Theorem of Calculus

Today we spoke about the Fundamental Theorem(s) of Calculus, which justifies the integration techniques you've been using all your lives. The term "fundamental" really is appropriate, since we've seen by now that integration and differentiation are completely different things, and yet in nice situations there is a direct relation between them. Still, it is important to understand that these theorems only apply under certain assumptions, which are not always necessarily satisfied.

Note that the book does not separate the Fundamental Theorems of Calculus into a "first" one and a "second" one as many other sources do, but instead lists them as two parts of the same theorem. Here we list them separately.

Warm-Up. Suppose that f is integrable on $[a, b]$ with integral equal to I. Let P_n be any sequence of partitions of $[a, b]$ such that $\|P_n\| \to 0$ and take any collection of sample points $t_{n,j}$ for each of these. We claim that the sequence of Riemann sums $S(f, P_n, t_{n,j})$ converges to I.

Let $\epsilon > 0$. Since f is Riemann integrable on $[a, b]$ there exists $\delta > 0$ such that for any partition P with $\|P\| < \delta$, we have

$$|S(f, P, t_j) - I| < \epsilon$$

for any sample points t_j. For our sequence of partitions, since $\|P_n\| \to 0$ there exists $N \in \mathbb{N}$ such that

$$\|P_n\| < \delta \text{ for } n \geq N.$$

Thus for the sample points $t_{n,j}$ we're looking at, we have

$$|S(f, P_n, t_{n,j}) - I| < \epsilon \text{ for } n \geq N,$$

so the sequence $(S(f, P_n, t_{n,j}))$ converges to I as claimed.

Remark. Note that if we didn't assume f is integrable beforehand, knowing that

$$S(f, P_n, t_{n,j}) \to I$$

alone is not enough to prove that f is Riemann integrable, since here we are only considering some particular partitions P_n with some specific sample points $t_{n,j}$, whereas the definition we saw last time of Riemann integrability requires knowing something about *all* partitions of small enough norm and *all* sample points. The point here is that, once we know that f is indeed integrable, we can in fact use a specific sequence of partitions and sample points to actually find the value of the integral. This is something that I alluded to last time: Riemann sums don't really help us when proving that a function is integrable, but they can be useful in actually computing integrals.

Application. We use the above result to compute the value of $\int_a^b x^2\,dx$, which we know from calculus should be $\frac{b^3-a^3}{3}$. Since $f(x) = x^2$ is continuous on $[a, b]$, it is integrable so by the Warm-Up we can compute the value of its integral by taking a sequence of partitions with norm converging to zero, taking any specific sample points we want, and finding the limit of the Riemann sums for these. We use the sequence P_n where P_n is the evenly-spaced partition:

$$x_0 = a < a + \frac{b-a}{n} < a + \frac{2(b-a)}{n} < \cdots < a + \frac{(n-1)(b-a)}{n} < b = x_n.$$

So, each subinterval of P_n has length $\frac{b-a}{n}$ and hence $\|P_n\| \to 0$ as required. For sample points we use the right-endpoints of each subinterval, so

$$t_{n,k} = a + \frac{k(b-a)}{n}.$$

We compute:

$$\begin{aligned}
S(f,P_n,t_{n,k}) &= \sum_{k=1}^{n} f(t_k)\Delta x_k \\
&= \sum_{k=1}^{n}\left(a+\frac{k(b-a)}{n}\right)^2\left(\frac{b-a}{n}\right) \\
&= \sum_{k=1}^{n}\left(a^2+\frac{2ak(b-a)}{n}+\frac{k^2(b-a)^2}{n^2}\right)\left(\frac{b-a}{n}\right) \\
&= \frac{a^2(b-a)}{n}\sum_{k=1}^{n}1+\frac{2a(b-a)^2}{n^2}\sum_{k=1}^{n}k+\frac{(b-a)^3}{n^3}\sum_{k=1}^{n}k^2.
\end{aligned}$$

The first sum is n, the second is

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

and the third is

$$1^2+2^2+3^2+\cdots+n^2=\frac{n(n+1)(2n+1)}{6}.$$

Thus

$$S(f,P_n,t_{n,k})=a^2(b-a)+\frac{a(b-a)^2(n+1)}{n}+\frac{(b-a)^3(n+1)(2n+1)}{6n^2}.$$

As $n\to\infty$ this converges to

$$a^2(b-a)+a(b-a)^2+\frac{(b-a)^3}{3}=\frac{b^3-a^3}{3}.$$

Hence since $\|P_n\|\to 0$ and $S(f,P_n,t_{n,k})\to\frac{b^3-a^3}{3}$, the Warm-Up implies that $\int_a^b x^2\,dx=\frac{b^3-a^3}{3}$.

First Fundamental Theorem of Calculus. Suppose that f is differentiable and that f' is integrable on $[a,b]$. Then

$$\int_a^b f'(x)\,dx=f(b)-f(a).$$

The book has a proof, but I'll give a different proof using the Warm-Up result.

Proof. Take any partition P of $[a,b]$. Note that (recall $x_n=b$ and $x_0=a$)

$$\begin{aligned}
f(b)-f(a) &= [f(x_n)-f(x_{n-1})]+[f(x_{n-1})-f(x_{n-2})]+\cdots+[f(x_2)-f(x_1)]+[f(x_1)-f(x_0)] \\
&= \sum_{k=1}^{n}[f(x_k)-f(x_{k-1})].
\end{aligned}$$

By the Mean Value Theorem, for each k there exists $t_k\in(x_{k-1},x_k)$ such that

$$f(x_k)-f(x_{k-1})=f'(t_k)(x_k-x_{k-1}).$$

Substituting this above gives

$$f(b)-f(a)=\sum_{k=1}^{n}f'(t_k)\Delta x_k=S(f',P,t_k).$$

Now, apply this to a sequence of partitions P_n of $[a, b]$ with $\|P_n\| \to 0$. By the Warm-Up we have that $S(f', P_n, t_{n,k})$ converges to $\int_a^b f'(x)\,dx$. But each of these Riemann sums equals $f(b) - f(a)$, so this sequence also converges to $f(b) - f(a)$. Hence

$$\int_a^b f'(x)\,dx = f(b) - f(a)$$

as claimed. □

Anti-differentiation and integration are different. Note that the assumption that f' is integrable is crucial, and highlights an important point: in general, the process of "integration" and "anti-differentiation" are quite different. In particular, there are examples of functions which have an antiderivative but are not integrable, contrary to whatever a previous calculus course may have led you to believe. For instance, let f be the function

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

This is differentiable and its derivative is

$$f'(x) = \begin{cases} 2x \sin \frac{1}{x^2} - \frac{2}{x} \cos \frac{1}{x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Due to the $\frac{2}{x}$ term, f' is unbounded on $[-1, 1]$ and so is not integrable on $[-1, 1]$. However, f' does have an antiderivative, namely f.

Similarly, there are examples of functions which are integrable but which do not have an antiderivative. Take the function g which is 0 for $x \leq 0$ and 1 for $x > 0$. This is integrable, but there can be no function G with $G' = g$ since this would imply (based on a fact we saw about derivatives a while ago) that $G' = g$ had no jump discontinuities, which it does.

Remark. Let me emphasize the chain of results which have now led us to this: the Fundamental Theorem of Calculus depends on the Mean Value Theorem, which depends on the Extreme Value Theorem, which depends on the Bolzano-Weierstrass Theorem. I mentioned way back when that the Bolzano-Weierstrass Theorem—the fact that bounded sequences always have convergent subsequences—was going to be crucial to developing calculus, and now hopefully you can see why.

Integrals are always continuous. Suppose that f is integrable on $[a, b]$ and define the function F on $[a, b]$ by

$$F(x) = \int_a^x f(t)\,dt.$$

To be clear, the value of F at $x \in [a, b]$ is the integral of f up to x itself. Then F is continuous on $[a, b]$. This is a basic property of integrals; the book has a proof, but here we give a slightly different one.

Proof. Since f is integrable on $[a, b]$ it is bounded on this interval; let M be a bound. Then for any $x, x_0 \in [a, b]$ we have

$$|F(x) - F(x_0)| = \left| \int_a^x f(t)\,dt - \int_a^{x_0} f(t)\,dt \right|$$

$$
\begin{aligned}
&= \left| \int_{x_0}^{x} f(t)\, dt \right| \\
&\le \int_{x_0}^{x} |f(t)|\, dt \\
&\le \int_{x_0}^{x} M\, dt \\
&= M|x - x_0|.
\end{aligned}
$$

Thus F is Lipschitz, and a previous homework problem shows that F is hence uniformly continuous, and so continuous on $[a, b]$ as well. □

Second Fundamental Theorem of Calculus. Suppose that f is continuous on $[a, b]$, and define the function F on $[a, b]$ by

$$F(x) = \int_a^x f(t)\, dt.$$

Then F is differentiable and $F'(x) = f(x)$. (In fact, even without f being continuous on all of $[a, b]$, F will be differentiable at any point at which f *is* continuous.) The proof of this is in the book, which you should read, but here's the basic idea.

We want to show that

$$\lim_{x \to x_0} \frac{F(x) - F(x_0)}{x - x_0} \text{ exists and equals } f(x),$$

which is the same as showing that

$$\lim_{x \to x_0} \frac{F(x) - F(x_0) - f(x)(x - x_0)}{x - x_0} = 0.$$

Now, we can rewrite the numerator of this fraction as

$$F(x) - F(x_0) - f(x)(x - x_0) = \int_a^x f(t)\, dt - \int_a^{x_0} f(t)\, dt - \int_{x_0}^x f(x)\, dt = \int_{x_0}^x f(t)\, dt - \int_{x_0}^x f(x)\, dt$$

where we use the fact that $\int_{x_0}^x f(x)\, dt = f(x) \int_{x_0}^x dt = f(x)(x - x_0)$ since $f(x)$ is constant with respect to t. Thus

$$F(x) - F(x_0) - f(x)(x - x_0) = \int_{x_0}^x [f(t) - f(x)]\, dt$$

and so

$$\frac{F(x) - F(x_0) - f(x)(x - x_0)}{x - x_0} = \frac{1}{x - x_0} \int_{x_0}^x [f(t) - f(x)]\, dt.$$

Using the fact that f is continuous (in particular, the fact that we can make $f(t) - f(x)$ as small as we want by taking x close enough to x_0), you can now show that the limit as $x \to x_0$ of the right-hand side is indeed 0. Again, check the book for details.

Remark. The First Fundamental Theorem of Calculus is the one which loosely says "the integral of the derivative is the original function" and the Second Fundamental Theorem of Calculus is the one which loosely says "the derivative of the integral is the original function". Taken together they say that "differentiation" and "integration" are in some sense inverse operations to each other, but again it is important to realize that this only holds under certain assumptions: namely that f' is integrable in the First Fundamental Theorem and that f is continuous in the Second Fundamental Theorem. For functions where these assumptions don't hold, integration and differentiation really have nothing to do with one another.

Lecture 24: More on the Fundamental Theorem of Calculus

Today we continued talking about the Fundamental Theorem of Calculus, looking at some uses and consequences. This is the end of possible material for the final.

Geometric intuition behind the 2nd Fundamental Theorem. Before moving on, note that there is a nice picture one can draw which illustrates why the second Fundamental Theorem of Calculus should be true. Recall the statement: if f is continuous on $[a, b]$, then

$$F(x) = \int_a^x f(t)\,dt$$

is differentiable and $F'(x) = f(x)$, which comes down to showing

$$\lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = f(x).$$

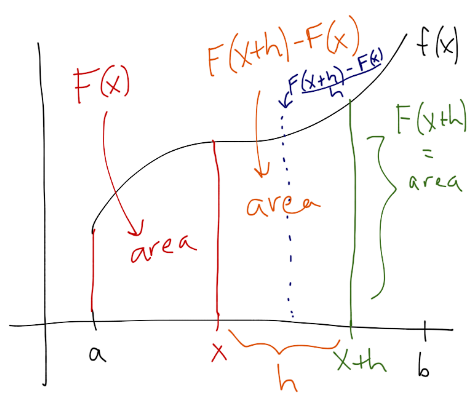
Now,

$$F(x+h) - F(x) = \int_x^{x+h} f(t)\,dt$$

represents the area under the graph of f between x and $x + h$. This region has length h, so

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

is measuring the "height" of this region (since area = length × height). (Of course, this height is not necessarily constant, but we're only trying to get some intuition here after all so that's not a big deal.)



The claim is that as the endpoint $x + h$ approaches the point x, this "height" should approach the "height" at x itself, which is $f(x)$.

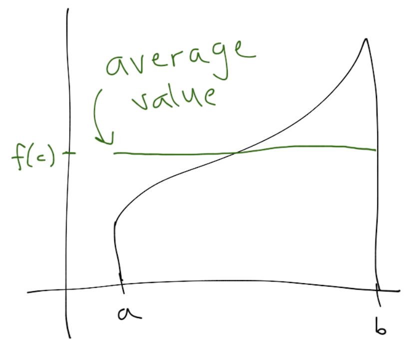
Warm-Up. Suppose that f is continuous on $[a, b]$. We claim that there exists $c \in (a, b)$ such that

$$f(c) = \frac{1}{b-a} \int_a^b f(t)\,dt.$$

The expression on the right is called the *average value* of f over $[a, b]$ since it is somewhat analogous to a usual average: the integral "adds" up all values of $f(t)$ and there are "$b - a$" many such things we are adding up. The claim is that this average value is actually attained at some point c. Rewriting the above expression as

$$f(c)(b-a) = \int_a^b f(t)\,dt$$

says that the area under f is equal to the area of some rectangle with base $[a, b]$ and height $f(c)$, which is another way to think about why this is an "average" value.



Define F on $[a, b]$ by

$$F(x) = \int_a^x f(t)\,dt.$$

Since f is continuous, the Second Fundamental Theorem of Calculus F is differentiable and $F'(x) = f(x)$ for all x. By the Mean Value Theorem, there exists $c \in (a, b)$ such that

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

We have

$$F(b) = \int_a^b f(t)\,dt \text{ and } F(a) = \int_a^a f(t)\,dt = 0,$$

so

$$\int_a^b f(t)\,dt = f(c)(b-a)$$

as claimed.

(This result is also often called the *Mean Value Theorem for Integrals*, or is at least one version of the mean value theorem for integrals.)

Example 1. Define the function F by

$$F(x) = \int_0^{x^2} e^{t^2}\,dt.$$

If the upper bound on the integral were simply x, the second Fundamental Theorem would give $F'(x) = e^{x^2}$. With the upper bound being x^2, we just need a chain rule application:

$$F'(x) = 2xe^{x^4}.$$

To be clear, the $2x$ comes from the derivative of the upper bound x^2, and the e^{x^4} comes from differentiation

$$\int_0^u e^{t^2}\,dt$$

with respect to u and then substituting $u = x^2$.

Similarly, the derivative of

$$\int_0^{x\sin x} e^{t^2}\,dt$$

with respect to x is

$$(\sin x + x\cos x)e^{x^2\sin^2 x}.$$

Example 2. We want to compute the derivative of

$$F(x) = \int_{x^2}^{x^3} \sin(t^2)\,dt$$

with respect to x. Before we apply the chain rule as before, we need to rewrite our expression as

$$F(x) = \int_{x^2}^{0} \sin(t^2)\,dt + \int_0^{x^3} \sin(t^2)\,dt = -\int_0^{x^2} \sin(t^2)\,dt + \int_0^{x^3} \sin(t^2)\,dt$$

since it is these types of expressions to which the second Fundamental Theorem applies. Hence

$$F'(x) = -2x\sin(x^4) + 3x^2\sin(x^4).$$

Integration gives "nicer" functions. The second Fundamental Theorem of Calculus also implies that each time we integrate a function, the resulting function has "nicer" properties. For instance, the function $|x|$ is continuous but not differentiable at 0, and its integral:

$$\int_{-1}^{x} |t|\,dt$$

is now continuously differentiable everywhere. Integrating this:

$$\int_{-1}^{y}\left(\int_{-1}^{x} |t|\,dt\right)\,dx,$$

where now y is the variable, gives a function which is now continuously twice-differentiable. And so on, each integration gives a function which has a higher order of continuous differentiability.

Similarly, my favorite function is not continuous on all of $[0, 1]$, but it is integrable. Hence

$$\int_0^x (\text{my favorite function})\,dt$$

is continuous on $[0, 1]$. Integrating again gives:

$$\int_0^y \left(\int_0^x (\text{my favorite function})\,dt\right)\,dx$$

which is now differentiable with respect to y. Integrating once more gives:

$$\int_0^z \left(\int_0^y \left(\int_0^x (\text{my favorite function})\,dt\right)\,dx\right)\,dy$$

which is twice-differentiable with respect to z, and so on.

Final integration techniques. The Fundamental Theorem of Calculus can be used to justify some final integration techniques you know from calculus: integration by parts and the method of substitution in the case where your integrand is continuous. Check the book for these.

Substitution also applies when the integrand is only integrable and not necessarily continuous, but as you can see in the book, the proof in this general setting is pretty tough. Still, I want to point out one thing. The statement is the following. Suppose ϕ is continuously differentiable on $[a,b]$ with $\phi'(x) \neq 0$ for all x and that f is integrable on $[c,d] = \phi([a,b])$, which is the image of $[a,b]$ under ϕ. Then

$$\int_c^d f(x)\,dx = \int_a^b f(\phi(t))|\phi'(t)|\,dt.$$

Intuitively, this just comes from making the substitution $x = \phi(t)$. However, this would suggest that $dx = \phi'(t)\,dt$, so the question is: where did the absolute value come from? This is something which is usually glossed over in a calculus course, but is actually pretty important. After all, if you remember the change of variables formula for multivariable integrals, there you definitely end up needing to take the absolute value of some kind of Jacobian determinant, so why should the single-variable case be any different? Let's walk through a basic example.

Consider $\int_0^1 f(x)\,dx$. If we make the substitution $x = \phi(t) = -t^3$, then $dx = \phi'(t) = -3t^2\,dt$ and this integral becomes

$$\int_0^1 f(x)\,dx = \int_0^{-1} f(-t^3)(-3t^2)\,dt.$$

However, we usually write integrals with left endpoints as the lower bound and right endpoints as upper bounds, so here we have it backwards. To get it into the "right" form we have to write this as

$$\int_0^1 f(x)\,dx = -\int_{-1}^0 f(-t^3)(-3t^2)\,dt = \int_{-1}^0 f(-t^3)(3t^2)\,dt.$$

But note that $3t^2 = |-3t^2| = |\phi'(t)|$, so the absolute value actually *is* there. The reason why need an absolute value in general after making a substitution is to correct the fact that the change of variables we use might "flip" our interval around!

Defining logarithms. Finally, we use what we've built up to define $\log x$, and to prove that it has some properties we would expect. The point is that we can do this without making any reference to the fact that $\log x$ should be the inverse of e^x; indeed, many authors take the definition we give for $\log x$ as a given, and then use this to define e^x as its inverse function.

We define $\log x$ for $x > 0$ by

$$\log x = \int_1^x \frac{1}{t}\,dt.$$

Note that $\log 1 = 0$. Since $\frac{1}{t}$ is continuous on $(0,\infty)$, by the second Fundamental Theorem of Calculus $\log x$ is differentiable on $(0,\infty)$ and

$$\frac{d}{dx}(\log x) = \frac{1}{x}.$$

Since this derivative is always positive, $\log x$ is strictly increasing. We can also see this using the fact that for $x < y$,

$$\log y - \log x = \int_x^y \frac{1}{t}\,dt > 0$$

since we are integrating a positive function.

Now, fix $y > 0$ and consider the function

$$g(x) = \log(xy) - \log x - \log y.$$

Differentiating with respect to x gives

$$g'(x) = \frac{1}{xy}(y) - \frac{1}{x} = 0 \text{ for all } x > 0,$$

so g is constant. Since $g(1) = \log x - \log x - \log 1 = 0$, this constant is 0 so

$$\log(xy) - \log x - \log y = 0 \text{ for all } x > 0.$$

Since $y > 0$ was arbitrary, we conclude that

$$\log(xy) = \log x + \log y$$

for all $x, y > 0$, which is another well-known property of logarithms. Continuing, you can justify all properties of $\log x$ you would like using only its definition as the integral of $\frac{1}{t}$.

Lecture 25: Measure Theory and the Lebesgue Integral

Today we did not have class, but I promised anyway to give a brief introduction to measure theory and the Lebesgue integral, simply to pique your interest in these topics. If nothing else, they're important topics in the historical development of analysis, since many further applications (such as advanced probability and statistics theory) depend on the fact that there are more general types of integrals than the Riemann integral we have been studying. Enjoy.

Flaws of Riemann integration. The integral we've been studying, the *Riemann integral*, was historically the first rigorous integral introduced, and gave a precise meaning to the "area" under the graph of a function. However, consider the following example. Let $f : [0,1] \to \mathbb{R}$ be the function

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases}$$

This is usually called the *characteristic function* of $\mathbb{Q}$ on $[0,1]$ since it "picks out" the elements of $\mathbb{Q}$ by assigning 1 to them and 0 to everything else. We've seen that this function is not Riemann integrable over $[0,1]$, since for instance all upper sums are 1 and all lower sums are 0. Hence $\int_0^1 f(x)\,dx$ is not defined.

However, I claim that this function *does* have a well-defined area under its graph and that this "area" should be zero. This is bad, since up until now we've said that the Riemann integral should give the area under the graph of a function, and yet in this case this area exists without the corresponding integral existing. The point is that this example and others point out that the Riemann integral is not "strong enough" to handle all possible types of areas one would want to compute. Essentially, the function f above is too badly behaved for its Riemann integral to exist, yet not so badly behaved that the area under its graph doesn't. We need a better type of integral if we want to be able to handle more general types of areas.

Why should the region under the graph of f have a well-defined area and why should it be zero? Take any $\epsilon > 0$. Since $\mathbb{Q}$ is countable, we can list the rationals in $[0,1]$ in some order, say

$$r_1,\ r_2,\ r_3,\ r_4, \ldots.$$

Take an interval I_1 around r_1 of length $\frac{\epsilon}{2}$, an interval I_2 around r_2 of length $\frac{\epsilon}{2^2}$, and in general an interval I_n around r_n of length $\frac{\epsilon}{2^n}$. For the interval I_n, take a rectangle R_n with base I_n and of height 1; then the area of R_n is simply the length of I_n. The rectangles $\{R_n\}$ all together complete cover the region under the graph of f, so the "area" of this region should be $\leq$ the total area enclosed by these rectangles. This total area is at most:

$$\sum_{n=1}^{\infty}(\text{area of } R_n) = \sum_{n=1}^{\infty}(\text{length of } I_n) = \sum_{n=1}^{\infty}\frac{\epsilon}{2^n} = \epsilon\sum_{n=1}^{\infty}\frac{1}{2^n} = \epsilon,$$

so we find that for any $\epsilon > 0$ the area of the region under the graph of f should be $\leq \epsilon$. Thus this area should be exactly zero! Intuitively, we are enclosing the region under the graph of f by rectangles of height 1 which can be made arbitrarily thin, so this region should have a well-defined area of zero.

Lebesgue to the rescue. So, the Riemann integral is not strong enough to capture all types of areas we need to be able to compute. How do we fix this? The answer is actually pretty simple: in the definition of the Riemann integral, we used areas of rectangles whose bases were intervals, so what we do is simply replace these by a more general type of "rectangle" where the base no longer has to be an interval, but rather can be any subset of $\mathbb{R}$ with a well-defined length. To be clear, these geometric figures are no longer "rectangles" in the usual sense, but rather something with a fixed height but a possibly strange looking base. If we consider such a "rectangle" of fixed height and base given by some $A \subseteq \mathbb{R}$, then the "area" of this "rectangle" is still

$$(\text{length of base})(\text{height}).$$

Geometrically, this "rectangle" is what you get when you take vertical line segments of length 1 above each point of A and put them all together. Note that this new type of "rectangle" still includes the old type when we take A to be an interval, which is what will imply that the new type of "integral" we get really is a generalization of the Riemann integral.

To make all this work, we need to have a notion of "length" for arbitrary subsets of $\mathbb{R}$ and not just for intervals. This is where "measure theory" comes into play.

Definitions. Let $A \subseteq \mathbb{R}$. We say that a countable collection of open intervals $\{(a_i, b_i)\}$ *covers* A if A is contained in their union:

$$A \subseteq \bigcup_i (a_i, b_i).$$

The *Lebesgue measure* of A, denoted by $\mu(A)$, is:

$$\mu(A) = \inf\left\{\sum_i (b_i - a_i) \;\middle|\; \{(a_i, b_i)\} \text{ covers } A\right\}.$$

Some remarks are in order. The (possibly infinite) sum $\sum_i (b_i - a_i)$ is adding up the lengths of all intervals in the given collection, and is called the *total length* of the collection. For any such collection which covers A, intuitively the "length" of A should be $\leq$ the total length of the covering, and the "length" of A is then defined to be the infimum of all such possible total lengths.

Actually, to be precise what we've defined is really the *Lebesgue outer measure* of A since we are measuring the length of A from the "outside" by considering intervals which cover A. Analogously, using intervals "inside" A (so considering collections whose union is contained in A) and taking the supremum of their total lengths gives the *Lebesgue inner measure* of A. These two "measures" can be different, which is related to what it means for A to be a "measurable" subset of $\mathbb{R}$; we'll skip any discussion of this here however and focus only on the outer measure. (In truth, what follows only works for subsets where the outer measure and inner measure are the same. So, the Lebesgue integral itself has some drawbacks!)

Lebesgue measure of intervals. If we are saying that Lebesgue measure is a sort of general notion of "length" it had better be the case that sets which already have a well-defined length have a Lebesgue measure which agrees with that length. This is true: the Lebesgue measure of each of $(a,b),[a,b),(a,b]$, and $[a,b]$ is $b-a$. For disjoint intervals (a,b) and (c,d), the union $(a,b)\cup(c,d)$ has Lebesgue measure $(b-a)+(d-c)$. And so on.

Step functions. So, the notion of "Lebesgue measure" generalizes the notion of length to now include arbitrary subsets of $\mathbb{R}$. Using this we now start building up a new type of integral.

First, for a subset $A\subseteq\mathbb{R}$ we let f_A denote its *characteristic function*; that is, the function which is 1 at points of A and 0 elsewhere. The integral of f_A should be the area of the region under its graph, but this region consists of vertical line segments of height 1 above each point of A. Hence this region is a "rectangle" of height 1 and base A, so intuitively the area of this region should be

$$(\text{length of base})(\text{height}) = (\mu(A))(1) = \mu(A).$$

We define the *Lebesgue integral* of f_A to be this value:

$$\int f_A\,d\mu := \mu(A).$$

The notation $d\mu$ indicates that we are integrating with respect to Lebesgue measure.

Next we consider functions which are "linear combinations" of characteristic functions. To be clear, for a finite collection of characteristic functions $f_{A_1},\dots,f_{A_n}$ and scalars $c_1,\dots,c_n\in\mathbb{R}$, the function

$$s = c_1 f_{A_1}+\cdots+c_n f_{A_n}$$

is what is known as a "step function". Graphically, the graph of f looks like a collection of "steps" where the n-th step occurs over A_n and is of height c_n. This is analogous to a piecewise constant function like

$$g(x)=\begin{cases}1 & \text{if } x\in[0,1]\\ 2 & \text{if } x\in(1,2]\\ 3 & \text{if } x\in(2,3]\end{cases}$$

only now we allow the "pieces" to occur not only over intervals but over arbitrary subsets of $\mathbb{R}$. The area of the region under the graph of f should simply be the sum of the areas of the regions under each individual piece (as it is for a piecewise constant function), so we define:

$$\int s\,d\mu := c_1\int f_{A_1}\,d\mu+\cdots+c_n\int f_{A_n}\,d\mu = c_1\mu(A_1)+\cdots+c_n\mu(A_n).$$

But now notice something else. For a piecewise constant function, the expression you get out of this integral is precisely the type of thing you get with a Riemann sum since all we're doing is adding up areas of various "rectangles". So, we should view this expression above as analogous to a Riemann sum, except again that we are allowing a more general type of rectangle. To be precise, say we view these expressions above as being analogous to the lower sums we've been considering. Then the integral of our function is obtained as the supremum of these lower sums, and so the

"Lebesgue integral" of any function should now be obtained as the supremum of this new type of "lower sum". This leads to our final definition.

Definition. Let $f : S \to \mathbb{R}$ be a function defined on some $S \subseteq \mathbb{R}$. The *Lebesgue integral* of f over S is defined to be:

$$\int_S f\,du = \sup \left\{ \int s\,du \;\middle|\; s \text{ is a step function such that } s \leq f \right\}.$$

That is, we consider all step functions "below" f, compute their Lebesgue integrals, and take the supremum of these values. Intuitively, we are computing the area under the graph of f by approximating this area from below using a more general type of "rectangle", which is directly analogous to taking the supremum of all lower sums in the construction of the Riemann integral. Such a supremum does not always exist, but when it does we say that f is *Lebesgue integrable.*

So, the upshot is that the Lebesgue integral is defined a very similar manner as the Riemann integral, only using a more notion of "rectangle", which we can do after we've defined a more general notion of "length".

Lebesgue Integration is better than Riemann Integration. First and foremost, we have the following fact: any function which is Riemann integrable is Lebesgue integrable and its Lebesgue integral and Riemann integral agree. Thus, we do not "lose" any integrals when we move from Riemann to Lebesgue integration. This result is essentially due to the previously mentioned fact that usual rectangles are counted among the more general type of "rectangle" we now consider.

But, the point is that we can now integrate functions which were not integrable (in the Riemann sense) before! Indeed, the characteristic function of $\mathbb{Q}$ over $[0, 1]$ is Lebesgue integrable and its Lebesgue integral is zero, as we said should be the case earlier when pointing out flaws of the Riemann integral. There are tons of other examples of functions which are not Riemann integrable but now are Lebesgue integrable, showing that Lebesgue integration really is a stronger theory of integration than Riemann integration is.

So What? After all this, it is fair to ask why we care about developing a more general type of integral. Consider the following. In practice, when dealing with a large numbers of "data points" it is not very likely that this data will fit together to form a "nice" geometric shape, such as an interval in 1-dimension or a rectangle in two. More often, these data points will be some seemingly random looking subset of $\mathbb{R}$ (or $\mathbb{R}^2$ depending on the situation), and if you want to be able to perform integral computations with this data (which you need to do when discussing probability and statistics), you need an "integral" which can handle more complicated sets.

The Lebesgue integral is perhaps the most widely used such integral, although there are others which also show up in practice. Hopefully you will learn about some of this in later courses; in particular, if you're planning on going on to graduate school in economics, don't be surprised if you see Lebesgue integrals popping up every now and then. Thanks for reading.

The Riemann-Lebesgue Theorem. Lebesgue measure actually has some direct application to Riemann integration itself without moving on to the more advanced Lebesgue integral. Indeed, it gives us another characterization of functions which are Riemann integrable: a bounded function f on $[a, b]$ is Riemann integrable if and only if the set of points where f is not continuous has Lebesgue measure zero. So, integrable functions are those which are continuous almost everywhere, in the sense that the set points where they are not continuous has overall "length" zero. See my notes on the The Riemann-Lebesgue Theorem at `http://math.northwestern.edu/~scanez/courses/berkeley/math104/fall11/handouts/riemann-lebesgue.pdf` for more info.