Math 25b: Theoretical Linear Algebra and Real Analysis II

Kenny Gu
kgu@college.harvard.edu

Spring 2022

Abstract

Just some notes for Math 25b! Mostly so that I can get more familiar and play around with LaTeX but hopefully people might be able to get some use out of them too! All content by Wes Cain.

Course description: *A rigorous treatment of basic analysis. Topics include: convergence, continuity, differentiation, the Riemann integral, uniform convergence, the Stone-Weierstrass theorem, Fourier series, differentiation in several variables. Additional topics, including the classical results of vector calculus in two and three dimensions, as time allows.*

Contents

1 January 24

Welcome to Math 25b, taught by Wes Cain in Spring 2022! We'll be using Marsden and Hoffman's *Elementary Classical Analysis*, along with Abbott's *Understanding Analysis*, and Spivak's *Calculus on Manifolds*. Readings for next time: 1.4 and 1.5 from M&H.

1.1 Constructing the Naturals

We'll be constructing $\mathbb{N}+0$ using sets. Begin by assigning $\emptyset$ to the the number we associate with 0. Now, define a successor function $S(n)$ defined by $S(n) = n \cup \{n\}$. Now, we have

$$S(0) = \emptyset \cup \{\emptyset\} = \{\emptyset\}$$
$$S(1) = \{\emptyset\} \cup \{\{\emptyset\}\} = \{\emptyset, \{\emptyset\}\}$$

and so on and so forth. We won't spend so much time on this process, even though other classes like Math 112 do. Read 8.6 in Abbott for the construction of $\mathbb{R}$ from $\mathbb{Q}$ using Didekind cuts.

1.2 Completeness and the Reals

Axiom 1.1. *(Axiom of Completeness) Every nonempty subset of $\mathbb{R}$ that is bounded above has a least upper bound, which we will identity as $\sup s$, where s is our subset.*

Proposition 1.1. *(Archimedean Property of $\mathbb{R}$) Given any $x \in \mathbb{R}$, $\exists n \in \mathbb{N}$ such that $x < n$.*

Proof. (Contradiction) Suppose $\exists x \in \mathbb{R}$ such that $\forall n \in \mathbb{N}$, we have $x \geq n$. Then, x is an upper bound, so $\mathbb{N}$ has a supremum: call it $s = \sup \mathbb{N}$. So, $s - 1$ is *not* an upper bound on $\mathbb{N}$. So, $\exists n \in \mathbb{N}$ such that $s - 1 < n$, so $s < n + 1$. But, $(n+1) \in \mathbb{N}$. $\lightning$ $\square$

Corollary 1.1.1. *Given a real $\epsilon > 0$, $\exists n \in \mathbb{N}$ such that $\frac{1}{n} < \epsilon$.*

Proof. Choose $x = \frac{1}{\epsilon}$. Then, from the Archimedean Property, $\exists n > x$. Then, $\frac{1}{\epsilon} < n \implies \frac{1}{n} < \epsilon$. $\square$

Definition 1.1. *A sequence on a set S is a function $f : \mathbb{N} \to S$. Rather than $f(n)$, we'll use subscripts for notation (e.g. x_n). For specificity, we will occasionally also use $\{x_n\}_{n=1}^{\infty}$.*

Definition 1.2. *A sequence $\{x_n\}$ converges to a limit L if given any real $\epsilon > 0$, $\exists N \in \mathbb{N}$ such that $|L - x_n| < \epsilon$, $\forall n \geq N$.*

Example 1.1. *We will show that $\{\frac{1}{n}\}_{n=1}^{\infty}$ converges to 0.*

Proof. Let $\epsilon > 0$ be given. We want to produce $N = N(\epsilon)$ such that $|\frac{1}{n} - 0| < \epsilon$, $\forall n \geq N$, so we need $\frac{1}{n} < \epsilon$. Choose an integer $N > \frac{1}{\epsilon}$. $\square$ $\square$

Proposition 1.2. *The limit of a sequence in $\mathbb{R}$, if it exists, is unique.*

Proof. Suppose $\{x_n\} \to L_1$ *and* $\{x_n\} \to L_2$. We want to show that $L_1 = L_2$. Now, let $\epsilon > 0$ be given. Since $\{x_n\} \to L_1$, $\exists N_1 \in \mathbb{N}$ such that $\forall n \geq N_1$, $|L_1 - x_n| < \epsilon$. Similarly, $\exists N_2 \in \mathbb{N}$ such that $|L_2 - x_n| < \epsilon$, $\forall n \geq N_2$.

Let $N = \max\{N_1, N_2\}$. Note then that $|L_1 - L_2| = |(L_1 - x_n) + (x_n - L_2)| \leq |L_1 - x_n| + |x_n - L - 2| < 2\epsilon$. But ϵ was arbitrary, so $L_1 = L_2$. $\square$

Definition 1.3. *Suppose that $\{x_n\}$ is a sequence in $\mathbb{R}$. We say the sequence is monotone increasing if $x_1 \leq x_2 \leq x_3 \leq \cdots$*

Proposition 1.3. $\mathbb{R}$ *has the following monotone sequence property: every monotone increasing sequence bounded above converges.*

2 January 26

Second day of Math 25b! Before next lecture, make sure to read 2.1 and 2.2 of M+H.

Recall. $\mathbb{R}$ has the monotone sequence property: every monotone increasing sequence that is bounded above converges. Also, every convergent sequence in $\mathbb{R}$ is bounded; that is, $\exists M$ such that $|x_n| \leq M$, $\forall n \in \mathbb{N}$.

Proposition 2.1. *If $\{x_n\} \to x$ and $\{y_n\} \to y$, show $\left\{x_n y_n\right\} \to xy$.*

Proof. We want to show that, given $\epsilon > 0$, $\exists N = N(\epsilon) \in \mathbb{N}$ such that $\forall n \geq N$, $|x_n y_n - xy| < \epsilon$. Let's rewrite this using a nice trick: adding and subtracting the same term so we add 0.

$$\begin{aligned} |x_n y_n - xy| &= |x_n y_n - x_n y + x_n y - xy| \\ &= |x_n(y_n - y) + y(x_n - x)| \\ &\leq |x_n||y_n - y| + |y||x_n - x| \end{aligned}$$

Now, choose $M > 0$ such that $|x_n| \leq M, \forall n \in \mathbb{N}$. We know this is possible since $\{x_n\}$ is bounded. Let $\epsilon > 0$ be given. Choose $N_1 \in \mathbb{N}$ such that $|y_n - y| < \frac{\epsilon}{2M}$, $\forall n \geq N_1$. Now, choose $N_2 \in \mathbb{N}$ such that $|x_n - x| < \frac{\epsilon}{2y}$, $\forall n \geq N_2$. Now, let $N = \max\{N_1, N_2\}$. Then, for $n \geq N$

$$|x_n y_n - xy| \leq |x_n||y_n - y| + |y||x_n - x| \leq |x_n|\frac{\epsilon}{2M} + |y|\frac{\epsilon}{2y} = \epsilon$$

□

2.1 Cauchy Sequences

We will now characterize convergent real sequences without reference to "limits" by relying upon Cauchy sequences.

Definition 2.1. *A sequence $\{x_n\}$ of real numbers is called Cauchy if $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$ such that $|x_m - x_n| < \epsilon$ $\forall m, n \geq N$.*

Q : Does Cauchy imply convergent?

A : In $\mathbb{R}$, yes (we'll get to that in a second). In $\mathbb{Q}$, no. Just consider $1, 1.4, 1.41, 1.414, \ldots$

Q : Does convergent imply Cauchy?

A : If $\{x_n\} \to L$, then given $\epsilon > 0$, $\exists N = N(\epsilon)$ such that $|x_n - L| < \frac{\epsilon}{2}$ $\forall n \geq N$. Let $m, n \geq N$. Then

$$\begin{aligned} |x_m - x_n| &= |x_m - L + L - x_n| \\ &\leq |x_m - L| + |L - x_n| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

Definition 2.2. *If $\{x_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$, a subsequence is a sequence of the form $x_{k_1}, x_{k_2}, x_{k_3}, \ldots$ where $k_1 < k_2 < k_3 < \cdots$ is an increasing sequence in $\mathbb{N}$.*

Proposition 2.2. *(Bolzano-Weierstrass). Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

Proof. We're going to play a game of bisecting to construct our convergent subsequence. Say that our sequence is bounded by the interval $I_1 = [-M, M]$. Bisect I_1 as $[-M, 0] \cup [0, M]$. Because we have an infinite number of elements in the sequence, at least one of these intervals must have an infinite number of points. Call this half I_2.

Choose $k_1 = 1$ and use x_{k_1} as the first member of our subsequence. Choose $k_2 > k_1$ with $x_{k_2} \in I_2$. Bisect again with a similar procedure and choose $k_3 > k_2$ with $x_{k_3} \in I_3$. Now, keep bisecting so that $I_1 \supset I_2 \supset I_3 \supset \cdots$

If we call a_l the left endpoint of I_l, then $\{a_l\}$ must be monotone increasing and bounded above by M. So, $\{a_l\}$ must converge to a point per Prop 1.3. Call this point L.

We also know that $x_{k_m} \in I_m$ and I_m has length $\frac{M}{2^{m-2}}$. We then know that $|x_{k_m} - a_m| \leq \frac{M}{2^{m-2}}$ so we can find an N_1 such that $|x_{k_n} - a_n| \leq \frac{M}{2^{n-2}} < \frac{\epsilon}{2}$ for all $n \geq N_1$. Then, since $\{a_l\} \to L$, we know that $\exists N_2$ such that $|a_l - L| < \epsilon$, $\forall l \geq N_2$.

From these two together, let $N = \max\{N_1, N_2\}$. Then, for $m \geq N$,

$$\begin{aligned}
|x_{k_m} - L| &= |x_{k_m} - a_m + a_m - L| \\
&\leq |x_{k_m} - a_m| + |a_m - L| \\
&< \epsilon
\end{aligned}$$

□

Proposition 2.3. *If a sequence is a Cauchy sequence, then it is also bounded.*

Proof. Suppose $\{x_n\}$ is Cauchy. Let $\epsilon = 1$ be given. Now, this makes it possible to choose $N \in \mathbb{N}$ such that $|x_m - x_n| < 1$ for $m, n \geq N$. Then, if $n \geq N$, $|x_n - x_N| < 1$, $\forall n \geq N$ which implies that $|x_n| \leq |x_N| + 1$. Then, if we set $M = \max\{|x_1|, |x_2|, \ldots, |x_{N-1}|, |x_N + 1|\}$, we know that $|x_n| \leq M$ for every $n \in \mathbb{N}$. □

Theorem 2.1. *In $\mathbb{R}$, if a sequence is Cauchy, then it is also convergent.*

Proof. Let $\{x_n\}$ be Cauchy. Then $\{x_n\}$ is bounded by Prop 2.3. By Bolzano-Weierstrass, we can then choose a convergent subsequence $\left\{x_{k_j}\right\}_{j=1}^{\infty} \to L$. Now, we must make the final step: prove that the whole sequence $\{x_n\} \to L$.

Let $\epsilon > 0$ be given. Since $\{x_n\}$ is Cauchy, we can choose $N \in \mathbb{N}$ such that $|x_n - x_m| < \frac{\epsilon}{2}$, $\forall n, m \geq N$. Additionally, since the subsequence $\left\{x_{k_j}\right\} \to L$, we can choose an element of subsequence such that $|x_{k_j} - L| < \frac{\epsilon}{2}$ and $k_j \geq N$. Then,

$$\begin{aligned}
|x_n - L| &= |x_n - x_{k_j} + x_{k_j} - L| \\
&\leq |x_n - x_{k_j}| + |x_{k_j} - L| \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \quad \forall n \geq N
\end{aligned}$$

□

2.2 Cluster Points

We were running out of time here but had enough to go over a definition (and an example!) quickly.

Definition 2.3. *Let $\{x_n\}$ be a sequence in $\mathbb{R}$. A cluster point (or accumulation point) is a number x such that, $\forall \epsilon > 0$, $|x_n - x| < \epsilon$ for infinitely many n.*

Example 2.1. $\left\{(-1)^n\right\}$ *has 2 cluster points but no limit.*

3 January 28

Recall. A cluster point of a real sequence $\{x_n\}$ is a number x such that, for every $\epsilon > 0$, there are infinitely many x_n such that $|x_n - x| < \epsilon$.

Proposition 3.1. *Show that x is a cluster point of $\{x_n\}$ if and only if $\exists$ a subsequence of $\{x_n\}$ that converges to x.*

Proof. Suppose x is a cluster point. Now, let $\epsilon = 1$. Then, by definition, there exist infinitely many x_n such that $|x_n - x| < 1$. Pick a $k_1 \in \mathbb{N}$ such that $|x_{k_1} - x| < 1$.

Next, let $\epsilon = \frac{1}{2}$. Using similar logic, we know that there are infinitely many x_n such that $|x_n - x| < \frac{1}{2}$. Now choose $k_2 > k_1$ with $|x_{k_2} - x| < \frac{1}{2}$. Rinse and repeat with two conditions:

$$k_m > k_{m-1} \quad \text{and} \quad |x_{k_m} - x| < \frac{1}{2^{m-1}}$$

Then, we know that the subsequence $\left\{x_{k_m}\right\}_{m=1}^{\infty}$ that we constructed converges to x.[1] $\square$

We won't prove the converse right now, but it's not too bad; simply use the definition of convergence. Also, note that by Bolzano-Weierstrass, any bounded sequence $\{x_n\}$ contains a convergent subsequence, and by the proof above, if $\{x_n\}$ contains a convergent subsequence, then the subsequence converges to a cluster point of $\{x_n\}$. From these two statements, we can conclude that every bounded sequence has a cluster point.

3.1 Limit Inferior and Superior

Definition 3.1. *Let $\{x_n\}$ be a bounded sequence in $\mathbb{R}$. The limit superior (denoted $\limsup_{n\to\infty}\{x_n\}$) is defined as*

$$\lim_{n\to\infty}\left\{\sup_{m\geq n} x_m\right\}$$

Similarly, the limit inferior (denoted $\liminf_{n\to\infty}\{x_n\}$) is defined as

$$\lim_{n\to\infty}\left\{\inf_{m\geq n} x_m\right\}$$

Note that:

i. The sequence $\left\{\sup_{m\geq n} x_m\right\}$ is monotone decreasing because each element is at most equal to the element that came before it.

ii. Additionally, since we assumed that $\{x_n\}$ was a bounded sequence in $\mathbb{R}$, the limit superior and limit inferior must exist per the Completeness Axiom.

iii. The sequence converges if and only if $\liminf\{x_n\} = \limsup\{x_n\}$.

Example 3.1. *Consider the sequence $\{x_n\}$ defined by $x_n = (-1)^n\left[1 + \frac{1}{n}\right]$. To find the limit superior of the sequence, we will calculate*

$$\lim_{n\to\infty}\left\{\sup_{m\geq n} x_m\right\}$$

From a brief look at the sequence, we can tell that

$$\sup_{m\geq n} x_m = \begin{cases} 1 + \frac{1}{n} & n \text{ is even} \\ 1 + \frac{1}{n+1} & n \text{ is odd} \end{cases}$$

and then we can tell that this sequence $\left\{\sup_{m\geq n} x_m\right\} \to 1$ as $n \to \infty$.

[1] With the second condition we listed, note that any term that converges to 0 also works.

Proposition 3.2. *If $\{x_n\}$ is a bounded sequence, then*

$$\liminf_{n\to\infty}\{x_n\} \le \limsup_{n\to\infty}\{x_n\}$$

and then the sequence $\{x_n\}$ converges to the limit L if and only if

$$\liminf_{n\to\infty}\{x_n\} = \limsup_{n\to\infty}\{x_n\} = L$$

3.2 Metric Space Topology

Definition 3.2. *A metric space (M,d) is a set M together with a metric $d : M \times M \to \mathbb{R}$ that satisfies:*

1. *(Positivity and definiteness) $d(x,y) \ge 0$, $\forall x, y \in M$, with equality if and only if $x = y$.*
2. *(Symmetry) $d(x,y) = d(y,x)$, $\forall x, y \in M$.*
3. *(Triangle inequality) $d(x,y) \le d(x,z) + d(z,y)$, $\forall x, y, z \in M$.*

Example 3.2. *If $M = \mathbb{R}^n$, then M is a metric space with the metric*

$$d(x,y) = \|x - y\| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

which is the Euclidean norm, the metric we will use with $\mathbb{R}^n$ unless otherwise specified.

Definition 3.3. *If (M,d) is a metric space, $x \in M$ and $\epsilon > 0$, the ϵ-ball or ϵ-neighborhood of x is*

$$B(x,\epsilon) = \{y \in M : d(x,y) < \epsilon\}$$

Definition 3.4. *A subset $U \subseteq M$ is open if, given any $x \in U$, $\exists \epsilon > 0$ such that $B(x,\epsilon) \subseteq U$.*

Definition 3.5. *A neighborhood of $x \in M$ is any open set containing x.*

Example 3.3. *Let $U = \{x \in \mathbb{R}^n : |x_n| < 2\}$ with the usual metric. Prove that U is open.*

Proof. Pick a $x \in U$. Let $\epsilon = \min\{2 - x_n, x_n - (-2)\} > 0$. Then, if $y \in B(x,\epsilon)$, then we know that $\|y - x\| < \epsilon$.

If $y_n \ge 2$, then $y_n - x_n = (y_n - 2) + (2 - x_n) \ge \epsilon$, which demonstrates that it cannot be in the ball because $\|y - x\| \ge |y_n - x_n| \ge \epsilon$ which contradicts that $y \in B(x,\epsilon)$. Similarly, we can find a contradiction if $y_n < -2$. □

Example 3.4. *If $U \subseteq \mathbb{R}$ has a countably infinite number of elements, U cannot be open. For instance, $|x - y| < \epsilon$ defines an interval $(x - \epsilon, x + \epsilon)$ which must have an uncountable number of elements.*

Proposition 3.3. *If (M,d) is a metric space and $x \in M$, then an open ball $B(x,\epsilon)$ is open.*

Proof. Pick $y \in B(x,\epsilon)$. We want to produce a $\delta > 0$ such that $B(y,\delta) \subseteq B(x,\epsilon)$. Let's choose $\delta = \epsilon - d(x,y)$ and form $B(y,\delta)$. We'll demonstrate that $B(y,\delta) \subseteq B(x,\epsilon)$ through our usual technique: that if $z \in B(y,\delta)$, then $z \in B(x,\epsilon)$. If $z \in B(y,\delta)$, then $d(y,z) < \delta = \epsilon - d(x,y)$. Additionally, by the Triangle Inequality, we have

$$\begin{aligned} d(x,z) &\le d(x,y) + d(y,z) \\ &< d(x,y) + \epsilon - d(x,y) = \epsilon \end{aligned}$$

so $z \in B(x,\epsilon)$. □

Proposition 3.4. *If (M,d) is a metric space, then*

1. *$\emptyset$ and M are open.*
2. *Intersections of finitely many open sets are open.*
3. *Arbitrary unions of open sets are open.*

Example 3.5. *Consider the following set: $\bigcap_{n=1}^{\infty} U_n$ where $U_n = \left(-\frac{1}{n}, \frac{1}{n}\right)$. Then, the set is equal to $\{0\}$ which is not open.*

Definition 3.6. *A subset K of a metric space (M,d) is closed if its complement $M \setminus K = K^c$ is open.*

Note, for instance, that $\emptyset$ and M are both closed *and* open.

Example 3.6. *In $\mathbb{R}^n$, a singleton set $\{x\}$ is closed. Most sets, however, are neither open nor closed.*

Proposition 3.5. *If (M,d) is a metric space, a set $K \subseteq M$ is closed if and only if every convergent sequence of points $x_n \in K$ converges to a point in K. That is, $d(x_n, x) \to 0$ as $n \to \infty$.*

Example 3.7. *Define $\{x_n\} = \{1 - \frac{1}{n}\} \to 1$. Then, we know that $\{x_n\} \subseteq [0,1)$, but it converges to something not in $[0,1)$.*

3.3 Limit Points and Interior Points

Definition 3.7. *Let (M,d) be a metric space, and let $E \subseteq M$. Call a point $p \in M$ a limit point if every open neighborhood of p contains a point $q \neq p$ with $q \in E$.*

Proposition 3.6. *If $E \subseteq M$ (M a metric space), p is a limit point of E if and only if $p = \lim_{n\to\infty} \{x_n\}$ for some sequence $x_n \in E \setminus \left\{p\right\}$. $\forall n \in \mathbb{N}$*

Proof. ($\Longrightarrow$) If p is a limit point, consider the open balls $B(p, \frac{1}{n})$ for each $n \in \mathbb{N}$. From $B(p, \frac{1}{n})$, we can choose a point $x_n \in E$ such that $x_n \neq p$. Then, we know that, given $\epsilon > 0$, we can pick a sufficiently large N such that $|x_n - p| < \frac{1}{n} < \epsilon$ for all $n \geq N$. □

For homework, we'll be asked to prove that $K \subseteq M$ is closed if and only if K contains all of its limit points.

Definition 3.8. *Let $E \subseteq M$ where (M,d) is a metric space. A point $x \in E$ is an interior point of E if $\exists$ an open set U such that $x \in U \subseteq E$. The interior of E, denoted $\text{int}(E)$, is the set of all interior points of E.*

Example 3.8. *The intervals $(0,1), (0,1], [0,1]$ all have the interior $(0,1)$.*

Proposition 3.7. *If (M,d) is a metric space and $E \subseteq M$, then $\text{int}(E)$ is the union of all open subsets of E.*

Proof. Let U be the union of all open subsets of E. We will show that $U \subseteq \text{int}(E)$.

If $x \in U$, then x is in some open subuset of E. Thus, $x \in \text{int}(E)$. So, $U \subseteq E$.

Now, if $x \in \text{int}(E)$ then $\exists$ an open set V with $x \in V \subseteq E$. Since $V \subseteq U$, then $x \in U$. So, $\text{int}(E) \subseteq U$. □

Corollary 3.7.1. *$\text{int}(E)$ is open and is the largest open subset of E. Also, E is open if and only if $\text{int}(E) = E$.*

4 January 31

Recall. If (M,d) is a metric space and $E \subseteq M$, we say $x \in E$ is an interior point of E if $\exists$ an open set U such that $x \in U \subseteq E$. The interior of E is the set of all interior points, denoted $\text{int}(E)$.

Example 4.1. *If $M = \mathbb{R}$ with $d(x,y) = |x-y|$, then* $\text{int}[0,1) = (0,1)$.

Example 4.2. *Let $M = \mathbb{R}$ with the discrete metric*

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

Then, we know that, $\forall x \in M$,

$$B(x,\epsilon) = \{x\}, 0 < \epsilon < 1 \quad \text{and} \quad B(x,\epsilon) = \mathbb{R}, \epsilon > 1$$

From this, we then know that every set is open because for a single element, we can choose an ϵ such that the ϵ-ball is the set itself, and we also know that the union of an arbitrary number of open sets is also open.

Then, we also know that the complement of any set (which is a set) is also open, which implies that all sets are also closed.

All sets, therefore, are both closed and open (they're clopen*!). Additionally, we also then have* $\text{int}(E) = E$.

4.1 Closure

Definition 4.1. *Let (M,d) be a metric space and let $E \subseteq M$. The closure of E, denoted* $\text{cl}\ (E)$, *is the intersection of all closed sets containing E.*

Note here that $\text{cl}\ (E)$ is a closed set (the smallest closed set that contains E) because intersections of an arbitrary number of closed sets are also closed.

Example 4.3. *For $M = \mathbb{R}$, $d(x,y) = |x-y|$, then* $\text{cl}\ (0,1] = [0,1]$.

Example 4.4. *Let $M = \mathbb{R}^n$ with the standard Euclidean metric. Now, consider*

$$B(x,\epsilon) = \left\{ y \in \mathbb{R}^n : \|x-y\| < \epsilon \right\}$$

Then, we know that $\text{cl}\ B(x,\epsilon)$ *is*

$$\overline{B(x,\epsilon)} = \left\{ y \in \mathbb{R}^n : \|x-y\| \leq \epsilon \right\}$$

which is a closed ball of radius ϵ centered at x.[2] *Let's check this.*

Proof. We will proceed by contradiction. Suppose that $\text{cl}(B,\epsilon) = K$ were a proper subset of $\overline{B(x,\epsilon)}$. Then, $\exists$ some point $y^* \notin K$ such that $\|y^* - x\| = \epsilon$.

We will now construct a convergent sequence to find a limit point. Consider the sequence $\left\{y_n\right\}$ where

$$y_n = x + (y^* - x)\left(1 - \frac{1}{n}\right)$$

□

[2] We will use the notation $\overline{B(x,\epsilon)}$ for these closed balls.

which essentially inches closer and closer from x to y^*. We know that $y_n \in B(x,\epsilon), \forall n \in \mathbb{N}$ because

$$\begin{aligned}
\|y_n - x\| &= \|(y^* - x)\left(1 - \frac{1}{n}\right)\| \\
&= \|(y^* - x)\|\left(1 - \frac{1}{n}\right) \\
&= \epsilon\left(1 - \frac{1}{n}\right) < \epsilon
\end{aligned}$$

But we also know then that $\left\{y_n\right\} \to y^*$, so y^* is a limit point of $B(x,\epsilon)$. K could not have been closed. $\lightning$

4.2 Boundaries

Definition 4.2. *Let (M,d) be a metric space and $E \subseteq M$. The boundary of E is the set $\partial E = (\mathrm{cl}E) \cap (\mathrm{cl}E^c)$*

Example 4.5. $\partial(0,1) = \{0,1\}$

Example 4.6. *Let $x \in \mathbb{R}$ and consider $E = \{x\}$. Then, we know that $\mathrm{cl}(E) = \{x\}$ and $\mathrm{cl}(E^c) = \mathbb{R}$. Together, these imply that $\partial E = \{x\}$.*

Now, let's try to chararcterize ∂E:

Proposition 4.1. *Let (M,d) be a metric space and have $E \subseteq M$. Then, $x \in \partial E$ if and only if $\forall \epsilon > 0, B(x,\epsilon) \cap E \neq \emptyset$ and $B(x,\epsilon) \cap E^c \neq \emptyset$.*

Lemma 4.1.1. *We will show that $\mathrm{cl}(E) = E \cup \left\{\text{limit points of } E\right\}$.*

Proof. This is an incomplete proof.

Let $S = \mathrm{cl}(E) = E \cup \left\{\text{limit points of } E\right\}$. Per problem set 1, any closed set containing E must also contain S. Now, we just need to show that S is closed by showing that, if y is a limit point of S, then $y \in S$. $\square$

Proof. Now back to the proof for Prop 4.1. Suppose $x \in \partial E = \mathrm{cl}E \cap \mathrm{cl}E^c$. Then, we know that either $x \in E$ or $x \in E^c$ simply by the definition of complement.

Suppose first that $x \in E$, which implies that $x \notin E^c$. But, we know that $x \in \mathrm{cl}E^c$, which implies that x is a limit point of E^c per the lemma. By the definition of a limit point, every open neighborhood of x contains points of E^c. other than x.

A similar argument works for $x \in E^c$. $\square$

4.3 Sequences and Completeness Revisited

Definition 4.3. *Let (M,d) be a metric space, and $\{x_n\}$ a sequence in M. Then, we say $\{x_n\}$ converges to $x \in M$ if given any open set containing x, $\exists N \in \mathbb{N}$ such that $x_k \in U$, $\forall k \geq N$. Equivalently, given $\epsilon > 0$, $\exists N \in \mathbb{N}$ such that $x_k \in B(x,\epsilon)$, $\forall k \geq N$.*

Proposition 4.2. *If $M = \mathbb{R}^n$ with the usualy Euclidean metric, convergence $\{x_k\} \to x$ is equivalent to component-wise converge.*

Proof. For notation, we will write $x_k = \left(x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(n)}\right)$ where $x_k^{(i)}$ is the ith component of x_k.

Let $\sqrt{\epsilon} > 0$ be given. If $\{x_m\} \to x$, then $\exists N \in \mathbb{N}$ such that $\|x_m - x\| < \sqrt{\epsilon}$, $\forall m \geq N$. Then,

$$\begin{aligned} \epsilon &> \|x_m - x\|^2 \\ &= \sum_{j=1}^{n} \left(x_m^{(j)} - x^{(j)} \right)^2 \\ &\geq \left(x_m^{(l)} - x^{(l)} \right)^2 \geq 0 \end{aligned}$$

So $\left\{ x_m^{(l)} \right\} \to x^{(l)}$. $\square$

5 February 2

Last time, we generalized notions from $\mathbb{R}$ to arbitrary metric spaces. We'll continue doing so today, so some definitions may seems familiar.

5.1 Generalizing from Reals to Metric Spaces

Proposition 5.1. *Let (M,d) be a metric space. Then, $K \subseteq M$ is closed if and only if for every convergent sequence $\{x_n\}$ with $x_n \in K$ the limit of the sequence lies in K.*

Proof. Proof is in the homework. □

Definition 5.1. *If (M,d) is a metric space and $\{x_n\}$ is a sequence in M, we say that $\{x_n\}$ is Cauchy if given $\epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall m,n \geq N$, we have $d(x_m,x_n) < \epsilon$.*

Definition 5.2. *We say M is complete if every Cauchy sequence in M converges to some element of M.*

Definition 5.3. *A sequence is bounded if $\exists x^* \in M$ and a $R > 0$ such that $d(x_n,x^*) \leq R$, $\forall n \in \mathbb{N}$.*

Proposition 5.2. *In a metric space (M,d), the following are true:*

1. *Convergent $\implies$ Cauchy*
2. *Convergent $\implies$ bounded*
3. *Cauchy $\implies$ bounded*

Recall from Math 25a: in a normed vector space, the norm induces a metric $d(x,y) = \|x-y\|$. We'll use this by default when applicable, unless otherwise specified.

5.2 Series and Convergence

Later in the course, we'll want to define functions like e^x, $\ln x$, $\cos x$ and derive their properties. We'll define these in terms of infinite series of functions as terms. But, before we get there, we'll start by understanding series of numbers.

Definition 5.4. *A series of real numbers $\displaystyle\sum_{k=1}^{\infty} a_k$ converges to a limit L if the sequence of its partial sums $s_n = \displaystyle\sum_{k=1}^{n} a_k$ converges to L as $n \to \infty$. That is, $\{s_n\}_{n=1}^{\infty} \to L$. A series that doesn't converge thus diverges .*

Example 5.1. *The harmonic series defined as*

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverges. Here, we can easily show that the partial sums are strictly monotone increasing, and we can also show that given a real $\xi > 0$, $\exists M$ with $s_m > \xi$, so it is unbounded.

Definition 5.5. *A series $\sum_{n=1}^{\infty} a_n$ is geometric if $\exists r \in \mathbb{R}$ such that*

$$a_{n+1} = ra_n \quad \forall n \in \mathbb{N}.$$

Then, if $r \neq 0$, r is called the common ratio .

Proposition 5.3. *A geometric series $\sum_{n=1}^{\infty} a_n$ converges if and only if $|r| < 1$.*

Proof. If $a_1 = 0$, this proof is trivial. The rest of the proof will assume that $a_1 \neq 0$. Then, for $k \in \mathbb{N}$, write

$$S_k = a_1 + ra_1 + r^2 a_1 + \cdots + r^{k-1} a_1$$
$$rS_k = ra_1 + r^2 a_1 + \cdots + r^{k-1} a_1 + r^k a_1$$

so then $S_k - rS_k = a_1 - r^k a_1$ which simplifies to $(1-r)S_k = a_1(1-r^k)$. Then, if $r = 1$, $S_k = kn$, so $\{s_k\}_{k=1}^{\infty}$ diverges because $a_1 \neq 0$. Else, we have

$$s_k = \frac{a_1(1-r^k)}{1-r}$$

By the limit property of sequences, $\{s_k\} = \frac{a_1}{1-r}$. Now, we simply prove that if $|r| > 1$ or if $r = -1$, then the sequence diverges. ☐

Proposition 5.4. *(Comparison Test v.1.0) Suppose that $\sum_{k=1}^{\infty} a_k$ is a convergent series of real numbers with non-negative terms. If $0 \leq b_k \leq a_k$ for each $k \in \mathbb{N}$, then $\sum_{k=1}^{\infty} b_k$ converges. Similarly, if $\sum_{k=1}^{\infty} a_k$ diverges and $c_k \geq a_k$ for $k \in \mathbb{N}$, then $\sum_{k=1}^{\infty} c_k$ diverges as well.*

Proof. Assume that $\sum_{k=1}^{\infty} a_k$ converges to L. Look at partial sums

$$A_n = \sum_{k=1}^{n} a_k \qquad B_n = \sum_{k=1}^{n} b_k$$

which must then have $0 \leq B_n \leq A_n \leq L$. Then, B_n is monotone increasing and bounded above by L, so, by the completeness axiom, $\sum_{k=1}^{\infty} b_k$ converges. ☐

Definition 5.6. *A series $\sum_{k=1}^{\infty} a_k$ is alternating if $a_k a_{k+1} < 0$, $\forall k \in \mathbb{N}$. (i.e. consecutive terms have opposite signs).*

Example 5.2. *Consider*

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

Then, the partial sums are $S_n = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + (-1)^{n+1}\frac{1}{n}$. But, consider the rearrangement $1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \cdots$ where we delay the negative elements until the end. But $\frac{1}{2n-1} > \frac{1}{2n} = \frac{1}{2}\left(\frac{1}{n}\right)$ and we know that that series diverges. So, it seems as though rearranging this series causes it to diverge.

Proposition 5.5. *If $\sum_{n=1}^{\infty} a_n$ converges, then $\{a_n\} \to 0$ as $n \to \infty$.*

Proof. We will approach via the contrapositive. Suppose that $\{a_n\}$ does not converge to 0. Then, $\exists \epsilon > 0$ such that $\forall n \in \mathbb{N}$, $|a_k| \geq \epsilon$ for some $k \geq N$. Now, given $N \in \mathbb{N}$, pick $k \geq N$ such that $|a_{k+1}| \geq \epsilon$. But, $a_{k+1} = S_{k+1} - S_k$, so $|S_{k+1} - S_k| \geq \epsilon$ with both k and $k+1$ at least N. So $\{s_k\}_{k=1}^{\infty}$ is not Cauchy, and since we're working in $\mathbb{R}$, it is also not convergent, so $\{s_k\}$ diverges. ☐

Definition 5.7. *A series of real numbers $\sum_{k=1}^{\infty} a_k$ converges absolutely if $\sum_{k=1}^{\infty} |a_k|$ converges. If a series of real numbers $\sum_{k=1}^{\infty} a_k$ converges but $\sum_{k=1}^{\infty} |a_k|$ diverges, then the series is said to converge conditionally .*

Example 5.3. *The alternating harmonic series described in example 5.2 is convergent but not absolutely. It is thus conditionally convergent – conditional upon the order of the terms.*

Proposition 5.6. *Absolute convergence implies convergence.*

Proof. We won't prove this fully, but we could do so by showing that the limit of the series is the limit of a Cauchy series of partial sums. ☐

Proposition 5.7. *(Ratio Test). Suppose $\sum_{n=1}^{\infty} a_n$ is a series whose terms are non-zero real numbers, and that*

$$L = \lim_{n\to\infty} |\frac{a_{n+1}}{a_n}|$$

exists. If $L < 1$, then the series converges absolutely and if $L > 1$, the series diverges.

Proof. Suppose $L < 1$ and $\lim_{n\to\infty} |\frac{a_{n+1}}{a_n}| = L$. Choose b such that $L < b < 1$. Since this limit is L, $\exists N \in \mathbb{N}$ such that $|\frac{a_{n+1}}{a_n}| < b$, $\forall n \geq N$. Thus, $|a_{N+1}| < b|a_N|$. More generally, for each $p \in \mathbb{N}$, $|a_{N+p}| < b^p|a_N|$.

We know that $\sum_{p=1}^{\infty} |a_{N+p}|$ is a sequence of positive terms and $\sum_{p=1}^{\infty} b^p|a_N|$ is a convergent geometric series because $b < 1$. So, the comparison test implies that $\sum_{p=1}^{\infty} |a_{N+p}|$ also converges. Then, by reindexing, $\sum_{n=N+1}^{\infty} |a_n|$ converges. So,

$$|a_1| + |a_2| + \cdots + |a_N| + \sum_{n=N+1}^{\infty} |a_n|$$

also converges, so the initial series converges absolutely. □

Note that, when working with the convergence and divergence of series, it's perfectly fine to throw out finitely many terms.

Proposition 5.8. *(Alternating Series Test) Suppose that $\sum_{n=1}^{\infty} a_n$ is an alterating series such that $|a_{n+1}| \leq |a_n|$ for each n and $\lim_{n\to\infty}\{a_n\} = 0$. Then, the series $\sum_{n=1}^{\infty} a_n$ converges.*

Proof. Rewrite $\sum_{n=1}^{\infty} a_n$ as $\sum_{n=1}^{\infty} (-1)^{n+1} b_n$ where $b_n > 0$ for each n. (Here, WLOG, assumed $a_1 > 0$) Note, $b_n = |a_n|$. So $b_1 \geq b_2 \geq b_3 \geq \ldots$ and $\lim_{n\to\infty}\{b_n\} = 0$. Look at the odd and even indexed partial sums. $S_{2n} = (b_1 - b_2) + (b_3 - b_4) + \cdots + (b_{2n-1} - b_{2n})$ so we have $S_2 \leq S_4 \leq S_6 \leq \ldots$

Similarly, $S_{2n+1} = b_1 - (b_2 - b_3) - (b_4 - b_5) - \cdots - (b_{2n} - b_{2n+1})$. Terms in parentheses are non-negative, so $S_1 \geq S_3 \geq S_5 \geq \ldots$

Note that $S_{2n+1} = S_{2n} + b_{2n+1} \geq S_{2n}$ for each n. Each even-indexed partial sum is thus bounded above by any given odd-indexed partial sum. So, by the monotone convergence theorem, $\{S_{2n}\}_{n=1}^{\infty}$ converges to some limit – call it L_{even}. Likewise, $\{S_{2n+1}\}_{n=1}^{\infty}$ converges to some limit – call it L_{odd}.

We then know that

$$\begin{aligned} |L_{odd} - L_{even}| &\leq |S_{2n+1} - S_{2n}| \\ &= |a_{2n+1}| \\ &= b_{2n+1} \end{aligned}$$

which goes to 0 as $n \to \infty$. □

6 February 4

Concluding remarks about infinite series: we can generalize some notions to series in a normed vector space V. For instance, consider $\sum_{n=1}^{\infty} a_k$ where $a_k \in V$. The series converges absolutely if $\sum_{k=1}^{\infty} \|a_k\|$ converges, which we also know is a series of real numbers. Then, we can also show that absolute convergence implies convergence in a complete normed vector space.

6.1 Compactness

Today, we'll be continuing with topology, with a focus on compactness, an especially important property in formulating some big theorems like the Extreme Value Theorem.

Let's begin by defining what compactness is, which will first require some other definitions:

Definition 6.1. *Let (M,d) be a metric space and $A \subseteq M$. A collection $\{\Omega_\alpha\}_{\alpha\in I}$ of open sets is called an open cover of A if*

$$\bigcup_{\alpha\in I} \Omega_\alpha \supseteq A$$

Here, I is an index set that is not necessarily countable.

Definition 6.2. *A subcover of an open cover is a subcollection of open sets from an open cover that also covers A.*

Definition 6.3. *A subset K of a metric space is compact if every open cover of K has a finite subcover.*

Example 6.1. *Consider the interval $(0,1] \subset \mathbb{R}$ with the usual metric. We claim that it's not compact.*

Proof. Let $\Omega_\alpha = \left(\frac{\alpha}{2}, 2\right)$ for $\alpha \in (0,1]$. Then, we know that

$$\bigcup_{\alpha\in(0,1]} \Omega_\alpha \supseteq (0,1]$$

because if $x \in (0,1]$ then $x \in \Omega_x$.

But, what if there was a finite subcover? If $\Omega_{\alpha_1}, \Omega_{\alpha_2}, \ldots, \Omega_{\alpha_n}$ were a finite subcover and $\alpha_1 < \alpha_2 < \cdots < \alpha_n$, then we know that

$$\Omega_{\alpha_1} \supseteq \Omega_{\alpha_2} \supseteq \cdots \supseteq \Omega_{\alpha_n}$$

Then, we also know that Ω_{α_1} doesn't contain $\frac{\alpha_1}{4}$, so $(0,1]$ is not compact. □

Example 6.2. *The interval $[0,1]$ is also compact, but that's harder to show!*

Example 6.3. *Consider the interval $[0,\infty)$, which is a closed subset of $\mathbb{R}$. Then, we have $\Omega_n = \left(n - \frac{2}{3}, n + \frac{2}{3}\right)$ for each $n = 0,1,2,\ldots$. Then, $\bigcup_{n=0}^{\infty} \Omega_n$ covers $[0,\infty)$, and Ω_n is open, but there is no finite subcover. Thus, the interval is not compact!*

Now, let's find an equivalent way of detecting compactness using the language of sequences.

Definition 6.4. *Let (M,d) be a metric space and let $A \subseteq M$. We say that A is sequentially compact if every sequence in A has a subsequence that converges in A.*

Example 6.4. *Consider the interval $[0,1]$, which is a subset of $M = \mathbb{R}$ with the usual metric. Since it is bounded, by Bolzano-Weierstrass, every sequence in $[0,1]$ has a convergent subsequence. We also know that the limit must be in $[0,1]$ because $[0,1]$ is closed.*

We now turn our attention to showing that $A \subseteq M$ is compact if and only if A is sequentially compact in the metric space.

Proposition 6.1. *If $A \subseteq M$ is compact, then A is closed.*

Proof. We will show that A^c is open.

Pick some arbitrary point $x \notin A$. Then, let $\Omega_n = \left\{y \in M : d(x,y) > \frac{1}{n}\right\}$. It's easy to show that each Ω_n is open and that

$$\bigcup_{n=1}^{\infty} \Omega_n = M - \{x\}.$$

But, $x \notin A$, so $\bigcup_{n=1}^{\infty}$ is an open cover of A. Since A compact, we can pick a finite subcover $\Omega_{n_1}, \Omega_{n_2}, \ldots, \Omega_{n_k}$ where $n_1 < n_2 < \cdots < n_k$. Then, we have $\Omega_{n_k} \supseteq \Omega_{n_{k-1}} \supseteq \cdots \supseteq \Omega_{n_1}$.

Since $\Omega_{n_k} = \left\{y \in M : d(x,y) > \frac{1}{n_k}\right\}$, we can form a ball $B(x, \frac{1}{n_k})$ which is contained in $M \backslash A$, so $M \backslash A$ is open. $\square$

Proposition 6.2. *If (M,d) is a metric space and M is compact, then any closed $K \subseteq M$ is also compact.*

Proof. If $\bigcup_{\alpha \in I} \Omega_\alpha$ is an open cover of K. Since K is closed, we know $M \setminus K$ is open. Then, simply append $M \setminus K$ to an open cover of K. Since M is compact, we can choose a finite subcover of M, which we then know is also a finite subcover of K. $\square$

Theorem 6.1. *(Bolzano-Weierstrass) A subset of a metric space is compact if and only if it is sequentially compact.*

Proof. ($\Longrightarrow$) We will proceed by contradiction. Suppose $A \subseteq M$ is compact but not sequentially compact. Then, $\exists$ a sequence $\{x_n\}_{n=1}^{\infty}$ in A with no convergent subsequence in A. Note that $\{x_n\}$ has infinitely many distinct points (otherwise, we can simply pull a subsequence with all terms being the same). Then, WLOG $\{x_n\}_{n=1}^{\infty}$ has all distinct points.

We will now build an open cover of A with no finite subcover. Given a point x_k in the sequence ,choose an $\epsilon_k > 0$ such that $B(x_k, \epsilon_k)$ contains no points in the sequence (other than x_k), which we know is possible by the assumption of no convergent subsequences in A).

Now, if we look at the set $K = \{x_1, x_2, x_3, \ldots\}$, we know it's closed because it contains all of its limit points (vacuously, because it there are no limit points, so it must contain all of them). So, K is closed, and since $K \subseteq A$ and A is compact, we know K is also compact.

But, $\left\{B(x_k, \epsilon_k\right\}_{k=1}^{\infty}$ is an open cover of K, and we can't have a finite subcover because throwing out $B(x_l, \epsilon_l)$ leaves x_l uncovered, which contradicts our original assumption. $\lightning$

We will not consider the converse at this point. $\square$

Proposition 6.3. *Let (M,d) be a metric space. If $K \subseteq M$ is compact, then K is bounded.*

Proof. Pick any $x \in K$. Define $\Omega_n = B(x,n)$ for $n \in \mathbb{N}$. Then,

$$\bigcup_{n=1}^{\infty} \Omega_n = M \supseteq K$$

so it forms an open cover. We know that K is compact though, so we can choose a finite subcover: $\Omega_{n_1}, \Omega_{n_2}, \ldots, \Omega_{n_l}$ with $n_1 < n_2 < \cdots < n_l$. Then, $\Omega_{n_l} = B(x, n_l) \supseteq K$, so K is bounded. $\square$

We have so far demonstrated that compactness implies closed and bounded in any metric space. In $M = \mathbb{R}^n$ with the usual Euclidean topology, the converse is true as well.

Theorem 6.2. *(Heine-Borel) In $\mathbb{R}^n$, with the usual metric, $A \subset \mathbb{R}^n$ is compact if and only if A is both closed and bounded.*

7 February 7

Now, jumping back to where we left off, we will prove the Heine-Borel Theorem.

7.1 Heine-Borel Theorem

Theorem 7.1. *(Heine-Borel, backwards) If $M = \mathbb{R}^n$ with the usual Euclidean metric, if $K \subset M$ and K is closed and bounded, then K is compact.*

Proof. Let $K \subset \mathbb{R}^n$ be closed and bounded. We will show that K is sequentially compact and therefore compact.

Pick a sequence $\{x_m\}_{m=1}^{\infty}$ in K. Write

$$x_m = \left(x_m^{(1)}, x_m^{(2)}, \dots, x_m^{(n)}\right)$$

where $x_m^{(i)}$ is the ith component of x_m. Then look at the sequence of first components – the sequence $\left\{x_m^{(1)}\right\}_{m=1}^{\infty}$, which is bounded by hypothesis. Then, by Bolzano-Weierstrass, we can choose $1 \le m_1 < m_2 < m_3 < \cdots$ such that $\left\{x_{m_l}^{(1)}\right\}_{l=1}^{\infty}$ converges to some limit $x^{(1)}$.

Now, move on to the second component. Look at $\left\{x_{m_l}^{(2)}\right\}_{l=1}^{\infty}$ and extract a further convergent subsequence that converges $x^{(2)}$.

Repeat this process for each component. Now, using the indices obtained at the very end of this process, we get a subsequence of $\{x_m\}_{m=1}^{\infty}$ that converges to $x \in \mathbb{R}^n$. Since K is closed, we know that $x \in K$, so K is sequentially compact. □

7.2 Connectedness

Connectedness is a concept that won't be used as much compared to some of the other concepts we've introduced thus far, but it will be useful for some integral theorems.

Definition 7.1. *Let (M, d) be a metric space and $E \subseteq M$. We say two open sets Ω_1 and Ω_2 separate E if*

1. $E \subseteq \Omega_1 \cup \Omega_2$
2. $\Omega_1 \cap E \neq \emptyset$ *and* $\Omega_2 \cap E \neq \emptyset$
3. $(\Omega_1 \cap E) \cap (\Omega_2 \cap E) = \emptyset$

Note that a singleton set cannot be separated.

Definition 7.2. *E is disconnected if there exists a separation.*

Definition 7.3. *E is connected if it is not disconnected.*

Example 7.1. *To separate $\mathbb{Q}$, use $\Omega_1 = (-\infty, \sqrt{2})$ and $\Omega_2 = (\sqrt{2}, \infty)$.*

Definition 7.4. *A set $I \subseteq \mathrm{R}$ is an interval if whenever $x, y \in I$ with $x < y$, we have $z \in I$ whenever $x < z < y$.*

Example 7.2. *Intervals include $\emptyset$, singletons, (a, b), $(a, b]$, $[a, b)$, $[a, b]$, $(-\infty, a)$, $(-\infty, a]$, (b, ∞), $[b, \infty)$, $\mathbb{R}$.*

Proposition 7.1. *A nonempty subset of $\mathbb{R}$ is connected if and only if it is an interval.*

Proof. ($\Longrightarrow$) Contraposition. Assume A not an interval. Choose $x, y \in A$ with $x < y$ and $z \notin A$ with $x < z < y$. Use $\Omega_1 = (-\infty, z)$ and $\Omega_2 = (z, \infty)$. Then, we know that the two are nonempty open sets with $\Omega_1 \cup \Omega_2 \supseteq A$ and $(\Omega_1 \cap A) \cap (\Omega_2 \cap A) = \emptyset$.

($\Longleftarrow$) Contraposition. Assume A is not connected. Then, since A is not connected, we can choose open sets Ω_1 and Ω_2 that separate A. Since $A \cap \Omega_1 \neq \emptyset$ and $A \cap \Omega_2 \neq \emptyset$, choose $a \neq b$ as elements of sets. Then, without loss of generality, assume that $a < b$. We also know that $a \neq b$.

Now, define $s = \sup(\Omega_1 \cap [a, b])$. We know then that $a \in \Omega_1 \cap [a, b]$ and that $\Omega_1 \cap [a, b]$ is bounded above by b, so s makes sense. We will then show that $a < s < b$ and that $s \notin A$, which will be sufficient to demonstrate that A is not an interval.

First, let's show that $a < s$. Let $x \in \Omega_1 \cap [a, b)$. Since Ω_1 is open and $a < b$, we can choose an $\epsilon > 0$ such that $[x, x + e) \subseteq \Omega_1 \cap [a, b) \subseteq \Omega_1$. It is impossible for $x = s$ because then $x + \frac{\epsilon}{2} \in \Omega_1 \cap [a, b)$, which would contradict x being the supremum because then $a \leq x < x + \frac{\epsilon}{2} < s$.

Next, let's show that $x < b$. Choose some $y \in (a, b] \cap \Omega_2$. Since Ω_2 is open and $a < b$, we can choose $\delta > 0$ such that $(y - \delta, y] \subseteq (a, b] \cap \Omega_2 \subseteq U_2$. Then, we know that $(y - \delta, y]$ is disjoint from Ω_1 because $(\Omega_1 \cap A) \cap (\Omega_2 \cap A) = \emptyset$. But $y \neq s$ because $y - \frac{\delta}{2}$ would be a smaller upper bound on $\Omega_1 \cap [a, b]$. Thus, since y was an arbitrary element of $(a, b] \cap \Omega_2$, we can conclude that $s \notin [a, b] \cap \Omega_2$ and that $s < b$.

Now, we know that

1. $a < s < b$
2. $s \notin \Omega_1$ and $s \notin \Omega_2 \implies s \ni n\Omega_1 \cup \Omega_2$.

But $\Omega_1 \cap \Omega_2 \subseteq A$, which then implies that A is not an interval, since $s \notin A$.

7.3 Path Connectedness

Connectedness is much harder to prove directly for other spaces like $A \subseteq \mathbb{R}^2$. So, we will now define path connectedness and show that path connectedness implies connectedness.

Definition 7.5. *Let (M, d) be a metric space, and let $a < b$ be real. A function $\varphi : [a, b] \to M$ is continuous if for every sequence $\{t_k\}$ in $[a, b]$ that converges to $t \in [a, b]$, we have $\left\{\varphi(t_k)\right\}_{k=1}^{\infty} \to \varphi(t)$ in M.*

Definition 7.6. *A continuous path joining $x \in M$ to $y \in M$ is a continuous function $\varphi : [a, b] \to M$ with $\varphi(a) = x$ and $\varphi(b) = y$.*

Definition 7.7. *We say that $E \subseteq M$ is path connected if, given any $x, y \in E$, there exists a continuous path joining x and y with $\varphi(t) \in E, \forall t \in [a, b]$.*

Example 7.3. *Let $r > 0$. Then, $B(x, r) \subseteq \mathbb{R}^n$ is path connected. Define $[0, 1] \to B(x, r)$ by $\varphi(t) = y + t(z - y)$. Then, for $t \in [0, 1]$, we have*

$$\begin{aligned}
\|\varphi(t) - x\| &= \|y + t(z - y) - x\| \\
&= \|(1 - t)(y - x) + t(z - x)\| \\
&\leq (1 - t)\|y - x\| + t\|z - x\| \\
&< (1 - t)r + tr = r
\end{aligned}$$

Now, for continuity, suppose that $\{t_k\} \to t$ in $[0, 1]$. Then,

$$\begin{aligned}
\|\varphi(t_k) - \varphi(t)\| &= \|y + t_k(z - y) - y - t(z - y)\| \\
&= |t_k - t|\|z - y\| \to 0 \text{ as } k \to \infty
\end{aligned}$$

□

8 February 9

Last time, we said that, if (M,d) was a metric space and $A \subseteq M$, if A is path connected then A is connected. Later in this course, the connectedness of a domain will be important for certain theorems.

8.1 Limits

We're here! The notion of "limits" is what distinguishes calculus-based math from other kinds of math. For functions $f : A \subseteq \mathbb{R} \to \mathbb{R}$, many of us interpreted $\lim_{x\to x_0} f(x) = L$ as "what would you think $f(x)$ is approaching if you weren't allowed to look at x_0?" Today, we'll approach limits more rigorously.

Setup: let (M, d_M) and (N, d_N) be metric spaces. Then, let $A \subseteq M$ and x_0 a limit point of A.

Definition 8.1. *If $f : A \to N$, then we say that $L \in N$ is the limit of f at x_0 if, given any $\epsilon > 0$, $\exists \delta > 0$ such that $\forall x \in A$, we have $0 < d_M(x, x_0) < \delta \implies d_N(f(x), L) < \epsilon$.*

Example 8.1. *Let $A = \mathbb{R} \setminus \{1\}$ and $f : A \subseteq \mathbb{R} \to \mathbb{R}$ defined as*

$$f(x) = \frac{x^2-1}{x-1}$$

for $x \in \mathbb{R} \setminus \{1\}$. Then, we claim that $\lim_{x\to 1} f(x) = 2$.

Proof. Let $\epsilon > 0$ be given. If $x \neq 1$, then

$$\begin{aligned}|f(x) - L| &= \left|\frac{x^2-1}{x-1} - 2\right| \\ &= \left|(x+1) - 2\right| \\ &= |x-1|\end{aligned}$$

so if $\delta = \epsilon$, then $0 < |x-1| < \delta$ implies that $|f(x) - 2| < \delta = \epsilon$, so $\lim_{x\to 1} f(x) = 2$. □

Proposition 8.1. *If $\lim_{x\to x_0}$ exists, then the limit is unique.*

Proof. We want to show that if $\lim_{x\to x_0} f(x) = L_1$ and $\lim_{x\to x_0} f(x) = L_2$, then $L_1 = L_2$. Given some $\epsilon > 0$, choose a $\delta_1 > 0$ such that $0 < d_M(x, x_0) < \delta \implies d_N(f(x), L_1) < \frac{\epsilon}{2}$. Similarly, choose $\delta_2 > 0$ such that $0 < d_M(x, x_0) < \delta \implies d_N(f(x), L_2) < \frac{\epsilon}{2}$. We know that these δ_1, δ_2 exist by the definition of limit.

Now, let $\delta = \min\{\delta_1, \delta_2\}$. Then, if $0 < d_M(x, x_0) < \delta$, then we know that

$$\begin{aligned}d_N(L_1, L_2) &\leq d_N(L_1, f(x)) + d_N(f(x), L_2) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon\end{aligned}$$

□

Example 8.2. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ where $A = \mathbb{R} \setminus \{0\}$ and $f(x) = \sin\left(\frac{1}{x}\right)$. Then, $\lim_{x\to 0} f(x)$ does not exist.*

Proof. Look at $x_k = \frac{1}{k\pi}$ with $k \in \mathbb{N}$. Then, $\{x_k\} \to 0$ as $k \to \infty$ and $f(x_k) = \sin(k\pi) = 0$ for each k. But, if $y_k = \frac{1}{2\pi k + \pi/2}$, then $f(y_k) = 1$ for $k \in \mathbb{N}$ and then $\left\{y_k\right\} \to 0$. □

8.2 Continuity

Definition 8.2. *Let $f : A \subseteq M \to N$ and $x_0 \in A$. We say that f is continuous at x_0 if given any $\epsilon > 0$, $\exists \delta > 0$ such that, if $x \in A$ satisfies $d_M(x, x_0) < \delta$, then we have $d_N(f(x), f(x_0)) < \epsilon$.*

Note that here we do not insist that $0 < d_M(x, x_0)$. We then say f is continuous on A if f is continuous.

Example 8.3. *Consider $f : [0, \infty) \subseteq \mathbb{R} \to \mathbb{R}$ given by $f(x) = \sqrt{x}$. We claim that f is continuous on $(0, \infty)$.*

Proof. Pick $x_0 \in (0, \infty)$. Then, let $\epsilon > 0$ be given.

Sidework: we want to look at $|f(x) - f(x_0)|$ where $x \geq 0$. So,

$$\begin{aligned} |f(x) - f(x_0)| &= |\sqrt{x} - \sqrt{x_0}| \\ &= \left| \frac{(\sqrt{x} - \sqrt{x_0})(\sqrt{x} + \sqrt{x_0})}{\sqrt{x} + \sqrt{x_0}} \right| \\ &= \left(\frac{1}{\sqrt{x} + \sqrt{x_0}} \right) |x - x_0| \end{aligned}$$

so then we know that $\sqrt{x} + \sqrt{x_0} \geq \sqrt{x_0} > 0$, so $\frac{1}{\sqrt{x}+\sqrt{x_0}} < \frac{1}{\sqrt{x_0}}$.

Now, choose $\delta = \delta(\epsilon)$ as[3] $\delta = \min\{x_0, \epsilon\sqrt{x_0}\}$, then we get

$$|f(x) - f(x_0)| < \frac{1}{\sqrt{x} + \sqrt{x_0}} (\epsilon\sqrt{x_0}) < \epsilon$$

$\square$

Proposition 8.2. *Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be linear. Then, T is continuous.*

Proof. Pick $x_0 \in \mathbb{R}^m$ and let $\epsilon > 0$ be given. Then,

$$\begin{aligned} \|Tx - Tx_0\| &= \|T(x - x_0)\| \\ &\leq \|T\| \|x - x_0\| \end{aligned}$$

where $\|T\| = \sup\limits_{\|y\| \leq 1} \|Ty\|$, which also happens to be equal to the largest singular value, as we learned in Math 25a. So, if $\delta = \frac{\epsilon}{1+\|T\|}$, then we know that $\|x - x_0\| = \delta \implies \|Tx - Tx_0\| < \epsilon$, so T is continuous. $\square$

Proposition 8.3. *Let $f : A \subseteq M \to N$ with (M, d_M) and (N, d_N) metric spaces. Then, f is continuous on A if and only if for every convergent sequence $\{x_k\} \to x_0$ in A, we have $\left\{f(x_k)\right\} \to f(x_0)$ in N.*

Proof. Pick any $x_0 \in A$ and assume that $\{x_k\} \to x_0$. Then, let $\epsilon > 0$ be given. Choose $\delta > 0$ such that if $x \in A$, then $d_N(f(x_k), f(x_0)) < \epsilon$. Then, since $x_k \to x_0$ in A, we can choose $K \in \mathbb{N}$ such that $d_M(x_k, x_0) < \delta, \forall k \geq K$. Then, for $k \geq K$, $d_N(f(x_k), f(x_0)) < \epsilon$, so $\left\{f(x_k)\right\} \to f(x_0)$. $\square$

[3] The $\delta(\epsilon)$ notation just helps to reinforce that our choice of δ is dependent on our choice of ϵ.

9 February 11

Last time, we introduced continuity and provided the $\epsilon - \delta$ definition, as well as an alternative characterization using sequences. Now, we will state another way of characterizing continuity using sets.

9.1 Continuity and Open Sets

In the following propositions and proofs, the term "relative to A" is used, where A is a set in a metric space. Later, Wes and Taeuk circulated notes demonstrating that subsets of metric spaces can be considered metric spaces in their own right, inheriting their metrics from their parent spaces and eliminating the "relative" problem.

Proposition 9.1. *Let $f : A \subseteq M \to N$. Then, f continuous on A if and only if, for every open set $\Omega \subseteq N$, the preimage $f^{-1}(\Omega) = \left\{x \in A : f(x) \in \Omega\right\}$ is open relative to A. That is, $f^{-1}(\Omega) = A \cap U$ where $U \subseteq M$ is open. (This also works if we replace "open" with "closed.")*

Proof. ($\Longrightarrow$) Assume f is continuous on A and let $\Omega \subset N$ be open. We'll argue that $f^{-1}(\Omega)$ is open. Pick $x_0 \in f^{-1}(\Omega)$. We want a $\delta > 0$ such that $B(x_0, \delta) \subseteq f^{-1}(\Omega)$.

Now, choose an $\epsilon > 0$ such that $B(f(x_0), \epsilon) \subseteq \Omega$. By the continuity of f, $\exists \delta > 0$ such that if $x \in A$ is in $B(x_0, \delta)$, then $f(x) \in B(f(x_0), \epsilon)$. We now claim that $B(x_0, \delta) \subseteq f^{-1}(\Omega)$. Here, note that $f(B(x_0, \delta)) \subseteq B(f(x_0), \epsilon) \subseteq \Omega$. Now, repeat this for all x_0 and union them (note that arbitrary unions of open sets are open). $\square$

Proposition 9.2. *Let $f : A \subseteq M \to N$ and $g : B \subseteq N \to P$ be continuous with M, N, P all metric spaces. Now, assume $f(A) \subseteq B$. Then, $g \circ f : A \subseteq M \to P$ is continuous.*

Proof. We will use the previous proposition. We essentially want to demonstrate that, if $\Omega_1 \subseteq P$ is open, then $(g \circ f)^{-1}(\Omega_1)$ is open. Since g continuous, we know that $g^{-1}(\Omega_1) = \Omega_2$ is open in B. Then, since f continuous, $f^{-1}(\Omega_2) = \Omega_3$ is similarly open in A. Then, $(g \circ f)^{-1}(\Omega_1) = \Omega_3$, an open set. $\square$

Side remark: Now, let $f : A \subseteq M \to N$ and $g : A \subseteq M \to N$. If N is a normed vector space, we have access to sums and scalar multiples. If $d_N(y_1, y_2) = \|y_1 - y_2\|$, then we can define $(f+g) : A \subseteq M \to N$ and $(cf)x = cf(x)$. Then, if f and g continuous on A, so are cf and $f + g$. If N has some product, we can also show that fg is also continuous, etc.

This finding is incredibly powerful because it allows us to show that functions created from continuous building blocks are also continuous.

9.2 Continuity and Compact Sets

Proposition 9.3. *If $f : M \to N$ is continuous and M and N are metric spaces and K is compact, then $f(K)$ is also compact.*

Proof. This proof is extremely straightforward if we invoke Bolzano-Weierstrass, which tells us that compact is equivalent to sequentially compact. So, we will show that $f(K)$ is sequentially compact.

Let $\left\{y_k\right\}$ be a sequence in $f(K)$. Then, we know that $y_m = f(x_m)$ for some $x_m \in K$. So, $\{x_m\}$ is a sequence in K, which is sequentially compact, so we can choose a subsequence $\left\{x_{m_l}\right\}_{l=1}^{\infty}$ that converges to $x_0 \in K$. By continuity of f, we know that $\left\{f(x_{m_l})\right\}_{l=1}^{\infty} \to f(x_0)$, where $y_{m_l} = f(x_{m_l}$. Then, since $x_0 \in K$, $f(x_0) \in f(K)$, so $f(K)$ is sequentially compact. $\square$

9.3 Continuity and Connected Sets

Proposition 9.4. *If $f : A \subseteq M \to N$ is continuous and $A \subseteq M$ is connected, then $f(A)$ is also connected.*

Proof. We will proceed by contraposition. Suppose $f(A)$ is not connected. Choose separation with open sets Ω_1 and Ω_2 in N. Then, we know that

1. $\Omega_1 \cup \Omega_2 \supseteq f(A)$
2. $\Omega_1 \cap f(A) \neq \emptyset$ and $\Omega_2 \cap f(A) \neq \emptyset$
3. $\Omega_1 \cap \Omega_2 \cap f(A) = \emptyset$.

Now, create a separation of A using $f^{-1}(\Omega_1)$ and $f^{-1}(\Omega_2)$, which are open by continuity of f. So, then

1. $f^{-1}(\Omega_1) \cup f^{-1}(\Omega_2) = f^{-1}(\Omega_1 \cup \Omega_2)$. But $\Omega_1 \cup \Omega_2 \supseteq f(A)$, so $f^{-1}(\Omega_1 \cup \Omega_2) \supseteq A$.
2. $A \cap f^{-1}(\Omega_1) = f^{-1}(f(A)) \cap f^{-1}(\Omega_1) = f^{-1}(f(A) \cap \Omega_1) \neq \emptyset$ and similarly $A \cap f^{-1}(\Omega_2) \neq \emptyset$.
3. $A \cap f^{-1}(\Omega_1) \cap f^{-1}(\Omega_2) = f^{-1}(f(A) \cap \Omega_1 \cap \Omega_2) = f^{-1}(\emptyset) = \emptyset$.

so then $f^{-1}(\Omega_1)$ and $f^{-1}(\Omega_2)$ do indeed form a separation, so A is not connected. □

Example 9.1. *If $f : \mathbb{R}^m \to \mathbb{R}^n$ is continuous, images of open sets need not be open. Consider, for instance, $f(x) = 0, \forall x \in \mathbb{R}^m$. Then, the image of any non-empty open set is not open. Images of compact and connected sets are compact and connected, respectiely.*

Definition 9.1. *If $f : M \to N$ is a continuous bijection with a continuous inverse, then f is a homomorphism and M and N are homeomorphic .*

10 February 14

10.1 Extreme and Intermediate Value Theorems

Theorem 10.1. *(Extreme Value Theorem). Let (M,d) be a metric space and suppose $f : A \subseteq M \to \mathbb{R}$ is continuous. If $K \subseteq A$ is nonempty and compact, then $f(K)$ is bounded. Moreover, $\exists x, y \in K$ such that $f(x) = \inf f(K)$ and $f(y) = \sup f(K)$.*

Proof. Since K is compact and f is continuous, $f(K)$ is a compact subset of $\mathbb{R}$. So, $f(K)$ is closed and bounded (and non-empty since $K \neq \emptyset$). Since $f(K)$ is closed, $\sup f(K) \in f(K)$. So, $\exists y \in K$ such that $f(y) = \sup f(K)$.

The proof to show that $\exists x \in K$ such that $f(x) = \inf f(K)$ is similar. □

Definition 10.1. *Using the same notation as above, $f(x)$ is called the absolute minimum of f on K and $f(y)$ is the absolute maximum .*

Example 10.1. *Consider $f : K \subseteq \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x,y) = \frac{y}{x^2+1}$ on the domain $K = \left\{(x,y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq 4\right\}$. Then, we can show that f is continuous and that K is compact, so f achieves an absolute maximum and absolute minimum somewhere on K.*

Theorem 10.2. *(Intermediate Value Theorem). Let $f : A \subseteq M \to \mathbb{R}$ be continuous. If $K \subseteq A$ is connected and $x, y \in K$, then, for every $c \in \mathbb{R}$ such that $f(x) < c < f(y)$, $\exists z \in K$ such that $f(z) = c$.*

Proof. We will proceed by contradiction. Suppose that for some $c \in \mathbb{R}$ with $f(x) < c < f(y)$ there is no $z \in K$ with $f(z) = c$. Then, let $\Omega_1 = (-\infty, c)$ and $\Omega_2 = (c, \infty)$. Then,

1. Both sets are open.
2. Their union covers $f(K)$.
3. $f(x) \in \Omega_1$ so $\Omega_1 \cap f(K) \neq \emptyset$, and similarly $\Omega_2 \cap f(K) \neq \emptyset$.
4. $\Omega_1 \cap \Omega_2 \cap f(K) = \emptyset$.

so Ω_1 and Ω_2 form a separation of $f(K)$. But, f continuous should imply that $f(K)$ is connected since K is connected. ↯ □

Example 10.2. *Show that there exists a $x \in R$ such that $x = \cos x$.*

Proof. Consider $g(x) = x - \cos x$. Then, $x = \cos x$ has a solution if and only if $g(x)$ has a root. But, we know that

$$g(0) = -1 \quad \text{and} \quad g\left(\frac{\pi}{2}\right) = \frac{\pi}{2}$$

and we know that g is continuous on the connected domain $\left[0, \frac{\pi}{2}\right]$. Then, since $-1 < 0 < \frac{\pi}{2}$, we know that $\exists z \in \left[0, \frac{\pi}{2}\right]$ such that $g(z) = 0$. □

10.2 Uniform Continuity

Suppose $f : A \subseteq M \to N$ with (M, d_M) and (N, d_N) being metric spaces. Then, we said that f is continuous on A if $\forall x_0 \in A$, given $\epsilon > 0$, $\exists \delta = \delta(\epsilon, x_0) > 0$ such that if $x \in A$ and $d_M(x_0, x) < \delta$, then $d_N(f(x_0), f(x)) < \epsilon$. But, what if we can find a δ that doesn't depend upon x_0?

Definition 10.2. *If $f : A \subseteq M \to N$ as above, then f is uniformly continuous on A if given any $\epsilon > 0$, $\exists \delta = \delta(\epsilon) > 0$ such that if $x_0, x \in A$, then $d_M(x, x_0) < \delta \implies d_N(f(x), f(x_0)) < \epsilon$.*

Example 10.3. *Consider $f : A \subseteq \mathbb{R} \to \mathbb{R}$ where $A = \left[\frac{1}{100}, 1\right]$ and $f(x) = \frac{1}{x}$. Then, we claim that f is uniformly continuous on this domain.*

Proof. Pick x_0, x in the domain, and do the following scrap work:

$$\begin{aligned} |f(x) - f(x_0)| &= \left|\frac{1}{x} - \frac{1}{x_0}\right| \\ &= \left|\frac{x_0 - x}{xx_0}\right| \\ &= \frac{1}{xx_0}|x_0 - x| \end{aligned}$$

but since $x, x_0 \geq \frac{1}{100}$, we know that $\frac{1}{xx_0} \leq 100^2$. So, given ϵ, choose $\delta = \frac{\epsilon}{100^2}$. Then, regardless of x, x_0, we know that $|x - x_0| < \delta \implies |f(x) - f(x_0)| < \epsilon$. □

Note here that, if we change the domain to $(0, 1]$, f is continuous but not uniformly continuous.

Proposition 10.1. *Let $f : A \subseteq M \to N$. Then, f is not uniformly continuous on A if and only if $\exists \epsilon > 0$ and two sequences $\{x_k\}$ and $\left\{y_k\right\}$ in A such that $\left\{d_M(x_k, y_k)\right\}_{k=1}^{\infty} \to 0$ as $k \to \infty$ but $d_N(f(x_k), f(y_k)) \geq \epsilon, \forall k \in \mathbb{N}$.*

Proof. ($\Longrightarrow$) Assume that f is not uniformly continuous. Then, $\exists \epsilon > 0$ such that $\forall \delta > 0$, $\exists x, y \in A$ with $d_M(x, y) < \delta$ but $d_N(f(x), f(y)) \geq \epsilon$.

Then, for each $k \in \mathbb{N}$, let $\delta_k = \frac{1}{k}$ and choose x_k, y_k such that $d_M(x_k, y_k) < \delta_k$ but $d_N(f(x_k), f(y_k)) \geq \epsilon$. Then, we know that $\left\{d_M(x_k, y_k)\right\}_{k=1}^{\infty} \to 0$ as $k \to \infty$ but $d_N(f(x_k), f(y_k)) \geq \epsilon$. □

Example 10.4. *Let $x_k = \frac{1}{k}$ and $y_k = \frac{1}{k+1}$. Then, $d_M(x_k, y_k) = \frac{1}{k(k+1)} \to 0$ as $k \to \infty$ but $|f(x_k) - f(y_k)| = 1$. Then, if we choose $\epsilon = \frac{1}{2}$, the proposition above implies that f is not uniformly continuous.*

Proposition 10.2. *If $f : A \subseteq M \to N$ is continuous and $K \subseteq A$ is compact, then f is uniformly continuous on K.*

Proof. We will proceed by contradiction. Suppose f is continuous on K but not uniformly. Since it is not uniformly continuous, the preceding proposition implies that $\exists \epsilon > 0$ and two sequences $\left\{x_j\right\}$ and $\left\{y_j\right\}$ with $\left\{d_M(x_j, y_j)\right\} \to 0$ as $j \to \infty$. Then, we can assume that $d_M(x_j, y_j) < \frac{1}{j}$ but $d_N(f(x_j), f(y_j)) \geq \epsilon$ for each j.

Now, since K is compact, we know that it is also sequentially compact. So, we can choose a convergent subsequence $\left\{x_{j_l}\right\}_{l=1}^{\infty} \to x$ in K. But, $\left\{y_{j_l}\right\}_{l=1}^{\infty} \to x$ as $l \to \infty$ because $d_M(y_{j_l}, x) \leq d_M(y_{j_l}, x_{j_l}) + d_M(y_{j_l}, x)$, which we can make as small as we want through our choice of a large l.

But, $\left\{f(x_{j_l})\right\}_{l=1}^{\infty} \to f(x)$ and $\left\{f(y_{j_l})\right\}_{l=1}^{\infty} \to f(x)$ contradicts the statement that $d_N(f(x_{j_l}), f(y_{j_l})) \geq \epsilon$ for each $l \in N$. $\lightning$ □

11 February 16

11.1 Differentiability

Definition 11.1. *Let $f : \Omega \subseteq \mathbb{R} \to \mathbb{R}$, where Ω is open. Then, we say that f is differentiable at $x_0 \in \Omega$ if $\lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0}$ exists. If it does, we denote this limit as $f'(x_0)$ and call it the derivative of f at x_0.*

Note that, if we allow $h = x - x_0$, then $f'(x_0)$ exists if and only if $\lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$ exists.

Example 11.1. *Consider $f : \Omega \in \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \sqrt{x}$ on $\Omega = (0, \infty)$. Then, given $x_0 \in \Omega$, we have*

$$\lim_{x\to x_0} \frac{\sqrt{x}-\sqrt{x_0}}{x-x_0} = \lim_{x\to x_0} \frac{1}{\sqrt{x}+\sqrt{x_0}} = \frac{1}{2\sqrt{x_0}}$$

which exists, so $f'(x_0)$ also exists.

Example 11.2. *Show that $f(x) = \sqrt[3]{x}$ on $\mathbb{R}$ is not differentiable at 0.*

Proof. We have

$$\lim_{h\to 0} \frac{f(h)-f(0)}{h} = \lim_{h\to 0} \frac{1}{h^{2/3}}$$

which doesn't exist, so f is not differentiable at 0. □

Proposition 11.1. *If f is differentiable at $x_0 \in \Omega$, then f is continuous at x_0. Note the converse of this proposition is false.*

Proof. If $f'(x_0) = \lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0}$ exists, we want to show that $\lim_{x\to x_0} f(x) = f(x_0)$. So,

$$\begin{aligned}
\lim_{x\to x_0} f(x) &= \lim_{x\to x_0} \left[\frac{f(x)-f(x_0)}{x-x_0} \cdot (x-x_0) + f(x_0)\right] \\
&= \lim_{x\to x_0} \left[f'(x_0) \cdot 0 + f(x_0)\right] \\
&= f(x_0)
\end{aligned}$$

□

Now, we introduce an ϵ-δ definition of differentiability. From our previous definition and the definition of a limit, we know that f is differentiable if $\forall \epsilon > 0$, $\exists \delta(\epsilon, x_0) > 0$ such that $0 < |x - x_0| < \delta$ and $x \in \Omega$, then

$$\left|\frac{f(x)-f(x_0)}{x-x_0} - f'(x_0)\right| < \epsilon$$

which we can rewrite as $|f(x) - f(x_0) - f'(x_0)(x-x_0)| < \epsilon(x - x_0)$ which gives us our definition:

$$0 < |x-x_0| < \delta \implies |f(x) - f(x_0) - f'(x_0)(x-x_0)| < \epsilon(x-x_0)$$

Proposition 11.2. *(Sum Rule). If f and g differentiable at x_0, then $(f+g)'(x_0) = f'(x_0) + g'(x_0)$.*

Proof. Let $\epsilon > 0$ be given. Then, choose $\delta_1 > 0$ such that if $x \in \Omega$ and $|x-x_0| < \delta_1$ then $|f(x)-f(x_0)-f'(x_0)(x-x_0)| \le \frac{\epsilon}{2}|x-x_0|$. Similarly, choose $\delta_2 > 0$ such that if $x \in \Omega$ and $|x-x_0| < \delta_2$ then $|g(x)-g(x_0)-g'(x_0)(x-x_0)| \le \frac{\epsilon}{2}|x-x_0|$.

Then, letting $\delta = \min\{\delta_1, \delta_2\}$, if $|x - x_0| < \delta$, then

$$\begin{aligned}
diff &= |(f+g)(x) - (f+g)(x) - (f'(x_0) + g'(x_0))(x - x_0)| \\
&= |f(x) + g(x) - f(x) - g(x) - f'(x_0)(x - x_0) - g'(x_0)(x - x_0)| \\
&\leq |f(x) - f(x) - f'(x_0)(x - x_0)| + |g(x) - g(x) - g'(x_0)(x - x_0)| \\
&\leq \frac{\epsilon}{2}|x - x_0| + \frac{\epsilon}{2}|x - x_0| \\
&= \epsilon|x - x_0|
\end{aligned}$$

□

Proposition 11.3. *(Product/Leibniz Rule).* $(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$

Proof. We will use the limit definition of the derivative.

$$\begin{aligned}
\lim_{x \to x_0} \frac{(fg)(x) - (fg)(x_0)}{x - x_0} &= \lim_{x \to x_0} \frac{f(x)g(x) - f(x_0)g(x_0)}{x - x_0} \\
&= \lim_{x \to x_0} \frac{f(x)g(x) - f(x)g(x_0) + f(x)g(x_0) - f(x_0)g(x_0)}{x - x_0} \\
&= \lim_{x \to x_0} \left[g(x) \cdot \frac{f(x) - f(x_0)}{x - x_0} + f(x) \cdot \frac{g(x) - g(x_0)}{x - x_0} \right] \\
&= g(x_0) \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} + f(x_0) \lim_{x \to x_0} \frac{g(x) - g(x_0)}{x - x_0} \\
&= f'(x_0)g(x_0) + f(x_0)g'(x_0)
\end{aligned}$$

□

Proposition 11.4. *(Chain Rule). Let $f : \Omega_1 \subseteq \mathbb{R} \to \mathbb{R}$ and $g : \Omega_2 \subseteq \mathbb{R} \to \mathbb{R}$ and $f(\Omega_1) \subseteq \Omega_2$ and Ω_1, Ω_2 open. Then, if f is differentiable at $x_0 \in \Omega_1$ and g is differentiable at $f(x_0) \in \Omega_2$, then $(g \circ f)'(x_0) = g'(f(x_0))f'(x_0)$.*

Note that f being differentiable does not necessarily imply that f' is continuous.

Example 11.3. *Let $f : \mathbb{R} \to \mathbb{R}$ be defined by*

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0. \end{cases}$$

Then, f' exists for $x \neq 0$. But, we claim that $f'(0)$ also exists.

Proof. $\lim_{h \to 0} \frac{f(h) - f(0)}{h} = \lim h \to 0 \frac{h^2 \sin\left(\frac{1}{h}\right)}{h} = \lim_{h \to 0} h \sin\left(\frac{1}{h}\right) = 0$, where the last step is true by the squeeze theorem because $-|h| \leq h \sin\left(\frac{1}{h}\right) \leq |h|$. □

11.2 Optimization Terminology

Definition 11.2. *Let $f : \Omega \subseteq \mathbb{R} \to \mathbb{R}$, where Ω is open. Then, f is increasing on an interval $I \subseteq \Omega$ (as long as I is not a singleton) if for $x, y \in I$, $x < y \implies f(x) \leq f(y)$.*

Definition 11.3. *Using the same terminology, if $x < y \implies f(x) < f(y)$, then f is strictly increasing in I.*

Definition 11.4. *Also, f has a local maximum at $x_0 \in \Omega$ if $\exists$an open interval $(x_0 - \epsilon, x + \epsilon)$ such that if x is in the interval, then $f(x_0) \geq f(x)$.*

Example 11.4. *If we use a modified version of f from Example 1.3, where*

$$f(x) = \begin{cases} \frac{x}{2} + x^2 \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0. \end{cases}$$

then $f'(0) = \frac{1}{2} + 0 = \frac{1}{2}$. So, there is no open interval containing 0 *in which f is increasing.*

12 February 18

12.1 Mean Value Theorem

Proposition 12.1. *Let $f : \Omega \subseteq \mathbb{R} \to \mathbb{R}$, where Ω is open. Then, if f has a local extremum at $x_0 \in \Omega$ and $f'(x_0)$ exists, then $f'(x_0) = 0$.*

Proof. Let us proceed by contradiction. For a local maximum (the case in which we handle a local minimum is analogous), suppose that $f'(x_0)$ exists but $f'(x_0) > 0$. By definition of differentiability at x_0, given an $\epsilon > 0$, $\exists \delta > 0$ such that

$$0 < |x - x_0| < \delta \implies \left| \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right| < \epsilon.$$

Now, choose $\epsilon = \frac{1}{2} f'(x_0) > 0$. Then, $\frac{f(x)-f(x_0)}{x-x_0} - f'(x_0) > -\epsilon = -\frac{1}{2} f'(x_0)$, which implies that $\frac{f(x)-f(x_0)}{x-x_0} > \frac{1}{2} f'(x_0)$. If $x > x_0$ and $|x - x_0| > \delta$, then $f(x) > f(x_0) + \frac{1}{2} f'(x_0)(x - x_0)$, which is positive. $\lightning$ $\square$

Not that, if $f'(x_0) = 0$, x_0 is not necessarily a local extremum. Simply consider the case of $f(x) = x^3$ at $x = 0$.

Corollary 12.1.1. *Let $f : I \subseteq \mathbb{R} \to \mathbb{R}$, where I is an interval. Then, if f has a local extremum at some interior point $c \in I$, then either $f'(c) = 0$ or $f'(c)$ doesn't exist.*

Proposition 12.2. *(Rolle's Theorem) Suppose that $f : [a, b] \to \mathbb{R}$ is continuous and differentiable on (a, b), and that $f(a) = f(b) = 0$. Then $\exists c \in (a, b)$ such that $f'(c) = 0$.*

Proof. If f is identically 0 on $[a, b]$, then we can choose any $c \in (a, b)$. Otherwise, assume WLOG that f takes on some positive values in (a, b). Then, by the Extreme Value Theorem, since f is continuous and $[a, b]$ is compact, f achieves its absolute maximum at some interior point $c \in (a, b)$. Since c is a local maximum and f is differentiable at c, we can say that $f'(c) = 0$. $\square$

Theorem 12.1. *(Mean Value Theorem). If $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), then $\exists$ some $c \in (a, b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$. Note that c is not necessarily unique.*

Proof. Let $L(x) = f(a) + \frac{f(b)-f(a)}{b-a} \cdot (x - a)$, which we know is differentiable on (a, b). Now, apply Rolle's Theorem to $g(x) = f(x) - L(x)$. $\square$

Corollary 12.2.1. *If f is continuous on $[a, b]$ and differentiable on (a, b) and $f'(x) = 0$ on (a, b), then f is constant on $[a, b]$.*

Proof. Pick any $x \in [a, b]$. We want to show that $f(x) = f(a)$. If $x = a$, then obviously $f(x) = f(a)$. Otherwise the Mean Value Theorem on $[a, x]$. Then, $\exists c \in (a, x)$ such that $f'(c) = \frac{f(x)-f(a)}{x-a}$, and we know that $f'(c) = 0$ by hypothesis, which implies that $f(x) = f(a)$. $\square$

Corollary 12.2.2. *Assume that f and g are continuous on $[a, b]$ and differentiable on (a, b). If $f'(x) = g'(x), \forall x \in (a, b)$, then $f(x) = g(x) + c$.*

Proof. Apply the previous corollary to $f - g$. $\square$

12.2 First Derivative Test

Proposition 12.3. *Assume f is continuous on $[a,b]$ and f is differentiable on (a,b). Then, f is increasing on $[a,b] \iff f'(x) \geq 0$ on (a,b). Similarly, f is decreasing on $[a,b] \iff f'(x) \leq 0$ on (a,b).*

Proof. ($\Longleftarrow$) Suppose that $f'(x) \geq 0$ on (a,b). We want to show that, given $x_1, x_2 \in [a,b]$ with $x_1 < x_2$, we have $f(x_1) < f(x_2)$.

By MVT on $[x_1,x_2]$, $\exists c \in (x_1,x_2)$ with $\frac{f(x_2)-f(x_1)}{x_2-x_1} = f'(c) \geq 0$. Since $x_2 - x_1 > 0$, then $f(x_2) - f(x_1) \geq 0$. ☐

Theorem 12.2. *(First Derivative Test) Suppose f continuous on $[a,b]$. Assume $c \in (a,b)$ and f differentiable on (a,c) and (c,b). If $\exists \delta > 0$ such that*

1. $(c-\delta, c+\delta) \subseteq (a,b)$
2. $f'(x) \geq 0$ *on* $(c-\delta, c)$
3. $f'(x) \leq 0$ *on* $c, c+\delta)$

then f has a local maximum at c.

Proof. Pick $x \in (c-\delta, c)$. By Prop 12.3, f is increasing on $(c-\delta, c)$, so $f(x) \leq f(c), \forall x \in (c-\delta, c)$. Likewise, $f(x) \geq f(c), \forall x \in (c, c+\delta)$. ☐

Note that the converse is not necessarily true.

Example 12.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be defined as*

$$f(x) = \begin{cases} 2x^4 + x^4 \sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0. \end{cases}$$

Then, f has an absolute minimum at 0 but f' takes positive and negative values in every open interval containing 0.

12.3 Construction of the Riemann Integral (Part I)

We will now construct the Riemann integral for bounded, real-valued functions on closed, bounded domains in $\mathbb{R}$.

Definition 12.1. *Let $I = [a,b]$ be a closed bounded interval with $a < b$. A partition of I is a finite ordered list $x_0, x_1, \ldots, x_n$ of points in I with $a = x_0 < x_1 < \cdots < x_n = b$.*

Let $f : I \to \mathbb{R}$ be a bounded function and P a partition with the above notation. Then, P partitions $[a,b]$ into subintervals $[x_0,x_1], [x_1,x_2], \ldots, [x_{n-1},x_n]$. For $k = 1, 2, \ldots, n$, we define

$$m_k = \inf\left\{f(x) : x \in [x_{k-1}, x_k]\right\}$$
$$M_k = \sup\left\{f(x) : x \in [x_{k-1}, x_k]\right\}$$

Definition 12.2. *We say that the upper and lower sum of f corresponding to P are, respectively,*

$$U(f,p) = \sum_{k=1}^{n} M_k(x_k - x_{k-1})$$

$$L(f,p) = \sum_{k=1}^{n} m_k(x_k - x_{k-1})$$

13 February 23

13.1 Construction of the Riemann Integral (Part II)

Let $I = [a,b] \subset \mathbb{R}$, $a < b$. Recall that a partition P of I is a finite, ordered list $a = x_0 < x_1 < x_2 < \cdots < x_n = b$. Let $f : [a,b] \to \mathbb{R}$ be bounded and a partition (using the above notation). Then, let

$$m_k = \inf\Big\{f(x) : x \in [x_{k-1}, x_k]\Big\}$$
$$M_k = \sup\Big\{f(x) : x \in [x_{k-1}, x_k]\Big\}$$

We previously defined the upper and lower sums as

$$U(f,P) = \sum_{k=1}^{n} M_k(x_k - x_{k-1})$$
$$L(f,P) = \sum_{k=1}^{n} m_K(x_k - x_{k-1})$$

Certainly then, $L(f,P) \le U(f,P)$, since $m_k \le M_k$ and $(x_k - x_{k-1}) > 0$ for each $k = 1, 2, \ldots, n$.

Definition 13.1. *If $P = \{x_0, x_1, \ldots, x_n\}$ is a partition of $I = [a,b]$, then another partition $Q = \{\widetilde{x_0}, \widetilde{x_1}, \ldots, \widetilde{w_n}\}$ is called a refinement of P if $P \subset Q$.*

Then, if Q is a refinement of P, then $L(f,P) \le L(f,Q)$, and $U(f,P) \ge U(f,Q)$. This is easy to see if we consider adding a single point to P.

If we include a new point $x_{k+1/2}$ to P then clearly we have

$$\sup_{x \in [x_k, x_{k+1/2}]} f(x) \le \sup_{x \in [x_k, x_{k+1}]} f(x)$$

and the same is true for the latter subinterval. Additionally, let

$$m = \inf\Big\{f(x) : x \in [a,b]\Big\} \text{ and } M = \sup\Big\{f(x) : x \in [a,b]\Big\}.$$

Then, for every partition P,

$$m(b-a) \le L(f,P) \le U(f,P) \le M(b-a)$$

Finally, let $f : [a,b] \to \mathbb{R}$ be bounded and let P, Q be any two partitions of $[a,b]$. Then, $L(f,P) \le U(f,Q)$.

Let $R = P \cup Q$. Then, R is a refinement of both P and Q, so we have

$$L(f,P) \le L(f,R) \le U(f,R) \le U(f,Q)$$

Definition 13.2. *If $f : (a,b] \to \mathbb{R}$ is bounded, then let $\mathcal{P}$ denote the set of all partitions of $[a,b]$. Then, the lower integral of f over $[a,b]$ is*

$$\underline{\int_a^b} f(x)\,\mathrm{d}x = \sup_{P \in \mathcal{P}}\Big\{L(f,P)\Big\}$$

and similarly the upper integral is

$$\overline{\int_a^b} f(x)\,\mathrm{d}x = \inf_{P \in \mathcal{P}}\Big\{U(f,P)\Big\}.I$$

These infinums and supremums exist by the Axiom of Completeness: we know that the sets are bounded, nonempty subsets of $\mathbb{R}$.

Proposition 13.1. *If $f : [a,b] \to \mathbb{R}$ is bounded, then $\underline{\int_a^b} f(x)\,\mathrm{d}x \leq \overline{\int_a^b} f(x)\,\mathrm{d}x$.*

Proof. If P and Q are partitions of $[a,b]$, then we know that $L(f,P) \leq U(f,Q)$. Then, since p is arbitrary, we know that

$$\sup_{P\in\mathcal{P}} \left\{L(f,P)\right\} \leq U(f,Q)$$

But Q is arbitrary, so

$$\sup_{P\in\mathcal{P}} \left\{L(f,P)\right\} \leq \inf_{Q\in\mathcal{P}} \left\{U(f,P)\right\}$$

□

Definition 13.3. *If $f : [a,b] \to \mathbb{R}$ is bounded, then we say that f is Riemann integrable on $[a,b]$ if $\underline{\int_a^b} f(x)\,\mathrm{d}x = \overline{\int_a^b} f(x)\,\mathrm{d}x$. Then, we write $\int_a^b f(x)\,\mathrm{d}x$ as the common value. Also, we set*

$$\int_a^a f(x)\,\mathrm{d}x = 0 \quad \text{and} \quad \int_b^a f(x)\mathrm{d}x = -\int_a^b f(x)\,\mathrm{d}x.$$

Example 13.1. *If $f : [a,b] \to \mathbb{R}$ is constant and we say $f(x) = c, \forall x$. Then, $\int_a^b f(x)\,\mathrm{d}x = c(b-a)$ because $U(f,P) = c(b-a)$ and we have $L(f,P) = c(b-a)$, $\forall P \in \mathcal{P}$.*

14 February 25

Last time, we said that $\overline{\int_a^b} f = \inf_{P \in \mathcal{P}} \Big\{ U(f,P) \Big\}$ and $\underline{\int_a^b} f = \sup_{P \in \mathcal{P}} \Big\{ L(f,P) \Big\}$ where $f : [a,b] \to \mathbb{R}$ is bounded. We also said that f was integrable if $\underline{\int_a^b} f = \overline{\int_b^a} f$.

14.1 Integrability and Continuity

Example 14.1. *Consider the restriction of the Direchlet function, $f : [0,1] \to \mathbb{R}$ defined as*

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

Then, this function is not Riemann integrable, because, regardless of our partition P,

$$L(f,P) = 0 \quad \text{and} \quad U(f,P) = 1$$

so the upper and lower integrals do not agree.

Example 14.2. *Consider $f : [0,1] \to \mathbb{R}$ defined*

$$f(x) = \begin{cases} 6 & x \geq \frac{1}{2} \\ 0 & x < \frac{1}{2} \end{cases}$$

so then consider the partition

$$P_\epsilon = \left\{ 0, \frac{1}{2} - \frac{\epsilon}{12}, \frac{1}{2} + \frac{\epsilon}{12}, 1 \right\}$$

which has

$$L(f,P_\epsilon) = 3 - \frac{\epsilon}{2} \quad \text{and} \quad U(f,P_\epsilon) = 3 + \frac{\epsilon}{2}$$

so then

$$\sup_{P \in \mathcal{P}} \Big\{ L(f,P) \Big\} \geq 3 \quad \text{and} \quad \inf_{P \in \mathcal{P}} \Big\{ U(f,P) \Big\} \leq 3$$

so we get $\int_0^1 f = 3$.

Proposition 14.1. *(Riemann). Let $f : [a,b] \to \mathbb{R}$ be bounded. Then, f is integrable on $[a,b]$ if and only if, given $\epsilon > 0$, $\exists$ partition P_ϵ of $[a,b]$ such that $U(f,P_\epsilon) - L(f,P_\epsilon) < \epsilon$.*

Proposition 14.2. *Suppose $f : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$. Then, f is integrable on $[a,b]$.*

Proof. f is continuous on the compact set $[a,b]$, so f is uniformly continuous. Given $\epsilon > 0$, we can choose a $\delta > 0$ such that $|x-y| < \delta$, then $|f(x) - f(y)| < \frac{\epsilon}{b-a}$. Choose N large enough such that $\frac{b-a}{N} < \delta$. Then, look at the partition $P = \{x_0, x_1, \ldots, x_N\}$ where $x_0 = a$ and $x_j = j\left(\frac{b-a}{N}\right) + a$.

Look at $[x_{k-1}, x_k]$ with length less than δ. Since compact, f achieves supremum M_k and infimum m_k on this subinterval. Since the length of the subinterval is under δ, we get $M_k - m_k < \frac{\epsilon}{b-a}$. Then, $U(f,P) = \sum_{k=1}^N M_k(x_k - x_{k-1})$ and $L(f,P) = \sum_{k=1}^N m_k(x_k - x_{k-1})$, so then

$$\begin{aligned} U(f,P) - L(f,P) &= \sum_{k=1}^N (M_k - m_k)(x_k - x_{k-1}) \\ &< \frac{\epsilon}{b-a} \sum_{k=1}^N (x_k - x_{k-1}) \\ &= \frac{\epsilon}{b-a}(b-a) = \epsilon \end{aligned}$$

□

Note that:

1. If f, g integrable on $[a,b]$ and $c \in \mathbb{R}$, we can show that $f+g$ and cf are integrable on $[a,b]$ and that $\int_a^b f + \int_a^b g$ and that $\int_a^b cf = c\int_a^b f$.
2. If f, g integrable on $[a,b]$ and $f(x) \leq g(x), \forall x \in [a,b]$, then $\int_a^b f \leq \int_a^b g$.
3. If f is integrable on $[a,r]$ and on $[r,b]$, then it is also integrable on $[a,b]$ and $\int_a^b f = \int_a^r f + \int_r^b f$.
4. If $|f|$ and f integrable on $[a,b]$, then $|\int_a^b f| \leq \int_a^b |f|$.

Theorem 14.1. *(FTC). Suppose $f : [a,b] \to \mathbb{R}$ is continuous. Then, f has an antiderivative F and $\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a)$.*

Proof. Claim F is defined by $F(x) = \int_a^x f(t)\,\mathrm{d}t$ is the antiderivative for t on (a,b) if $x_0 \in (a,b)$.

Then, consider

$$\lim_{x \to x_0} \frac{F(x) - F(x_0)}{x - x_0}.$$

Then, we have, for instance

$$\begin{aligned} F(x) - F(x_0) &= \int_a^x f(t)\,\mathrm{d}t - \int_a^{x_0} f(t)\,\mathrm{d}t \\ &= \int_{x_0}^x f(t)\,\mathrm{d}t \approx f(x_0)(x - x_0). \end{aligned}$$

□

14.2 Creating Transcendental Functions

How do we define e^2, $\cos x$, and similar functions? We will define these in terms of other functions, but before doing so, we need a way of assigning meaning to series of functions.

Let A be a set, and (N, d_N) a metric space.

Definition 14.1. *A sequence $\{f_k\}$ of functions $f_k : A \to N$ converges pointwise to $f : A \to N$ if given any $x \in A$, we have $\{f_k(x)\} \to f(x)$. That is, $\forall \epsilon > 0$, $\forall x \in A$, $\exists C = C(\epsilon, x)$ such that $k \geq C$ implies that $d_N(f_k(x), f(x)) < \epsilon$. We also say that $\{f_k\}$ converges uniformly to f if $\forall \epsilon > 0$, $\exists C = C(\epsilon)$ such that $k \geq C$ implies $d_N(f_k(x), f(x)) < \epsilon$, $\forall x \in A$.*

Obviously, uniform convergence implies pointwise converge.

Example 14.3. *Consider $f_k : [0,1] \to \mathbb{R}$ where $k \in \mathbb{N}$ defined as $f_k(x) = x^k$. Then,*

$$\lim_{k \to \infty} f_k(x) = \begin{cases} 1 & x = 1 \\ 0 & x \in [0,1). \end{cases}$$

If $f(x) = \lim_{k \to \infty} f_k(x)$, then $f_k \to f$ pointwise on $[0,1]$, but not uniformly.

For instance, let $\epsilon = \frac{1}{2}$. Then, consider $f_k(x)$ where $x = \sqrt[k]{0.9} \in [0,1)$, but $f_k(x) = 0.9 > \epsilon$.

Also note that, even though f_k is continuous for each k, f is not.

Proposition 14.3. *Let $f_k : A \subseteq M \to P$ where (M, d_M) and (P, d_P) are metric spaces. Assume that f_k is continuous for each k and that they converge uniformly to f. Then, f is continuous on A.*

Proof. Let $\epsilon > 0$ be given. Then, choose $N \in \mathbb{N}$ such that $\forall n \geq N$, $d_P(f_n(x), f(x)) < \frac{\epsilon}{3}$, $\forall x \in A$. Now, pick any $x_0 \in A$ and we will argue that f is continuous at x_0.

Note that
$$d_P(f(x), f(x_0)) \leq d_P(f(x), f_N(x)) + d_P(f_N(x), f_N(x_0)) + d_P(f_N(x_0), f(x_0)).$$
But, by the continuity of f, we can choose $\delta > 0$ such that $d_M(x, x_0) < \delta \implies d_P(f_N(x), f_N(x_0)) < \frac{\epsilon}{3}$, so then
$$\begin{aligned} d_P(f(x), f(x_0)) &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ &= \epsilon \end{aligned}$$

□

15 February 28

15.1 Cauchy Criterion and Weierstrass M-Test

Recall that a sequence $\left\{f_k\right\}$ of functions $f_k : A \to N$ converges pointwise to $f : A \to N$ if, given any $x \in A$, we have $\left\{f_k(x)\right\} \to f(x)$. We say that $\left\{f_k\right\}$ converges uniformly to f is $\forall \epsilon > 0$, $\exists C = C(\epsilon)$ such that, if $k \geq C$, then $d_N(f_k(x), f(x)) < \epsilon$, $\forall x \in A$.

Example 15.1. *The sequence $\left\{f_k\right\}$ where $f_k(x) = x^k$ does not converge uniformly (which is clear if we look at values very close to* 1*, as demonstrated in Ex. 14.3).*

For series of functions $\sum_{k=1}^{\infty} f_k$, we say that the series converges pointwise (respectively uniformly) if the sequence of partial sums converges pointwise (respectively uniformly) to f on A.

Example 15.2. *The series $\sum_{k=0}^{\infty}(-1)^k x^{2k}$ is geometric for each x on $(-1,1)$, so the series converges pointwise to $\frac{1}{1+x^2}$. If we look at the nth partial sum, S_n, we see that*

$$\begin{aligned} S_n(x) &= 1 - x^2 + x^4 + \cdots + (-1)^{n-1}x^{2(n-1)} \\ x^2 S_n(x) &= x^2 - x^4 + \cdots - (-1)^{n-1}x^{2n} \end{aligned}$$

so then

$$(1+x^2)S_n(x) = 1 - (-1)^{n-1}x^{2n}$$

and then we have

$$|S_n(x) - \frac{1}{1+x^2}| = \frac{x^{2n}}{1+x^2} \leq x^{2n} \to 0$$

as $n \to \infty$ for $x \in (-1,1)$.

Generally, however, we do not have a nice form for S_n, so we will need a better test for uniform convergence.

Proposition 15.1. *Let $f_k : A \to N$ where (N, d_N) is a complete metric space. Then, $\left\{f_k\right\}$ converges uniformly to f on A if and only if it satisfies the Cauchy criterion: that $\forall \epsilon > 0$, $\exists C \in \mathbb{N}$ such that if $m, n \geq C$, then $d_N(f_m(x), f_n(x)) < \epsilon$, $\forall x \in A$.*

Proof. ($\Longrightarrow$) Given some $\epsilon > 0$, $\exists C \in \mathbb{N}$ such that if $n \geq C$, $d_N(f_n(x), f(x)) < \frac{\epsilon}{2}$, $\forall x \in A$. Then, if $m, n \geq C$, $d_N(f_m(x), f_n(x)) \leq d_N(f_m(x)) + d_N(f_n(x), f(x)) < \epsilon$. □

Theorem 15.1. *(Weierstrass Test) If N is a complete normed vector space and $g_k : A \to N$ are functions such that $\exists$ constraints M_k with $\|g_k(x)\| \leq M_k$, $\forall x \in A$ and $\sum_{k=1}^{\infty} M_k$ converges, then $\sum_{k=1}^{\infty} g_k$ converges absolutely and uniformly on A. That is, $\forall x \in A$, $\sum_{k=1}^{\infty} \|g_k(x)\|$ converges.*

Proof. Since $\sum_{k=1}^{\infty} M_k$ is convergent, the sequence $\sigma_n = \sum_{k=1}^{n} M_k$ of partial sums converges as $n \to \infty$, so $\{\sigma_n\}$ is Cauchy. So, given $\epsilon > 0$, $\exists C \in \mathbb{N}$ such that $m, n \geq C \implies |\sigma_n - \sigma_M| < \epsilon$. Then, WLOG assume that $n \geq M$ and then write $n = m + p$. So, we get

$$|\sigma_n - \sigma_m| = |M_{m+1} + M_{m+2} + \cdots + M_{m+p}| < \epsilon$$

if $m \geq C$. Now, look at the partial sums $S_n = \sum_{k=1}^{n} g_k$. Then, for each $x \in A$,

$$\begin{aligned} \|S_n(x) - S_m(x)\| &= \|g_{m+1}(x) + \cdots + g_{m+p}(x)\| \\ &= \|g_{m+1}(x)\| + \cdots + \|g_{m+p}(x)\| \\ &= M_{m+1} + \cdots + M_{m+p} < \epsilon \end{aligned}$$

so then $\{S_n\}_{n=1}^{\infty}$ is Cauchy in N, and if N is complete, then $\left\{S_n(x)\right\}$ converges. The convergence is also uniform because ϵ was independent of x. □

Example 15.3. *Consider the series* $\sum_{k=0}^{\infty} \frac{x^k}{k!}$*. The series converges uniformly to a continuous function* $f : \mathbb{R} \to \mathbb{R}$ *on any interval* $[a,b] \subseteq \mathbb{R}$*.*

Proof. Pick $x_0 \in \mathbb{R}$ and consider $I = \{x \in \mathbb{R} : |x| \le 2|x_0| + 1\}$. Then, $f_k(x)$ is continuous on I, and

$$|f_k(x)| \le \frac{(2|x_0| + 1)^k}{k!} = M_k$$

and we know that $\sum_{k=0}^{\infty} M_k$ converges by the Ratio Test. Then, by Weierstrass M-test, $\sum_{k=0}^{\infty} \frac{x^k}{k!}$ converges absolutely and uniformly on I. And we know that f is continuous, because it has continuous partial sums and the sequence converges uniformly. □

However, it should be noted that, even though $\{f_n\} \to f$ uniformly and each f_n is differentiable, $\{f_n'\} \to f'$ isn't necessarily uniform.

Example 15.4. *Take, for instance, the sequence* $\{f_n\}$ *where* $f_n(x) = \frac{x^{n+1}}{n+1}$ *on* $(0,1)$*. Then,* $\{f_n\} \to 0 = f(x)$ *uniformly on* $(0,1)$ *but the sequence* $\{f_n'\}$ *where* $f_n'(x) = x^n$ *doesn't converge uniformly to* $f'(x) = 0$*.*

Example 15.5. *Let* $f_k : A \subseteq M \to N$ *where* (M, d_M) *and* N *are normed, vector spaces. If each* f_k *is continuous on* A *and* $\sum_{k=0}^{\infty} f_k$ *converges uniformly to* f *on* A*, then* f *is continuous on* A*, which gives us the following: if* $x_0 \in A$*, then*

$$\begin{aligned}
\lim_{x \to x_0} \sum_{k=1}^{\infty} f_k(x) &= \sum_{k=1}^{\infty} \lim_{x \to x_0} f_k(x) \\
&= \sum_{k=1}^{\infty} f_k(x_0) \\
&= f(x_0).
\end{aligned}$$

But this raises some questions:

1. Would this work if the series converged pointwise but not uniformly?
2. If $f_k : (a,b) \to \mathbb{R}$ is differentiable and $\{f_k\} \to f$ pointwise (or uniformly), is f differentiable? Does $\{f_k'\}$ converge to f' in some way?
3. What about integration?

15.2 Interchangeability of Limits and Integrals

Proposition 15.2. *Suppose* $\{f_k\}$ *where* $f_k : [a,b] \to \mathbb{R}$ *is integrable and bounded. If* $\{f_k\} \to f$ *uniformly on* $[a,b]$*, then*

$$\lim_{n \to \infty} \int_a^b f_n = \int_a^b \lim_{n \to \infty} f_n = \int_a^b f$$

Proof. We will begin by showing that f is integrable. By the Riemann criterion, f is integrable if and only if, given $\epsilon > 0$, $\exists$ a partition P of $[a,b]$ such that $U(f,P) - L(f,P) < \epsilon$. Now, let $\epsilon > 0$ be given. Since $\{f_k\} \to f$ uniformly, we can choose $N \in \mathbb{N}$ such that $n \ge N$ implies that

$$|f_n(x) - f(x)| < \frac{\epsilon}{4(b-a)}, \forall x \in [a,b].$$

Now, consider f_N, which we know is integrable. By the Riemann criterion, we can choose a partition P such that

$$U(f_N, P) - L(f_N, P) < \frac{\epsilon}{2}.$$

Then, let us estimate $U(f,P)$ in terms of $U(f_N,P)$ and similarly for $L(f,P)$ and $L(f_N,P)$. Using the notation of partitions, given some interval corresponding to the partition $[x_{k-1},x_k]$, we know that, if

$$M_k(f) = \sup_{x\in[x_{k-1},x_k]} f(x) \quad \text{and} \quad M_k(f_N) = \sup_{x\in[x_{k-1},x_k]} f_N(x)$$

then since $|f_N(x) - f(x)| < \frac{\epsilon}{4(b-a)}$ for all $x \in [a,b]$, we know that $f(x) < f_N(x) + \frac{\epsilon}{4(b-a)}$, $\forall x \in [a,b]$. Then, taking the supremum over $[x_{k-1},x_k]$, we get that $M_k(f) \le M_k(f_N) + \frac{\epsilon}{4(b-a)}$. But then we know that

$$\begin{aligned} U(f,P) &= \sum_{k=1}^{m} M_k(f)\cdot(x_k - x_{k-1}) \\ &\le \sum_{k=1}^{m}\left[M_k(f_N) + \frac{\epsilon}{4(b-a)}\right]\cdot(x_k - x_{k-1}) \\ &= U(f_N,P) + \sum_{k=1}^{m}\frac{\epsilon}{4(b-a)}\cdot(x_k - x_{k-1}) \\ &= U(f_N,P) + \frac{\epsilon}{4} \end{aligned}$$

and similarly, $L(f_N,P) - \frac{\epsilon}{4} \le L(f,P)$, so

$$L(f_N,P) - \frac{\epsilon}{4} \le L(f,P) \le U(f,P) \le U(f_N,P) + \frac{\epsilon}{4}$$

which thus implies that

$$U(f,P) - L(f,P) \le U(f_N,P) - L(f_N,P) + \frac{\epsilon}{2} < \epsilon.$$

Then, if f integrable, we know that

$$\lim_{n\to\infty}\int_a^b f_n = \int_a^b \lim_{n\to\infty} f_n = \int_a^b f.$$

□

Corollary 15.2.1. *If $f_k : [a,b] \to \mathbb{R}$ are bounded and integrable, and $\sum_{k=1}^{\infty} f_k$ converges uniformly to f on $[a,b]$, then we know that*

$$\sum_{k=1}^{\infty}\int_a^b f_k = \int_a^b \sum_{k=1}^{\infty} f_k = \int_a^b f$$

Proof. Apply the above proof to the partial sums of the series. □

Note, however, that pointwise convergence does not guarantee that this integral property applies.

Example 15.6. *Look at $\left\{f_k\right\}$ where $f_k : [a,b] \to \mathbb{R}$ defined by*

$$f_k(x) = \begin{cases} 2k^2x & 0 \le x \le \frac{1}{k} \\ 4k - 2k^2x & \frac{1}{k} < x \le \frac{2}{k} \\ 0 & \frac{2}{k} < x \le 1 \end{cases}$$

Then, $\int_0^1 f_n = 2$, but $\left\{f_n\right\} \to f$ converges pointwise to $f = 0$ on $[0,1]$.

16 March 2

Recall: If $f_k : [a,b] \to \mathbb{R}$ is bounded and integrable, and $\sum_{k=1}^{\infty} f_k$ converges uniformly to f on $[a,b]$, then f is integrable and $\int_a^b \sum_{k=1}^{\infty} f_k(x)\,\mathrm{d}x = \sum_{k=1}^{\infty} \int_a^b f_k(x)\,\mathrm{d}x$. Previously, we demonstrated that pointwise convergence is not enough to justify this!

Example 16.1. *Let $0 < a < 1$ and consider the sequence $\left\{f_k\right\}$ where $f_k(x) = (-1)^k x^{2k}$. Then, we know that*

$$|f_k(x)| \leq a^{2k} = M_k$$

and that $\sum_{k=0}^{\infty} a^{2k}$ converges by the geometric series test, so by the Weierstrass-M test, $\sum_{k=0}^{\infty} f_k$ converges uniformly on $[-a,a]$ to a continuous function. We also showed last time that it converges to $f(x) = \frac{1}{1+x^2}$ on $[-a,a]$. If we pick $s \in [-a,a]$, then

$$\begin{aligned}
\int_0^s \sum_{k=0}^{\infty} (-1)^k x^{2k}\,\mathrm{d}x = \arctan s - \arctan 0 \\
&= \sum_{k=0}^{\infty} (-1)^k \int_0^s x^{2k}\,\mathrm{d}x \\
&= \sum_{k=0}^{\infty} (-1)^k \frac{s^{2k+1}}{2k+1}
\end{aligned}$$

16.1 Differentiation on Series

Example 16.2. *Consider the sequence $\left\{f_k\right\}$ where $f_k : [0,1] \to \mathbb{R}$ is defined as $f_k(x) = x^k$, which is differentiable and continuous, but the sequence converges pointwise to a discontinuous function.*

Example 16.3. *Consider the sequence $\left\{f_k\right\}$ where $f_k(x) = \frac{x^{k+1}}{k+1}$ on $[0,1]$. Then, we can show that $\left\{f_k\right\} \to 0$ uniformly. But, $\left\{f_k'\right\} = \left\{x^k\right\}$, which we have shown does not converge to 0, uniformly or pointwise.*

Proposition 16.1. *Let $f_k : (a,b) \to \mathbb{R}$ be differentiable. Then, let $\left\{f_k\right\} \to f$ pointwise on (a,b). If the derivatives f_k' are continuous and $\left\{f_k'\right\} \to g$ uniformly, then f is differentiable, and $f' = g$.*

Proof. If $\left\{f_k'\right\} \to g$ uniformly on (a,b), then g is continuous, since each f_k' is continuous and the sequence converges uniformly.

Now, pick a $x_0 \in (a,b)$. Then, for each $x \in (a,b)$, we know that

$$\int_{x_0}^{x} f_k'(s)\,\mathrm{d}s \to \int_{x_0}^{x} g(s)\,\mathrm{d}s$$

as $k \to \infty$. Then, if let $k \to \infty$ and use pointwise convergence of $\left\{f_k\right\}$ to f, we find that

$$\begin{aligned}
f(x) - f(x_0) &= \int_{x_0}^{x} g(s)\,\mathrm{d}s \\
f(x) &= f(x_0) + \int_{x_0}^{x} g(s)\,\mathrm{d}s
\end{aligned}$$

which then implies that $f'(x) = g(x)$ per Fundamental Theorem of Calculus. And, since g is continuous, f has a continuous derivative. $\square$

Corollary 16.1.1. *Let f_k have continuous derivatives on (a,b). If $\sum_{k=1}^{\infty} f_k$ converges pointwise to some limit, and $\sum_{k=1}^{\infty} f_k'$ converges uniformly, then*

$$\left(\sum_{k=1}^{\infty} f_k\right)' = \sum_{k=1}^{\infty} f_k'$$

Proof. Apply the previous result to the sequence of partial sums. □

16.2 The Function E(x)

Example 16.4. *Consider $\sum_{k=0}^{\infty} \frac{x^k}{k!}$. We claim that this series converges to a differentiable function on all of $\mathbb{R}$.*

Proof. If we pick $x_0 \in \mathbb{R}$, then consider the interval $I = \{x \in \mathbb{R} : |x| \leq 2|x_0| + 1\}$. Now, use Weierstrass-M test with $M_k = \frac{(2|x_0|+1)^k}{k!}$. Since $\sum_{k=1}^{\infty} M_k$ converges by the Ratio Test, we know that $\sum_{k=0}^{\infty} \frac{x^k}{k!}$ converges uniformly on I to a continuous limit function.

Then, by the Power Rule, we know that $f_k'(x) = \frac{x^{k-1}}{(k-1)!}$ for $k \geq 1$ and $f_0' = 0$ for $k = 0$. Then,

$$\sum_{k=1}^{\infty} f_k'(x) = \sum_{k=1}^{\infty} \frac{x^{k-1}}{(k-1)!} = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

which we know also converges uniformly on I. Then, the corollary we proved above implies that $\sum_{k=0}^{\infty}$ converges to f on I, that f' exists, and that $\sum_{k=0}^{\infty} f_k' = f'$.

But, since x_0 we chose above was arbitrary, we know that $\sum_{k=0}^{\infty} \frac{x^k}{k!}$ is differentiable on all of $\mathbb{R}$. □

Proposition 16.2. *There is a function $E : \mathbb{R} \to \mathbb{R}$ such that:*

1. $E'(x) = E(x), \forall x \in \mathbb{R}$
2. $E(0) = 1$.

Proof. Define $E_0(x) = 1, \forall x \in \mathbb{R}$. Then, for each $n \in \mathbb{N}$, define $E_n(x) = 1 + \int_0^x E_{n-1}(s)\,\mathrm{d}s$. We know E_0 is integrable over any bounded, closed interval, and by induction, $E_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!}$.

Note that $E_n'(x) = E_{n-1}(x), \forall n \in \mathbb{N}, x \in \mathbb{R}$. Then, given any $R > 0$, consider $\{E_n\}$ on the interval $[-R, R]$. We said $\{E_n\}$ converges uniformly, and if we use $E(x) = \lim_{n\to\infty} E_n(x)$, then we can see that $E_n(0) = 1$, so $E(0) = \lim_{n\to\infty} E_n(0) = 1$.

Also, the previous example showed that $E'(x) = E(x), \forall x \in [-R, R]$. But, given any $x_0 \in \mathbb{R}$, $\exists R$ large enough such that $x_0 \in \operatorname{int}[-R, R]$, so we can conclude that $E'(x) = E(x)$, $\forall x \in \mathbb{R}$. □

Corollary 16.2.1. *The function E we just defined has derivatives of all orders, and $E^{(n)}(x) = E(x), \forall x \in \mathbb{R}, \forall n \in \mathbb{N}$.*

Proof. Use induction on n with the previous result as the base case. □

Proposition 16.3. *The function $E : \mathbb{R} \to \mathbb{R}$ satisfying the conditions in Prop 16.2 is unique.*

Proof. We won't prove this yet, but uniqueness will follow from applying Taylor's Theorem, which we'll introduce later. □

Definition 16.1. *This function is called the (natural) exponential function , usually written $f(x) = e^x$ or $\exp(x)$, instead of $E(x)$. Also $E(1)$ is called Euler's constant , noted as e.*

Before we introduce Taylor's Theorem, we will first introduce some notation that will make it clearer:

$$C([a,b],\mathbb{R}) = \Big\{ f : [a,b] \to \mathbb{R} : f \text{ continuous on } [a,b] \Big\}$$
$$C^k([a,b],\mathbb{R}) = \Big\{ f : [a,b] \to \mathbb{R} : f^k \text{ continuous on } [a,b] \Big\}, k \geq 0.$$

where f^k is the kth derivative of f.

Theorem 16.1. *(Taylor). Let $n \in \mathbb{N}$ and suppose $f \in C^{n+1}([a,b],\mathbb{R})$. Then, if $x_0 \in [a,b]$, then, for each $x \in [a,b]$ where $x \neq x_0$, $\exists$ a point c between x_0 and x such that*

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots + \frac{f^n(x_0)}{n!}(x-x_0)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1}$$

Note that Taylor's Theorem is something of an expansion of MVT.

17 March 4

Recall. Suppose $f \in C^{(n+1)}([a,b],\mathbb{R})$. If $x_0 \in [a,b]$, then for each $x \neq x_0$ in $[a,b]$, $\exists c$ between x_0 and x such that

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \cdots + \frac{f^n(x_0)}{n!}(x-x_0)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1}$$

17.1 More Properties of E(x)

Proposition 17.1. *The exponential function $E : \mathbb{R} \to \mathbb{R}$ satisfies $E(x) > 0, \forall x$.*

Proof. Recall that $E(0) = 1$. We will proceed by contradiction. Suppose indirectly that $E(x_0) = 0$ for some $x_0 \in \mathbb{R}$. Consider $E(x)$ on the compact interval I whose endpoints are 0 and x_0. By the EVT, $\exists M > 0$ such that $|E(x)| \leq M, \forall x \in I$. Use Taylor's Theorem to say that $\exists c \in \mathrm{int}(I)$ such that

$$1 = E(0) = E(x_0) + E'(x_0)(0-x_0) + \cdots + \frac{E^{(n)}(x_0)}{n!}(0-x_0)^n + \frac{E^{(n+1)}(c)}{(n+1)!}(0-x_0)^{n+1}$$

□

But, since $E(x_0) = 0$ by assumption, and $E'(x_0) = 0, \ldots, E^{(n)}(x_0) = 0$, we know that

$$1 = \frac{E^{(n+1)}(c)}{(n+1)!}(-x_0)^{n+1}$$

and applying absolute value to both sides we get

$$1 = \left|\frac{E^{(n+1)}(c)}{(n+1)!}\right| \left|(-x_0)^{n+1}\right| = \left|\frac{E^{(n+1)}(c)}{(n+1)!}\right| |x_0|^{n+1}$$

So then,

$$1 \leq \frac{M}{(n+1)!}|x_0|^{n+1}, \forall n \in \mathbb{N}$$

But, the right-hand side of this expression goes to 0 as $n \to \infty$, which forms a contradiction, so there does not exist $x \in \mathbb{R}$ such that $E(x) = 0$

Then, since $E(0) = 1$, E is continuous, and $E(x) \neq 0, \forall x \in \mathbb{R}$, we can invoke IVT to conclude that $E : \mathbb{R} \to \mathbb{R}$ satisfies $E(x) > 0, \forall x$.

Proposition 17.2. $E(x+y) = E(x)E(y), \forall x, y \in \mathbb{R}$

Proof. Fix $y \in \mathbb{R}$. Then, consider $q(x) = \frac{E(x+y)}{E(y)}$, which works because $E(y) \neq 0$ from above. Then,

$$q'(x) = \frac{E'(x+y)}{E(y)} = \frac{E(x+y)}{E(y)} = q(x)$$

But if $q'(x) = q(x)$ and $q(0) = 1$, by the uniqueness of the exponential function, know that $q(x) = E(x)$. □

Before the next proof, note that we say that $\lim_{x\to\infty} f(x) = L$ if and only if $\forall \epsilon > 0, \exists R \in \mathbb{R}$ such that $x > R \implies |f(x) - L| < \epsilon$. Similarly, $\lim_{x\to\infty} f(x) = \infty$ if and only if for each $M \in \mathbb{R}$, $\exists R \in \mathbb{R}$ such that $x > R \implies f(x) > M$.

Proposition 17.3. $\lim_{x\to\infty} E(x) = \infty$ *and* $E(x) = 0$.

Proof. Partial proof: we know that $E(m) = e^m > 2^m$ and that $E(-m) = e^{-m} < 2^{-m}$, from which we can show convergence to each limit. □

17.2 Function Spaces

Consider functions $f : A \subseteq M \to N$ where (M,d) is a metric space, and N is a normed real vector space with the norm $\|\cdot\|$. Then, if f and g are such functions, and λ is a scalar, then $f+g$ and λf are defined as

$$(f+g)(x) = f(x)+g(x) \quad \text{and} \quad (\lambda f)(x) = \lambda f(x).$$

We will also use the notation

$$C(A,N) = \left\{f : A \subseteq M \to N : f \text{ continuous on } A\right\}$$

which is a vector space with addition and scalar multiplication as defined above. Note, however, that we cannot use the supremum norm, defined as

$$\|f\|_\infty = \sup\left\{\|f(x)\| : x \in A\right\}$$

because functions need not be bounded on A. But, if we instead consider the subspace

$$C_B(A,N) = \left\{f : A \subseteq M \to N : f \text{ bounded and continuous on } A\right\}$$

then C_B is a normed vector space with the above supremum norm. Here, you might also notice that, if A is compact, then by EVT, $C_B(A,N) = C(A,N)$.

But do the suprema of this norm make sense? To show this, we will first demonstrate the following:

Proposition 17.4. $\|\cdot\| : N \to \mathbb{R}$ *is continuous.*

Proof. Pick some $x_0 \in N$. Then, $\forall \epsilon > 0$, we want to show that $\exists \delta > 0$ such that

$$\|x - x_0\| < \delta \implies \left|\|x\| - \|x_0\|\right| < \epsilon$$

By the Triangle Inequality, we know that

$$\begin{aligned} \|x\| &= \|x - x_0 + x_0\| \\ \|x\| &\leq \|x - x_0\| + \|x_0\| \\ \|x\| - \|x_0\| &\leq \|x - x_0\| \end{aligned}$$

so then, setting $\delta = \epsilon$, we have that, if $\|x - x_0\| < \delta$, then

$$\left|\|x\| - \|x_0\|\right| \leq \|x - x_0\| < \delta = \epsilon.$$

□

Now, since this is continuous, this makes sense because we map each compact set in M to a compact set in N to a compact set in $\mathbb{R}$, which must contain a supremum by EVT.

Definition 17.1. *If N is a complete normed vector space, then $C_B(A,N)$ is complete, and we call it a Banach space . Similarly, a complete inner product space is called a Hilbert space .*

Q: What do open, closed, and compact subsets of $C_B(A,N)$ look like?

Example 17.1. *Consider the nonempty and compact subset $K \subseteq \mathbb{R}^m$ and the corresponding space $C(K,\mathbb{R}^m)$. Then, there is no need for C_B since $C(K,\mathbb{R}^m) = C_B(K,\mathbb{R}^m)$ because K is compact. Then, the zero function is $C(K,\mathbb{R}^m)$ is just 0, and*

$$B(0,\epsilon) = \left\{f \in C(K,\mathbb{R}^n) : \|f\|_\infty < \epsilon\right\}.$$

Then, we claim that $B(0,\epsilon)$ is open.

Proof. To prove that $B(0,\epsilon)$ is open, we must prove that given some $g \in B(0,\epsilon)$, $\exists \eta > 0$ such that $B(g,\eta) \subset B(0,\epsilon)$. We know that $\|g\|_\infty < \epsilon$, so let $\|g\|_\infty = M$.

Now, let $\eta = \epsilon - M$. Then, $B(g,\eta) \subseteq B(0,\epsilon)$ because given some $h \in B(g,\eta)$, we know that

$$\begin{aligned}
\|h\|_\infty &= \|h - g + g\|_\infty \\
&\leq \|h - g\|_\infty + \|g\|_\infty \\
&< \epsilon - M + M = \epsilon
\end{aligned}$$

□

Example 17.2. *Now, consider the closed unit ball in* $C([0,1],\mathbb{R})$ *defined as*

$$\left\{ f \in C([0,1],\mathbb{R}) : \|f\|_\infty \leq 1 \right\}.$$

Then, this closed unit ball is not compact.

Proof. We will demonstrate that it is not sequentially compact. Consider the sequence $\left\{f_k\right\}$ where $f_k : [0,1] \to \mathbb{R}$ is defined $f_k(x) = x^k$. Then, we have showed that $\left\{f_k\right\}$ converges pointwise to f where

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

so $\exists \left\{f_k\right\}$ such that no subsequence converges to a continuous function. Then, because it is not sequentially compact, it is not compact. □

18 March 7

18.1 Equicontinuity and the Arzelà-Ascoli Theorem

Definition 18.1. *We say $\mathcal{B}$ is equicontinuous if $\forall \epsilon > 0$, $\exists \delta > 0$ such that if $x, y \in A$ and $d_M(x,y) < \delta$, then $\|f(x) - f(y)\| < \epsilon, \forall f \in \mathcal{B}$.*

Note that some texts might call this "uniform equicontinuity." We'll use this concept of equicontinuity to formula the closest analog of Heine-Borel for $C(A,N)$.

Example 18.1. *Consider the closed unit ball in $C([0,1],\mathbb{R})$. If $f \in C([0,1],\mathbb{R})$, use the sup norm as*

$$\|f\|_\infty = \sup_{x\in[0,1]} |f(x)|.$$

Then, the closed unit ball is

$$\left\{f \in C([0,1],\mathbb{R}) : \|f\|_\infty \leq 1\right\}$$

which is not compact because it is not sequentially compact: consider $\left\{f_k\right\}$ where $f_k(x) = x^k$.

Definition 18.2. *We say that $\mathcal{B}$ is pointwise bounded if for each fixed $x \in A$, the set $\left\{f(x) : f \in \mathcal{B}\right\}$ is bounded.*

Theorem 18.1. *(Arzelà-Ascoli) Let $\mathcal{B} \in C(K,\mathbb{R}^n)$, where K is a compact subset of $\mathbb{R}^m$. If $\mathcal{B}$ is equicontinuous and pointwise bounded, then every sequence in $\mathcal{B}$ has a uniformly convergent subsequence.*

Note that if $\left\{f_n\right\}$ is an equicontinuous and pointwise bounded sequence in $\mathcal{B}$, then the closure of $\left\{f_1, f_2, \ldots\right\}$ is compact.

Example 18.2. *Suppose $\left\{f_n\right\}$ is a sequence of functions $f_n : [a,b] \to \mathbb{R}$ uniformly bounded by M and whose derivatives exist on (a,b) and are uniformly bounded by M'. We claim that $\left\{f_n\right\}$ has a uniformly convergent subsequence.*

Proof. We are already given that the points in the set are pointwise bounded by M, $\forall n \in \mathbb{N}$, $x \in [a,b]$. So, we need only check equicontinuity of $\mathcal{B}$. Use the fact that $|f_n'(x)| \leq M'$ for each $x \in (a,b)$ and $\forall n \in \mathbb{N}$. By MVT,

$$|f_n(x) - f_n(y)| = |f_n'(c)||x - y|$$

for some c between x and y, but the derivative is uniformly bounded by M', so we can say instead

$$|f_n(x) - f_n(y)| \leq M'|x - y|$$

so, given $\epsilon > 0$, if we use $\delta = \frac{\epsilon}{M'+1}$, then if $x, y \in A$ and $|x - y| < \delta$, then $|f_n(x) - f_n(y)| < \epsilon$, $\forall n \in \mathbb{N}$, which shows equicontinuity. The result then follows by applying Arzelà-Ascoli. $\square$

18.2 Contraction Maps

Last time, we encountered the initial value problem $E'(x) = E(x)$, $E(0) = 1$. Now, we will set up a theorem that will help us deal with other initial value problems. In the follow examples, let (M,d) be a metric space.

Definition 18.3. *Let $\Phi : M \to M$. Then Φ is a contraction if $\exists c \in \mathbb{R}$ with $0 \leq c < 1$ such that $d(\Phi(x),\Phi(y)) \leq cd(x,y)$, $\forall x, y \in M$.*

Note that, if M is a normed vector space and Φ is a contraction map, then $\|\Phi(x) - \Phi(y)\| \leq \|x - y\|$, $\forall x, y \in M$, so Φ is Lipschitz with a constant $c < 1$.

Theorem 18.2. *(Banach) Contraction Mapping Principle. If M is a complete metric space and $\Phi : M \to M$ is a contraction, then Φ has a unique fixed point. That is, $\exists x^* \in M$ such that $\Phi(x^*) = x^*$.*

Proof. We will first show the existence of a fixed point and then show uniqueness. Pick any $x_0 \in M$. Then, define a recursive sequence where $x_n = \Phi(x_{n-1})$ for each $n \in \mathbb{N}$ if $x_1 = \Phi(x_0) = x_0$, then x_0 was a fixed point. Assume instead that $x_1 \neq x_0$. Choose $c \in [0,1)$ such that $d(\Phi x, \Phi y) \leq cd(x,y)$, which is possible by contraction map. Then, $d(x_1, x_2) \leq cd(x_0, x_1)$. By induction, $d(x_n, x_{n+1}) \leq c^n d(x_0, x_1)$.

Now, we claim the sequence $\{x_n\}$ is Cauchy. Pick $p \in \mathbb{N}$. Then,

$$\begin{aligned}
d(x_n, x_{n+p}) &\leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \cdots + d(x_{n+p-1}, x_{n+p}) \\
&\leq c^n d(x_0, x_1) + \cdots + c^{n+p-1} d(x_0, x_1) \\
&= c^n (1 + c + \cdots + c^{p-1}) d(x_0, x_1)
\end{aligned}$$

but then $1 \leq 1 + c + \cdots c^{p-1} \leq \frac{1}{1-c}$ where the last inequality is true because the left side is a partial sum of the geometric series, whose full sum is on the right.

Now, given $\epsilon > 0$, we can choose $M \in \mathbb{N}$ such that $c^M \frac{1}{1-c} d(x_0, x_1) < \epsilon$. Then, $d(x_n, x_{n+p}) < \epsilon$ whenever $n \geq M$. So, the sequence is Cauchy. Then, since M is complete, $\{x_n\}$ is also convergent. Let x^* be the limit.

We now claim that x^* is a fixed point of Φ. By definition,

$$x^* = \lim_{n\to\infty} x_n = \lim_{n\to\infty} \Phi(x_{n-1}).$$

Recall that Φ is uniformly continuous because it is Lipschitz (with constant c). Then, we can say that

$$\lim_{n\to\infty} \Phi(x_{n-1}) = \Phi\left(\lim_{n\to\infty} x_{n-1}\right) = \Phi(x^*)$$

which demonstrates that x^* is a fixed point.

Now, we will show that x^* is unique. If x^{**} is also a fixed point,

$$\begin{aligned}
d(\Phi(x^*), \Phi(x^{**})) &= d(x^*, x^{**}) \\
d(x^*, x^{**}) &\leq cd(x^*, x^{**}) \\
(1-c)d(x^*, x^{**}) &\leq 0
\end{aligned}$$

which implies that $d(x^*, x^{**}) = 0$, so $x^* = x^{**}$, so there must only be one unique fixed point. ☐

Example 18.3. *Consider the nonlinear equation $x = \frac{\cos x}{2}$. We claim there is a unique solution. Let $\Phi(x) = \frac{1}{2}\cos x$. Then, Φ is a contraction because we can choose $c = \frac{1}{2}$ as our Lipschitz constant – we know that $\frac{1}{2}$ is an upper bounded on $|\Phi'(x)| = \frac{1}{2}|\sin x|$. Then, by Banach's theorem, $\exists! x^*$ such that $x^* = \Phi x^*$.*

19 March 9

19.1 Picard-Lindelöf Theorem

Recall that we gave an example of an initial value problem (that is, a differential equation with one initial condition) which does not have a unique solution:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = 3x^{2/3} \qquad x(0) = 0.$$

Example 19.1. *Even if an initial value problem has a unique solution, it may not exist for a long time. Consider the problem*

$$\frac{\mathrm{d}x}{\mathrm{d}t} = x^2 \qquad x(0) = 1 \implies x(t) = \frac{1}{1-t}$$

but there exists an asymptote at $t = 1$.

More generally, consider initial value problems of the form $x'(t) = F(x(t))$, $x(0) = b$, where $F : \Omega \subseteq \mathbb{R} \to \mathbb{R}$, where Ω is open and $b \in \Omega$. We will refer to this general system as $(*)$. Does this initial value problem have a unique solution on some interval $\alpha < t < \beta$ that contains $t = 0$?

Let us reformulate $(*)$ is terms of an integral. If $x \in C'((\alpha,\beta),\Omega)$ is a solution of $(*)$, then we can write

$$\int_0^t x'(s)\,\mathrm{d}s = x(t) - x(0) \implies x(t) = b + \int_0^t F(x(s))\,\mathrm{d}s, \forall t \in (\alpha,\beta)$$

But if $x \in C^0((\alpha,\beta),\Omega)$ and $x(t) = b + \int_0^t F(x(s))\,\mathrm{d}s$ and F is continuous, by FTC, x is C^1 and x satisfies $(*)$ on (α,β).

Suppose $F : \Omega \subseteq \mathbb{R} \to \mathbb{R}$, where Ω is open and $b \in \Omega$.

Theorem 19.1. *(Picard-Lindelöf) If F is locally Lipschitz on Ω, then $\exists \eta > 0$ and a unique C^1 function $x : (-\eta,\eta) \to \Omega$ such that $x' = F(x)$ and $x(0) = b$.*

Proof. We'll reformulate $(*)$ is terms of an integral, set up a contraction on some appropriate Banach space, and then use the contraction mapping principle.

Step 1: Since Ω is open and f is locally Lipschitz and $b \in \Omega$, we can choose a $\delta > 0$ such that $[b-\delta, b+\delta] \subset \Omega$ and $F|_{[b-\delta,b+\delta]}$ is Lipschitz with constant L. Now, for some $\eta > 0$ and some δ, define

$$S_\eta = \left\{x \in C([-\eta,\eta],\mathbb{R}) : |x(t) - b| \leq \delta, \forall t \in [-\eta,\eta]\right\}$$

with the sup norm on S_η.

Step 2: Now, reformulate $(*)$ in terms of an integral. As demonstrated earlier, we want a function x that satisfies $x(t) = b + \int_0^t F(x(s))\,\mathrm{d}s$. Then, define a function $\Phi : S_\eta \to C([-\eta,\eta],\mathbb{R})$ where $\Phi(x(t)) = b + \int_0^t F(x(s))\,\mathrm{d}s$ for $t \in [-\eta,\eta]$. By definition of S_η, $|x(s) - b| \leq \delta$, so $x(s) \in [b-\delta, b+\delta] \subseteq \Omega$, so the integral makes sense. Then, a fixed point of Φ would satisfy the integral formulation of $(*)$.

Step 3: We claim that, if η is small enough, then $\Phi : S_\eta \to S_\eta$. Pick $x \in S_\eta$. How do we guarantee that $\Phi(x) \in S_\eta$? We know

$$\begin{aligned}
\|\Phi(x) - b\|_\infty &= \sup_{t\in[-\eta,\eta]} \left|\Phi(x)(t) - b\right| \\
&= \sup_{t\in[-\eta,\eta]} \left|\int_0^t F(x(s))\,\mathrm{d}s\right|
\end{aligned}$$

the value of which we hope to get under δ. Now, if we let $I(t) = [\min\{0,t\}, \max\{0,t\}]$, then the above expression becomes

$$\|\Phi(x) - b\|_\infty \leq \sup_{t\in[-\eta,\eta]} \int_{I(t)} |F(x(s))|\,\mathrm{d}s$$

But $x(s) \in I(t) \subseteq [-\eta, \eta]$ so $x(s) \in [b-\delta, b+\delta]$. But f is continuous on $[b-\delta, b+\delta]$ so, by EVT, $\exists M > 0$ with $|F(z)| \leq M$, $\forall z \in [b-\delta, b+\delta]$. So then

$$\sup_{t\in[-\eta,\eta]} |t|M = \eta M \leq \delta$$

so we can choose $0 < \eta < \frac{\delta}{M}$ so that $\Phi : S_\eta \to S_\eta$.

Step 4: Assuming $\eta \leq \frac{\delta}{M}$, we can further restrict η such that Φ is a contraction on S_η. We know that $F|_{[b-\delta,b+\delta]}$ is Lipscihtz, so pick $x, y \in S_\eta$ such that

$$\begin{aligned}
\|\Phi(x) - \Phi(y)\|_\infty &= \sup_{t\in[-\eta,\eta]} \left|\Phi(x)(t) - \Phi(y)(t)\right| \\
&= \sup_{t\in[-\eta,\eta]} \left|\int_0^t F(x(s)) - F(y(s))\,\mathrm{d}s\right| \\
&\leq \sup_{t\in[-\eta,\eta]} \int_{I(t)} \left|F(x(s)) - F(y(s))\right|\mathrm{d}s \\
&\leq \sup_{t\in[-\eta,\eta]} \int_{I(t)} L|x(s) - y(s)|\,\mathrm{d}s \\
&\leq \sup_{t\in[-\eta,\eta]} |t|L\|x-y\|_\infty \\
&= \eta L\|x-y\|_\infty.
\end{aligned}$$

So, if we insist that $0 < \eta \leq \min\left\{\frac{1}{L+1}, \frac{\delta}{M}\right\}$, then Φ is a contraction on S_η. But we can also show that S_η is a closed subset of the Banach space $C([-\eta,\eta],\mathbb{R})$, so S_η is also a Banach space, which allows us to invoke the contraction mapping principle. Thus, Φ has a unique fixed point, which then means that there is a unique C^1 function $x : (-\eta, \eta) \to \Omega$ such that $x' = F(x)$ and $x(0) = b$. $\square$

Note that η is constructed in the proof in a way that is generally way more restrictive than necessary.

20 March 11

20.1 Derivatives for Functions of Higher Dimensions

Recall that if $f : \Omega \subseteq \mathbb{R} \to \mathbb{R}$, with Ω open and $x_0 \in \Omega$, then we said f is differentiable at x_0 if

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

exists, which is equivalent to the existence of a number $f'(x_0)$ such that

$$\lim_{x \to x_0} \left| \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} \right| = 0$$

Then, if $L(x)$ is the linearization where $L(x) = f(x_0) + f'(x_0)(x - x_0)$, then the above limit is the same as

$$\lim_{x \to x_0} \frac{|f(x) - L(x)|}{|x - x_0|}$$

so the separation between the graph of f and its tangent line at $(x_0, f(x_0))$ becomes small relative to $x - x_0$ as $x \to x_0$.

Suppose now that we want to generalize this idea of differentiability to $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ where Ω is open and $x_0 \in \Omega$. We want the best linear approximation to f near x_0, so we will need a linear transformation $Df(x_0) : \mathbb{R}^n \to \mathbb{R}^m$ such that $f(x) \approx f(x_0) + Df(x_0)(x - x_0)$ near x_0.

Definition 20.1. *If $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $x_0 \in \Omega$, we say that f is differentiable at x_0 if $\exists$ a linear transformation $Df(x_0) : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\lim_{x \to x_0} \frac{\|f(x) - f(x_0) - Df(x_0)(x - x_0)\|}{\|x - x_0\|} = 0.$$

We then call $Df(x_0)$ the derivative of f at x_0.

Equivalently, f is differentiable at x_0 if $\forall \epsilon > 0$, $\exists \delta > 0$ such that if $x \in \Omega$ and $\|x - x_0\| < \delta$, then $\|f(x) - f(x_0) - Df(x_0)(x - x_0)\| \le \epsilon \|x - x_0\|$.

Proposition 20.1. *If $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x_0 \in \Omega$, then the derivative $Df(x_0) : \mathbb{R}^n \to \mathbb{R}^m$ is unique.*

Example 20.1. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and $x_0 \in \mathbb{R}^n$. We claim that $DT(x_0) = T$. We can check this by noticing that*

$$\frac{\|Tx - Tx_0 - T(x - x_0)\|}{\|x - x_0\|} = \frac{\|Tx - Tx_0 - Tx - Tx_0\|}{\|x - x_0\|} = \frac{0}{\|x - x_0\|}$$

which is equal to 0 as $x \to x_0$. By the uniqueness proposition, our claim is true.

20.2 Partial Derivatives

Definition 20.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$. Given a point $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, the partial derivative of f with respect to its jth component is*

$$\lim_{h \to 0} \frac{f(x_1, x_2, \dots, x_{j-1}, x_j + h, x_{j+1}, \dots, x_n) - f(x_1, x_2, \dots, x_n)}{h}$$

provided the limit exists.

Example 20.2. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x,y) = e^{-(x^2+y^2)}$. Then, we can calculate the partial derivatives with the following notation by regarding other variables as constant:*

$$\frac{\partial f}{\partial x} = -2xe^{-(x^2+y^2)} \qquad \frac{\partial f}{\partial y} = -2ye^{-2(x^2+y^2)}$$

so we then say that, at the point $(1,2) \in \mathbb{R}^2$,

$$\frac{\partial f}{\partial x} = -2e^{-5} \qquad \frac{\partial f}{\partial y} = -4e^{-5}$$

Warning: note that just because the partial derivatives exist does not necessarily imply that f is differentiable (or even continuous)!

Example 20.3. *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ where*

$$f(x,y) = \begin{cases} 1 & \text{if } x = 0 \text{ or } y = 0 \\ 0 & \text{otherwise} \end{cases}$$

which, graphed in 3D, essentially looks like the plane $z = 0$ with a plus cut out and elevated to $z = 1$.

If $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x_0 \in \Omega$, then the partials of f do exist at x_0, and they give us a nice way of finding the standard matrix for $Df(x_0)$. Write

$$f(x_1, x_2, \ldots, x_n) = \begin{bmatrix} f_1(x_1, x_2, \ldots, x_n) \\ \vdots \\ f_m(x_1, x_2, \ldots, x_n) \end{bmatrix}$$

and let

$$A = \Big[\ Df(x)e_1 \ \Big|\ \cdots \ \Big|\ Df(x)e_n \ \Big]_{m \times n}.$$

Then, we want to figure out how to find A_{ij}.

We know

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

In particular, suppose that $y = x + he_j$. Then, we get

$$\lim_{h \to 0} \frac{\|f(x + he_j) - f(x) - Df(x)(he_j)\|}{|h|} = 0.$$

So for the ith component of f,

$$\lim_{h \to 0} \frac{|f_i(x + he_j) - f_i(x) - A_{ij}h|}{|h|} = 0$$

So, it must be that $\frac{\partial f_i}{\partial x_j} = A_{ij}$.

Proposition 20.2. *If $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$, Ω open and $x \in \Omega$ and f differentiable at x, then $\frac{\partial f_i}{\partial x_j}$ exists and the standard matrix for $Df(x)$ is*

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

Definition 20.3. *The matrix above is also called the Jacobian of f at x. If $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, we usually write* $\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\mathrm{d} f}{\mathrm{d} x_n} \end{bmatrix}$.

Example 20.4. *If $f : \mathbb{R}^2 \to \mathbb{R}$ is $f(x,y) = 9 - x^2 - y^2$, then $\nabla f(x,y) = (-2x, -2y)$ and $\nabla f(1,2) = (-2,-4)$.*

If f is differentiable at $(1,2)$, then $(-2,-4)$ is the standard matrix for the derivative. The "best linear approximation" near that point then is

$$\begin{aligned} P(x,y) &= f(1,2) + \nabla f(1,2) \cdot \begin{bmatrix} x-1 \\ y-2 \end{bmatrix} \\ &= 4 + (-2,-4) \cdot \begin{bmatrix} x-1 \\ y-2 \end{bmatrix} \\ &= 4 - 2(x-1) - 4(y-2) \end{aligned}$$

Example 20.5. *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}^2$, which we can visualize as a vector field where*

$$\begin{bmatrix} x \\ y \end{bmatrix} \overset{f}{\longmapsto} \begin{bmatrix} x - xy \\ -y + xy \end{bmatrix}.$$

Then,

$$Df(x,y) = \begin{bmatrix} 1-y & -x \\ y & -1-x \end{bmatrix} \quad \text{and} \quad Df(0,0) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

which then helps us encode the best linear approximation near $(0,0)$ as

$$f(0,0) + Df(0,0) \begin{bmatrix} x-0 \\ y-0 \end{bmatrix} = \begin{bmatrix} x \\ -y \end{bmatrix}.$$

21 March 21

21.1 Continuity and Differentiability for Higher-Dimension Functions

Recall that, if $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ where Ω is open, we say that f is differentiable at $x_0 \in \Omega$ if $\exists$ a linear transformation $Df(x_0) : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{x \to x_0} \frac{\left\| f(x) - f(x_0) - Df(x_0)(x - x_0) \right\|}{\|x - x_0\|} = 0$$

We have shown in the past that, if $m = n = 1$, then being differentiable at x_0 implies being continuous at x_0. We have also shown more recently that, if f is differentiable on Ω, then the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist throughout Ω and the Jacobian at x_0 is the standard matrix for $Df(x_0)$.

But, can we similarly give sufficient conditions for differentiability in terms of partial derivatives? Also, does differentiability imply continuity?

Proposition 21.1. *If $f : \Omega : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable on the open set Ω, then f is continuous on Ω.*

Proof. Let $x_0 \in \Omega$ be given. Then, since f is differentiable at x_0, given some $\epsilon_0. > 0$, we can find some $\delta_0 > 0$ such that, if $\|x - x_0\| < \delta_0$ and $x \in \Omega$,

$$\begin{aligned} \left\| f(x) - f(x_0) - Df(x_0)(x - x_0) \right\| &\leq \epsilon_0 \|x - x_0\| \\ \left\| f(x) - f(x_0) \right\| - \left\| Df(x_0)(x - x_0) \right\| &\leq \epsilon_0 \|x - x_0\| \\ \left\| f(x) - f(x_0) \right\| &\leq (\epsilon_0 + \left\| Df(x_0) \right\|) \|x - x_0\| \end{aligned}$$

Then, we get continuity because given some $\epsilon > 0$, we can choose $\delta = \min\left\{\delta_0, \frac{\epsilon}{\epsilon_0 + \left\| Df(x_0) \right\|}\right\}$, which, applied to the inequality above, tells us that

$$\|x - x_0\| < \delta, x \in \Omega \implies \left\| f(x) - f(x_0) \right\| < \epsilon.$$

□

21.2 Curves in Euclidean Space

Definition 21.1. *A curve in $\mathbb{R}^n$ is a continuous function $\gamma : [a, b] \to \mathbb{R}^n$, where $a < b$. We say that γ is closed if $\gamma(a) = \gamma(b)$. We also say that γ is simple if γ is injective on $[a, b)$ and on $(a, b]$. A simple closed curve is called a Jordan curve.*

If $\gamma : [a, b] \to \mathbb{R}^n$ is differentiable on (a, b), then $D\gamma(x)$ has a matrix representation as

$$\begin{bmatrix} \partial \gamma_1 / \partial x \\ \vdots \\ \partial \gamma_n / \partial x \end{bmatrix}$$

for each $x_0 \in (a, b)$. Here, we'll often use prime notation and write $\gamma'(t)$ instead of $D\gamma(t)$ because the notation is not ambiguous. If $\gamma'(t) \neq 0$, then we can orient the curve and sketch a special tangent line to the curve.

Example 21.1. *Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by*

$$\gamma(t) = \begin{bmatrix} \cos t \\ \sin t \\ t \end{bmatrix}.$$

If γ is differentaible, then we know that

$$\gamma(t) = \begin{bmatrix} -\sin t \\ \cos t \\ 1 \end{bmatrix},$$

so

$$\gamma'\left(2\pi + \frac{\pi}{4}\right) = \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \\ 1 \end{bmatrix}$$

as a velocity vector. We can then parameterize the line tangent to the curve at the point $\gamma\left(2\pi + \frac{\pi}{4}\right)$ in the direction of $\gamma'\left(2\pi + \frac{\pi}{4}\right)$ as

$$L(t) = \gamma\left(2\pi + \frac{\pi}{4}\right) + t\gamma'\left(2\pi + \frac{\pi}{4}\right).$$

Example 21.2. *Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ defined by $\gamma(t) = \begin{bmatrix} t^3 \\ t^2 \end{bmatrix}$. Then, γ is injective. If γ is differentiable, we know*

$$\gamma'(t) = \begin{bmatrix} 3t^2 \\ 2t \end{bmatrix} \implies \gamma'(0) = 0.$$

Note that we get $y = x^{2/3}$ but there exists some sort of lack of smoothness at $x = 0$. A similar case exists if we parameterize the line $y = x$ as $\gamma(t) = \begin{bmatrix} t^2 \\ t^2 \end{bmatrix}$.

Recall the discussion from earlier that the existence of partials alone is not enough to conclude differentiability.

Proposition 21.2. *Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$, where Ω is open. Then, assume $\frac{\partial f_i}{\partial x_j}$ where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$ exist and are continuous on Ω. Then, f is differentiable on Ω.*

Proof. We will provide an incomplete sketch of the proof, whose full version can be found in M&H. Pick $x_0 \in \Omega$. We want to show that $\exists$ a linear transformation $Df(x_0)$ such that $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$$\left\| f(y) - f(x) - Df(x)(y - x) \right\| \leq \epsilon \left\| y - x \right\|$$

whenever $y \in \Omega$ and $\left\| y - x \right\| < \delta$. We make three observations that make this proof simple:

1. It suffices to assume that $m = 1$ because, if $z \in \mathbb{R}^m$, $\|z\| \leq |z_1| + |z_2| + \cdots + |z_m|$.
2. We know that, if $Df(x)$ exists, they the Jacobian must be the standard matrix.
3. Note that

$$\begin{aligned} f(y) - f(x) &= f(y_1, y_2, \ldots, y_n) - f(x_1, x_2, \ldots, x_n) \\ &= f(y_1, y_2, \ldots, y_n) - f(x_1, y_2, \ldots, y_n) \\ &\quad + f(x_1, y_2, \ldots, y_n) - f(x_1, x_2, y_3, \ldots, y_n) \\ &\quad + \cdots + f(x_1, x_2, \ldots, x_{n-1}, y_n) - f(x_1, \ldots, x_n) \\ &= Df(c_1, y_2, \ldots, y_n)(y_1 - x_1) + \cdots + Df(x_1, x_2, \ldots, x_{n-1}, c_n)(y_n - x_n) \end{aligned}$$

   where $c_i \in (x_i, y_i)$ for $i = 1, 2, \ldots, n$ by the MVT.

$\square$

21.3 Directional Derivatives Part 1

Definition 21.2. *Suppose $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ where Ω is open and $x_0 \in \Omega$. Then, given a unit vector $u \in \mathbb{R}^n$, the directional derivative of f at x_0 in the direction of u is*

$$D_u f(x_0) = \lim_{h \to 0} \frac{f(x_0 + hu) - f(x_0)}{h}$$

Example 21.3. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined as $f(x, y) = x^2 + y^2$. Then, to find the directional derivative in the direction of the vector $v = (3, 4)$ is found by first normalizing v to get $u = \left(\frac{3}{5}, \frac{4}{5}\right)$, from which we can calculate (omitted here) the result $D_u f(1, -1) = -\frac{2}{5}$.*

22 March 23

22.1 Directional Derivatives Part 2

Last time, we introduced the directional derivative but did not provide a relatively simple way to find it. We'll do that today.

Proposition 22.1. *Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be differentiable on the open set Ω and $x_0 \in \Omega$. Then, the derivatives $D_u f(x_0)$ exist in all directions of unit vectors u, and moreover, $D_u f(x_0) = \langle \nabla f(x_0), u \rangle$.*

Proof. Since f is differentiable at x_0, we know that

$$\lim_{x \to x_0} \frac{\left| f(x) - f(x_0) - \nabla f(x_0)(x - x_0) \right|}{\|x - x_0\|} = 0.$$

Then, let $x = x_0 + hu$, so then we get

$$\lim_{h \to 0} \left| \frac{f(x_0 + hu) - f(x_0)}{h} - \nabla f(x_0) \cdot u \right|.$$

$\square$

Note that the converse is not necessarily true!

Example 22.1. *Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined as $f(x, y) = x^2 + y^2$. Here, $\nabla f(x, y) = (2x \quad 2y)$, so $\nabla f(1, -1) = (2 \quad -2)$. If $u = \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix}$, then*

$$\langle \nabla f(1, -1), u \rangle = 2\left(\frac{3}{5}\right) + (-2)\left(\frac{4}{5}\right) = -\frac{2}{5}$$

22.2 Differentiation Rules

Example 22.2. *If $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $g : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ are differentiable on open set Ω $\lambda \in \mathbb{R}$, then $(f + g)$ and (λf) are differentiable on Ω, and if $x_0 \in \Omega$, then $D(f + g)(x_0) = Df(x_0) + Dg(x_0)$ and that $D(\lambda f)(x_0) = \lambda Df(x_0)$.*

In the following statement of the chain rule, let $\Omega_1 \subseteq \mathbb{R}^n$ be open and $\Omega_2 \subseteq \mathbb{R}^m$ also open. Then, suppose we have two functions $f : \Omega_1 \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $f : \Omega_2 \subseteq \mathbb{R}^m \to \mathbb{R}^p$ and that $f(\Omega_1) \subseteq \Omega_2$.

Proposition 22.2. *If f is differentiable at $x_0 \in \Omega_1$ and g is differentiable at $f(x_0) \in \Omega_2$, then $(g \circ f)$ is differentiable at x_0 and*

$$D(g \circ f)(x_0) = Dg(f(x_0)) \circ Df(x_0).$$

If we use Jacobians as the standard matrix for these derivatives, the chain rule reduces to matrix multiplication.

Proof. Let $\epsilon > 0$ be given. We want to exhibit $\delta > 0$such that, if $x \in \Omega_1$ and $\|x - x_0\| < \delta$, then

$$\left\| g \circ f(x) - g \circ f(x_0) - Dg(f(x_0)) \circ Df(x_0)(x - x_0) \right\| \le \epsilon \|x - x_0\|.$$

If we add and subtract $Dg(f(x_0))[f(x) - f(x_0)]$ inside the norm on the left, and then invoke the Triangle Inequality, we get

$$\|\cdot\| \le \left\| g(f(x)) - g(f(x_0)) - Dg(f(x_0))[f(x) - f(x_0)] \right\| + \left\| Dg(f(x_0))[f(x) - f(x_0) - Df(x_0)(x - x_0)] \right\|$$

Since f is differentiable at x_0, we can choose a $\delta_I > 0$ such that if $x \in \Omega_1$ and $\|x - x_0\| < \delta_I$, we can bound the second norm to be less than $\|Dg(f(x_0))\|\left\|f(x) - f(x_0) - Df(x_0)(x - x_0)\right\|$, where the first norm is the supremum norm, and then we can arrange for a $\delta_I I > 0$ such that, whenever $\|x - x_0\| < \delta_{II}$, we have

$$\left\|f(x) - f(x_0) - Df(x_0)(x - x_0)\right\| \leq \frac{\epsilon\|x - x_0\|}{2\left(\left\|Dg(f(x_0))\right\| + 1\right)}$$

so that we eventually get that, when $\|x - x_0\| < \min\{\delta_I, \delta_{II}\}$,

$$\left\|Dg(f(x_0))[f(x) - f(x_0) - Df(x_0)(x - x_0)]\right\| < \frac{\epsilon}{2}\|x - x_0\|.$$

To bound the first norm, we use the fact that differentiability implies continuity. We won't go through the specifics here, but, essentially:

1. We can use continuity to show that we can bring $f(x)$ and $f(x_0)$ as close as desired by bounding $\|x - x_0\|$.
2. Then, we invoke the differentiability of g at $f(x_0)$ to bound the first norm by $\frac{\epsilon}{2}\|x - x_0\|$.

Together, we now get that

$$\left\|g \circ f(x) - g \circ f(x_0) - Dg(f(x_0)) \circ Df(x_0)(x - x_0)\right\| < \epsilon\|x - x_0\|$$

which finishes the proof. □

Now that we have a couple rules familiar to us from single-variable calculus, what about the product rule? This is slightly trickier because of dimensions, so we'll do a relatively specific case here. Suppose that $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $g : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ where Ω is open.

Proposition 22.3. *If f, g, are differentiable on Ω, then $gf : \Omega \to \mathbb{R}^m$ is differential for each $x \in \Omega$ and*

$$D(gf)(x) = (Dg(x) \cdot v)f(x) + g(x)(Df(x) \cdot v), \forall v \in \mathbb{R}^n$$

22.3 Spherical Coordinates

The Cartesian coordinate system is useful in many ways, but other coordinate systems – such as the cylindrical coordinate system and the spherical coordinate system – can be especially helpful for tackling certain problems. Here, we deal with the spherical coordiante system, which uses ρ to measure distance from the origin, ϕ, which measures the angle measure from the z-axis, and θ, which measures the angle from the x-axis.

Then, we can create conversions in the form of

$$z = \rho\cos\phi,\ r = \rho\sin\phi,\ x = r\cos\theta = \rho\sin\phi\cos\theta,\ y = r\sin\theta = \rho\sin\phi\sin\theta$$

Example 22.3. *Suppose the temperate at a point with Cartesian coordinates (x, y, z) is given by $T(x, y, z) = x^2 + y^2 + z^2$. How sensitive is the temperature to small changes in ρ, ϕ, and θ? Consider the transformation*

$$\begin{bmatrix}\rho\\ \phi\\ \theta\end{bmatrix} \overset{f}{\longmapsto} \begin{bmatrix}\rho\sin\phi\cos\theta\\ \rho\sin\phi\sin\theta\\ \rho\cos\phi\end{bmatrix}$$

which we can compose with T to get T transformed into spherical coordinates. So, using the chain rule from above, we get

$$D(T \circ f)(\rho, \phi, \theta) = DT(f(\rho, \phi, \theta)) \circ Df(\rho, \phi, \theta)$$

so we have

$$\begin{aligned} D(T \circ f)(\rho, \phi, \theta) &= \begin{bmatrix} 2\rho \sin\phi \cos\theta & 2\rho \sin\phi \sin\theta & 2\rho \cos\phi \end{bmatrix} \begin{bmatrix} \sin\phi \cos\theta & \rho \cos\phi \cos\theta & -\rho \sin\phi \sin\theta \\ \sin\phi \sin\theta & \rho \cos\phi \sin\theta & \rho \sin\phi \cos\theta \\ \cos\phi & -\rho \sin\phi & 0 \end{bmatrix} \\ &= \begin{bmatrix} 2\rho & 0 & 0 \end{bmatrix}. \end{aligned}$$

23 March 25

Last time, we introduced the chain rule. Note the special case in which $f : \mathbb{R} \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}$ and both differentiable. Then, if we let $h(t) = g(f(t))$, we get that

$$\begin{aligned}\frac{\mathrm{d}h}{\mathrm{d}t} &= \nabla g(f(t)) \circ f'(t) \\ &= \langle \nabla g(f(t)), f'(t) \rangle\end{aligned}$$

But, how do we interpret this gradient vector geometrically?

23.1 Level Sets and Gradients

Consider $f : \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$ differentiable where Ω open. Then, the graph of f defined as all of the points in the set $\left\{(x, y, f(x, y)) \in \mathbb{R}^3 : (x, y) \in \Omega\right\}$ is a patch of surface. Given $c \in \mathbb{R}$, the level set, denoted L_c, is

$$\left\{(x, y) \in \Omega : f(x, y) = c\right\}.$$

Example 23.1. *Consider the function $f(x, y) = x^2 + y^2$ and the corresponding level sets:*

$$L_4 = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 4\right\}, \quad L_0 = \left\{(0, 0)\right\}, \quad L_{-1} = \emptyset.$$

If a level set happens to be a nice level curve, we ca also interpret the gradient vector at a point on that curve as the direction of steepest ascent – think of it as the fastest way of getting to another level set.

Definition 23.1. *A differentiable is regular if $\exists$ a parameterization γ so that $\gamma'(t) \neq 0$, $\forall t \in (a, b)$.*

Suppose that $f : \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$ is differentiable and that a level curve L_c admits a C^1 regular parameterization in the vicinity of some point (x_0, y_0) on L_c. Call it $\gamma(t)$ and assume $\gamma(t_0) = (x_0, y_0)$. Then, $f(\gamma(t)) = c, \forall t$ near t_0. Differentiate to get that $\langle \nabla f(\gamma(t)), \gamma'(t) \rangle = 0$. At $t = t_0$, get $\langle \nabla f(x_0, y_0), \gamma'(t_0) \rangle = 0$, so expect $\nabla f(x_0, y_0)$ to be normal to level curve through (x_0, y_0).

Example 23.2. *Consider the set of points $(x, y) \in \mathbb{R}^2$ such that*

$$\frac{y^2}{2} - \frac{x^2}{2} + \frac{x^4}{4} = 0$$

and call this $f(x, y)$. This is a level set of f. Now consider $(x_0, y_0) = \left(1, \frac{\sqrt{2}}{2}\right)$, which satisfies $f(x_0, y_0) = 0$. Note that $\nabla f(x, y) = \begin{bmatrix} x^3 - x & y \end{bmatrix}$ and so $\nabla f\left(1, \frac{\sqrt{2}}{2}\right) = \begin{bmatrix} 0 & \frac{\sqrt{2}}{2} \end{bmatrix}$. So, the tangent line must be horizontal. $y = \frac{\sqrt{2}}{2}$.

23.2 MVT and Taylor-Lagrange Theorem But Cooler (More Variables)

Generalizing the MVT can't be done naively.

Example 23.3. *Suppose $\gamma : [0, \pi] \to \mathbb{R}^2$ as $\gamma(t) = \begin{bmatrix} \cos t \\ \sin t \end{bmatrix}$. Does $\gamma(\pi) - \gamma(0) = \gamma'(c) \cdot (\pi \cdot 0)$ for some $c \in (0, \pi)$? This is equivalent to asking if $\begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} -\sin c \\ \cos c \end{bmatrix} \pi$, which cannot happen because $\pi \cos c = 0$ implies $c = \frac{\pi}{2}$, which doesn't work for $-\pi \sin c$.*

There are many ways that we could try to formulate a MVT for higher dimension functions, but only one is useful for us.

Definition 23.2. *Let $S \in \mathbb{R}^n$. We say that S is convex if given $x, y \in S$, every point of form $L(t) = x + t(y - x)$ for $t \in [0,1]$ is also in S.*

Proposition 23.1. *(MVT) Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be differential on the open set Ω. Let $x, y \in \Omega$ be such that $L(t) = x + t(y - x) \in \Omega$ for $t \in [0,1]$. Then, $\exists c \in [0,1]$ such that $f(y) - f(x) = \nabla(f(L(c)) \cdot (y - x)$.*

Proof. L is differential, so $g = f \circ L$ where $g : [0,1] \to \mathbb{R}$ is continuous on $[0,1]$ and differentiable on $(0,1)$. Apply the standard MVT to this to see that, $\exists c \in (0,1)$ such that

$$\begin{aligned} g(1) - g(0) &= g'(c)(1-0) \\ f(y) - f(x) &= \langle \nabla f(L(c)), L'(c) \rangle \\ &= \langle \nabla f(L(c)), y - x \rangle \end{aligned}$$

□

So, MVT isn't too hard to translate to higher dimensions, but translating Taylor's theorem is much, much harder.

Suppose $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ differentiable. If $x_0 \in \Omega$, $\exists Df(x_0) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Then, we have a mapping $x \mapsto Df(x)$, which comprises a function $Df : \Omega \subseteq \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Like f, Df maps from an open subset of $\mathbb{R}^n$ into a vector space, so it may also be differentiable.

Definition 23.3. *If Df is differentiable at $x_0 \in \Omega$, then $D(Df)(x_0)$ is called the second derivative of f at x_0 and is denoted $D^2 f(x_0)$.*

How exactly are we to make sense of and compare things with $D^2 f(x_0)$? Well, we know that $D^2 f(x_0) : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, so given $y \in \mathbb{R}^n$, we know that $D^2 f(x_0) y \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. If $z \in \mathbb{R}^n$, $(D^2 f(x_0) y) z \in \mathbb{R}^m$.

This is some awkward notation, so we'll sometimes write $D^2 f(x_0)(y, z)$ in place of the above notation. M&H will write $B_{x_0}(y, z) = (D^2 f(x_0) y)(z)$, where $B_{x_0} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ but this is even more confusing – this makes it seem like a ball! Boo!

23.3 Higher Partials and Bilinear Forms

Suppose $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is bilinear (i.e., linear in both arguments). Then, if e_j is the jth standard basis vector for $\mathbb{R}^n$, B is completely determined by how it acts on pairs (e_i, e_j). So, let $A_{ij} = B(e_i, e_j)$. Now, if we let

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n.$$

We can check that $B(x, y) = x^T A y$, where $A \in \mathbb{R}^{n \times n}$ has entry A_{ij}.

Example 23.4. *If $B(x, y) = \langle x, y \rangle$, then $A = I$.*

24 March 28

24.1 Second Derivatives and Mixed Partials

Assuming that $\frac{\partial f}{\partial x_i}$ exists throughout Ω, the partial $\frac{\partial f}{\partial x_i}$ may also have partials that exist as well. So, we get

$$\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

and we can also write

$$f_{x_i} = \frac{\partial f}{\partial x_i} \implies f_{x_i x_j} = \left(f_{x_i}\right)_{x_j}$$

Example 24.1. *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as $f(x,y) = x\cos\left(xy\right)$. Then, we see that*

$$\begin{aligned} f_x(x,y) &= \cos\left(xy\right) - xy\sin\left(xy\right) \\ f_{xx} &= -y\sin\left(xy\right) - y\sin\left(xy\right) - xy^2\cos\left(xy\right) \\ f_{xy} &= -x\sin\left(xy\right) - x\sin\left(xy\right) - x^2y\cos\left(xy\right) \\ f_{xy} &= f_{yx} \end{aligned}$$

Now, suppose that $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is twice differentiable on Ω and let $x \in \Omega$.

Proposition 24.1. *Consider the bilinear form $B_x(y,z) = D^2 f(x)(y,z)$. Then, with respect to the standard basis vectors, the matrix associated with $B_x(y,z)$ is*

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

and we can then invoke $B_x(y,z)$ as $D^2 f(x)(y,z) = y^T H z$.

Example 24.2. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x,y) = \sin(x) + \cos\left(y\right)$. Then, we we approximate $f(x,y)$ near $(0,0)$ as*

$$f(x,y) \approx f(0,0) + \nabla f(0,0) \cdot \begin{bmatrix} x-0 \\ y-0 \end{bmatrix} + \frac{1}{2}(x-0, y-0)H(0,0)\begin{bmatrix} x-0 \\ y-0 \end{bmatrix}$$

and we can find that

$$f_x = \cos(x),\ f_y = \sin\left(y\right),\ f_{xy} = 0 = f_{yx},\ f_{xx} = -\sin(x),\ f_{yy} = -\cos(x)$$

so our approximation becomes

$$\begin{aligned} f(x,y) &\approx 1 + \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{2}\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x & y \end{bmatrix} \\ &= 1 + x - \frac{y^2}{2} \end{aligned}$$

Theorem 24.1. *(Clairaut - Schwarz) If $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is twice differentiable on the open set Ω and $D^2 f$ is continuous, then for each $x \in \Omega$, the bilinear form $D_x(y,z) = D^2 f(x)(y,z)$ is symmetric. So, $D^2 f(x)(y,z) = D^2 f(x)(z,y)$, $\forall y,z \in \mathbb{R}^n$.*

Corollary 24.1.1. *Under the above hypothesis, we see that mixed partials have the property*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

.

Proof. Abbreviated. Use the Clairaut-Schwarz theorem above with $y = e_i$, $z = e_j$. □

Note that, if the second partials exist and are continuous in some open set U, then the mixed partials are equal throughout U.

24.2 Multivariable Taylor-Lagrange Theorem

Definition 24.1. *A function f is of class C^r on its domain if its first r derivatives exist and are continuous. We then say that, if $f \in C^\infty$, then f is of class C^r, $\forall r \in \mathbb{N}$. Finally, we'll say that f is smooth if $f \in C^\infty$.*

Theorem 24.2. *(Taylor-Lagrange) Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be of class C^{k+1} on the open set Ω. Now, suppose $x, x_0 \in \Omega$ such that, for each $t \in [0,1]$, the point $L(t) = x_0 + t(x - x_0) \in \Omega$. Then, $\exists c = L(t_0)$, where $t_0 \in [0,1]$, such that*

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + \frac{1}{2!}D^2 f(x_0)(x - x_0, x - x_0)$$
$$+ \cdots + \frac{1}{k!}D^{(k)} f(x_0)\overbrace{(x - x_0, \ldots, x - x_0)}^{k \text{ times}} + \frac{1}{(k+1)!}D^{(k+1)} f(c)\overbrace{(x - x_0, \cdots, x - x_0)}^{k+1 \text{ times}}$$

Also, if $x = x_0 + h$,

$$f(x) = f(x_0) + Df(x_0)h + \cdots + \frac{1}{k!}D^{(k)}\overbrace{f(x_0)(h, \ldots, h)}^{k \text{ times}} + E(x_0, h)$$

where we then have

$$\frac{\left\|E(x_0, h)\right\|}{\left\|h\right\|^k} \to 0 \quad \text{as} \quad \left\|h\right\| \to 0.$$

Example 24.3. *Consider $f : \mathbb{R} \to \mathbb{R}$ where f is class C^{k+1}.*

$$f(x) = f(x_0) + \cdots + \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \frac{f^{(k+1)}(c)}{(k+1)!}(x - x_0)^{k+1}.$$

If I is the closed interval with endpoints at x and x_0, then we can bound this last term with EVT. If we let

$$M^{(k+1)} = \max_{y \in I}\left|f^{(k+1)}(y)\right|$$

then we find that

$$\text{error} \le \frac{M^{(k+1)}(x - x_0)^{k+1}}{(k+1)!}$$

Example 24.4. *Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined as $f(x,y) = \sin x + \cos y + xy$. Then, we know that f has continuous partials of all orders (i.e. f is smooth) and the domain is convex. So, given any $(x,y) \in \mathbb{R}^2$, by the Taylor-Lagrange Theorem, if $h = \begin{bmatrix} x \\ y \end{bmatrix}$, then*

$$f(x,y) = f(0,0) + \nabla f(0,0)h + \frac{1}{2}D^2 f(0,0)(h,h) + \frac{1}{3!}D^3 f(c)(h,h,h)$$

where c is on the line segment connecting (x,y) to the origin. In other words,

$$f(x,y) = 1 + x - \frac{1}{2}y^2 + xy + \text{error term}$$

Definition 24.2. *Suppose, for $j \in \mathbb{N}$, that $D^{(j)} f_{x_0}$ exists. Then, the series*

$$\sum_{j=0}^{\infty} \frac{1}{j!} D^{(j)} f(x_0)(x - x_0, \cdots, x - x_0)$$

is the Taylor series for f centered at x_0. If the series converges in some open neighborhood of x_0, we say f is analytic at x_0.

Example 24.5. *Let $f : \mathbb{R} \to \mathbb{R}$ be defined as*

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

Then, f is C^{∞} but not analytic at 0.

25 March 30

25.1 Optimization and Extrema for Multivariable Functions

Today, we'll focus on maxima, minima, and optimization for functions of several variables.

Remember from previous lectures (and single-variable calculus) that a real-valued function has its local maximum at a point x_0 in its domain if there exists an open neighborhood U of x_0 such that, if $x \in U$ is in the domain of f, then $f(x) \le f(x_0)$. The local minimum is defined analogously.

If $f : (a,b) \to \mathbb{R}$ is differentiable, and if f has a local max or min at $c \in (a,b)$, then $f'(c) = 0$. Also, if $f : [a,b] \to \mathbb{R}$ is continuous, then EVT guarantees that f achieves its absolute extrema somewhere.

Definition 25.1. *Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, where Ω is open. We say that $x_0 \in \Omega$ is a critical point of f if either $\nabla f(x_0) = 0$ or if it does not exist.*

Proposition 25.1. *Suppose $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable on the open set Ω. If f has a local extremum at $x_0 \in \Omega$, then $\nabla f(x_0) = 0$.*

Proof. (Contradiction). Assume f has a local max at $x_0 \in \Omega$ but $\nabla f(x_0) \neq 0$. Now, let $\epsilon = \frac{1}{2}\left\|\nabla f(x_0)\right\| > 0$. Choose a $\delta > 0$ such that $B(x_0, \delta) \subseteq \Omega$ and

$$\|x - x_0\| < \delta \implies \left|f(x) - f(x_0) - \nabla f(x_0) \cdot (x - x_0)\right| \le \epsilon \|x - x_0\|$$

which is possible by the differentiability of f. Now, let

$$u = \frac{\nabla f(x_0)}{\left\|\nabla f(x_0)\right\|}$$

such that u is a unit vector parallel to the gradient vector. Then, if we let $x = x_0 + hu$, where $0 < h < \delta$, then $\|x - x_0\| < \delta$, so $\left|f(x_0 + hu) - f(x_0) - \nabla f(x_0) \cdot hu\right| \le \epsilon h$. Then, notice that

$$\begin{aligned} -\epsilon h &\le f(x_0 + hu) - f(x_0) - \nabla f(x_0) \cdot hu \\ &= f(x_0 + hu) - f(x_0) - h\left\|\nabla f(x_0)\right\| \end{aligned}$$

so then $f(x_0 + hu) \ge f(x_0) + \frac{1}{2}h\left\|\nabla f(x_0)\right\|$ for $0 < h < \delta$. So there are points arbitrarily close to x_0 at which $f(x) > f(x_0)$. ↯ ∎

Note also that, if $f : K \subseteq \mathbb{R}^n \to \mathbb{R}$ is continuous on a compact set K, then EVT implies that f achieves an absolute max/min somewhere on K. The absolute extrema must occur either on ∂K or at a critical point in $\text{int}(K)$.

Example 25.1. *Consider the function $f : K \subseteq \mathbb{R}^2 \to \mathbb{R}$ defined by $z \xmapsto{f} \|z\|$ on the domain*

$$K = \left\{(x,y) \in \mathbb{R}^2 : \frac{x^2}{4} + \frac{y^2}{9} \le 1\right\}$$

such that K is compact. Note that $f(x,y) = \sqrt{x^2 + y^2}$. So, we get

$$\frac{\partial f}{\partial x} = \frac{x}{\sqrt{x^2+y^2}}, \quad \frac{\partial f}{\partial y} = \frac{y}{\sqrt{x^2+y^2}}$$

so $\nabla f(x,y)$ is never equal to 0 but it fails to exist at $(0,0)$. Note that $f(0,0) = 0$. Then, we want to check ∂K, which we can parameterize as $\gamma(t) = \begin{bmatrix} 2\cos t \\ 3\sin t \end{bmatrix}$ for $0 \le t \le 2\pi$. Then,

$$f(\gamma(t)) = \sqrt{(2\cos t)^2 + (3\sin t)^2}.$$

Then, we can optimize f using techniques from single variable calculus.

Example 25.2. *Consider $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x^3$. Then, there is a critical point at $x = 0$ but clearly this is not a local extrema.*

Example 25.3. *Differentiable functions from $\mathbb{R}^n \to \mathbb{R}$ can have "saddle points" that are a generalization of the phenomenon from the previous example. Consider $f(x,y) = x^2 - y^2$ such that $\nabla f(x,y) = (2x \quad -2x)$. Then, $\nabla f(x,y) = 0$ if and only if $x = y = 0$, so $(0,0)$ is the critical point. In the plane $y = 0$, the cross section is a parabola $z = x^2$, and, for $x = 0$, we get $z = -y^2$. So, in every neighborhood of $(0,0) \in \mathbb{R}^2$, there are points (x,y) with $f(x,y) > f(0,0)$ and points with $f(x,y) < f(0,0)$. This is a saddle point.*

We now want a calculus-based test for classifying critical points as either local minima, local maxima, or saddle points.

Definition 25.2. *A bilinear form $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called positive definite if $B(x,x) \geq 0$, $\forall x \in \mathbb{R}^n$, with equality if and only if $x = 0$.*

Theorem 25.1. *(Second derivative test). Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be of class C^2 on the open set Ω. Now, assume f has a critical point at $x_0 \in \Omega$. If the bilinear form $B_{x_0}(y,z) = D^2 f(x_0)(y,z)$ is positive definite, then f has a local minimum at x_0.*

Proof. The goal is to show that $\exists \delta > 0$ such that $B(x_0, \delta) \subseteq \Omega$ and $\forall x \neq x_0$ in $B(x_0, \delta)$, we have $f(x) > f(x_0)$. As in the proof of the proposition earlier, we will single out a special direction.

Since we know $D^2 f(x_0)(x,x) > 0$ if $x \neq x_0$ and this holds for all unit vectors, consider $u \overset{\phi}{\longmapsto} D^2 f(x_0)(u,u)$ where $\|u\| = 1$. The set $K = \{u \in \mathbb{R}^n : \|u\| = 1\}$ is compact. The map ϕ is continuous, so EVT implies that $\exists u^*$ that minimizes $D^2 f(x_0)(u,u)$. Now, let

$$\epsilon = D^2 f(x_0)(u^*, u^*) > 0.$$

Then, for all unit vectors u, $D^2 f(x_0)(u,u) \geq \epsilon$. If $x \neq x_0$, we can write $x = \|x\| \frac{x}{\|x\|}$ and use bilinearity to get that

$$D^2 f(x_0)(x,x) \geq \epsilon \|x\|^2, \forall x \in \mathbb{R}^n. \tag{1}$$

Then, since Ω is open and $D^2 f$ is continuous, we can choose $\delta > 0$ such that $B(x_0, \delta) \subseteq \Omega$ and

$$\|x - x_0\| < \delta \implies \left\| D^2 f(x) - D^2 f(x_0) \right\|_\infty < \frac{\epsilon}{2}. \tag{2}$$

Now, given $x \in B(x_0, \delta)$, which is convex, we can apply the Taylor-Lagrange Theorem, to find that

$$\begin{aligned} f(x) &= f(x_0) + \nabla f(x_0) \cdot (x - x_0) + \frac{1}{2} D^2 f(c)(x - x_0, x - x_0) \\ &= f(x_0) + \frac{1}{2} D^2 f(c)(x - x_0, x - x_0) \end{aligned}$$

for some c on the line segment connecting x_0 and x. Now, use (1) and (2) to find that

$$D^2 f(c)(x - x_0, x - x_0) \geq \frac{\epsilon}{2} \|x - x_0\|^2.$$

If $\|x - x_0\| < \delta$, then $\forall x \in B(x_0, \delta)$, get

$$f(x) \geq f(x_0) + \frac{\epsilon}{4} \|x - x_0\|^2$$

so we have a local minimum at x_0. □

26 April 1

One last optimization example because we love them so much.

Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x,y) = \frac{y^2}{2} - \frac{x^2}{2} + \frac{x^4}{4}$. Then, $\nabla f(x,y) = (-x + x^3, \quad y)$.

So, there appear to be three critical points: $(\pm 1, 0), (0,0)$. But, if we check the Hessian matrix

$$H(x,y) = \begin{bmatrix} 3x^2 - 1 & 0 \\ 0 & 1 \end{bmatrix}$$

we see that

$$H(\pm 1, 0) = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

which has positive eigenvalues, indicating that the matrix is positive definite. So, f has local minima at $(-1,0)$ and $(1,0)$. Then, we also have

$$H(0,0) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

which has eigenvalues of opposite signs, so $(0,0)$ must be a saddle point.

26.1 Inverse Function Theorem

Recall that, if (X, d_1) and (Y, d_2) are metric spaces, a function $f : X \to Y$ that is continuous and has a continuous inverse is called a homomorphism.

Definition 26.1. *Let $\Omega_1 \subseteq \mathbb{R}^n$ and $\Omega_2 \subseteq \mathbb{R}^n$, both open. A map $f : \Omega_1 \to \Omega_2$ is called a diffeomorphism if f is a differentiable bijection with a differentiable inverse. If both f and f^{-1} are of class C^r with $r \geq 1$, we say it's a C^r-diffeomorphism . We say f is a local diffeomorphism if, for each x in the domain, there is an open neighborhood U of x contained in the domain of f such that $f(U)$ is open, and $f|_U$ is a diffeomorphism onto $f(U)$.*

Example 26.1. *Consider the function $f : (-2,2) \to (-8,8)$ be defined by $f(x) = x^3$. Then, f is a homeomorphism. f is also differentiable. But f^{-1} is not differentiable at 0, so f is not a diffeomorphism.*

Consider the single-variable version of the inverse function theorem. Suppose $f : \Omega \subseteq \mathbb{R} \to \mathbb{R}$ is class C^1 on an open set Ω. Let $x_0 \in \Omega$ and $y = f(x_0)$. If $f'(x_0) \neq 0$, then f is a local C^1-diffeomorphism from some open U containing x_0 onto $f(U)$ containing y_0.

It's important that f is C^1, not just differentiable.

Example 26.2. *Consider the following cautionary case:*

$$f(x) = \begin{cases} \frac{x}{2} + x^2 \sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

where f is differentaible and $f'(0) = \frac{1}{2} \neq 0$ but f is not locally invertible at 0.

Theorem 26.1. *(Inverse function theorem). Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be of class C^r where $r \geq 1$ on the open set Ω. Suppose $x_0 \in \Omega$ is such that $\det Df(x_0) \neq 0$. Then, $\exists$ an open neighborhood U of x_0 contained in Ω such that $W = f(U)$ is an open neighborhood of $y_0 = f(x_0)$ and $f|_U$ is a C^r-diffeomorphism of U onto W. Moreover, for each $y \in W$, $Df^{-1}(y) = \left[Df(x)\right]^{-1}$ where $x = f^{-1}(y)$.*

We will prove the theorem for $r = 1$ by invoking the contractoin mapping theorem on an appropriate domain to construct a local inverse.

Proof. First, we will begin with **preliminaries**.

Without loss of generality, we may assume that $Df(x_0) = I$. If $Df(x_0) \neq I$, we know that $\det Df(x_0) \neq 0$, so we can define $F(x) = Df(x_0)^{-1} f(x)$ on Ω. Then, $DF(x_0) = Df(x_0)^{-1} Df(x_0) = I$. So, the theorem applies to F. We can get f^{-1} by doing $f^{-1}(y) = F^{-1}(Df(x_0)^{-1}y)$ for each y. Note that F is C^r because f and F^{-1} are, so f^{-1} is C^r.

Also without loss of generality, we can assume that $x_0 = 0$ and $y_0 = f(x_0) = 0$. Suppose that we prove the theorem in this special case and are later given a function f with $x_0 \neq 0$ and $f(x_0) \neq 0$. Define $\Omega_- = \left\{y - x_0 : y \in \Omega\right\}$, which is open because Ω is open. Now, define h on Ω_- as $h(x) = f(x + x_0) - f(x_0)$. Now, $0 \in \Omega$ and $h(0) = 0$ and $h \in C^r$.

Now, we can begin our main proof.

First, we want to show the existence of a local inverse. We want to establish an open neighborhod U of $x_0 = 0$ and W of $y_0 = f(x_0) = 0$ such that, $\forall y \in W$, there exists a unique $x \in U$ with $f(x) = y$.

Now, we want to set up for the application of the contraction mapping theorem. Given $y \in \mathbb{R}^n$, define $g_y : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ as $g_y(x) = x - f(x) + y$. Then, the only way that g_y could have a fixed point is if

$$x^* = x^* - f(x^*) + y \implies f(x^*) = y.$$

For $x \in \Omega$, we also know that g_y is differentiable at x and $Dg_y(x) = I - Df(x)$. In particular, $Dg_y(0) = I - Df(0) = I - I = 0$ by assumption. Since $f \in C^1(\Omega)$, Df – and therefore Dg_y are continuous on Ω. Then, Dg_y continuous and $Dg_y(0) = 0$ implies that we can find an open ball U centered at $x_0 = 0$ such that $\left\|Dg_y(x)\right\| < \frac{1}{2}, \forall x \in U$. We will refine this U later.

We then claim that $Df(x)$ is invertible, $\forall x \in U$. If it was not, we could choose a unit vector $v \in \mathbb{R}^n$ such that $Df(x)v = 0$. But, then

$$\left\|Dg_y(x)\right\| = \sup_{\|u\|=1} \left\|Dg_y(x)u\right\| \geq \left\|Dg_y(x)v\right\| = \left\|Iv - Df(x)v\right\| = \|v\| = 1$$

which forms a contradiction, so the above claim holds.

Since U is convex and $\left\|Dg_y(x)\right\| < \frac{1}{2}$, by MVT, given $x_1, x_2 \in U$ and any y, we get that

$$\left\|g_y(x_2) - g_y(x_1)\right\| \leq \frac{1}{2}\|x_2 - x_1\|. \tag{1}$$

We're almost there, but to finish setting up the contraction mapping theorem, we need to have $g_y : K \to K$ where K is a closed subset of a Banach space.

As a brief aside, the map $\psi : GL_n(\mathbb{R}) \to GL_n(\mathbb{R})$ defined by $A \overset{\psi}{\longmapsto} A^{-1}$ is smooth, where $GL_n(\mathbb{R})$ is the set of invertible linear operators on $\mathbb{R}^n$.

Since Df is continuous and ψ is continuous and $\left\|Df(0)^{-1}\right\| = \|I\| = 1$, we can also assume that U was chosen such that $\left\|Df(x)^{-1}\right\| \leq 2$, $\forall x \in U$. So, write $U = B(0, r)$ where r is appropriately small such that the earlier inequality and this new inequality both hold.

Let $W = f(U)$. We claim that f is injective on U. Suppose indirectly that $\exists x_1, x_2 \in U$ such that $x_1 \neq x_2$ but $f(x_1) = f(x_2)$. But, by (1), this means that

$$\begin{aligned}
\left\|g_y(x_2) - g_y(x_1)\right\| &\leq \frac{1}{2}\|x_2 - x_1\| \\
\left\|x_2 - f(x_2) + y - x_1 + f(x_1) - y\right\| &\leq \frac{1}{2}\|x_2 - x_1\| \\
\|x_2 - x_1\| &\leq \frac{1}{2}\|x_2 - x_1\|
\end{aligned}$$

which forms a contradiction. If $W = f(U)$, then $f|_U : U \to W$ is bijective! So $f^{-1} : W \to U$ is well-defined.

Phase 3: Show that W is open.

Pick any $z \in W$. Want to exhibit $\delta > 0$ such that $B(z,\delta) \subseteq W$. Since f^{-1} is well-defined on W, let $\xi = f^{-1}(z)$. Choose $\delta > 0$ such that $K = \overline{B(\xi,\delta)} \subseteq U$. We claim that $B(z,\frac{\delta}{2}) \subseteq W$. Pick any $y \in B(z,\frac{\delta}{2})$ and will show that $y \in W$. Given any $x \in K$, we know

$$\begin{aligned}
\left\|g_y(x) - \xi\right\| &\leq \left\|g_y(x) - g_y(\xi)\right\| + \left\|g_y(\xi) - \xi\right\| \\
&\leq \frac{1}{2}\|x - \xi\| + \left\|\xi - f(\xi) + y - \xi\right\| \\
&= \frac{1}{2}\|x - \xi\| + \left\|y - z\right\| \\
&< \frac{1}{2}\delta + \frac{1}{2}\delta = \delta.
\end{aligned}$$

so $g_y : K \to K$. Finally, we can apply the contraction mapping theorem. There exists $x \in K$ such that $g_y(x) = x$, which is equivalent to $f(x) = y$. Since $x \in K \subseteq U$, this means that $y \in f(K) \subseteq f(U) = W$. So, W is open.

Stage 4: Show that f^{-1} is C^1 on W. Helpful to show that f^{-1} is continuous. Pick any $z_1 \in W$. Let $\xi_1 = f^{-1}(z_1)$. Given any other $z_2 \in W$, let $xi_2 = f^{-1}(z_2)$. Then,

$$\left\|f^{-1}(z) - f^{-1}(z_2)\right\| = \|\xi_1 - \xi_2\|$$

But $g_0(\xi_i) = \xi_i - f(\xi_i)$, so the left side of the equality becomes

$$\begin{aligned}
\left\|f^{-1}(z) - f^{-1}(z_2)\right\| &= \left\|g_0(\xi_1) + f(\xi_1) - g_0(\xi_2) - f(\xi_2)\right\| \\
&\leq \left\|g_0(\xi_1) - g_0(\xi_2)\right\| + \left\|f(\xi_1) - f(\xi_2)\right\| \\
&\leq \frac{1}{2}\|\xi_1 - \xi_2\| + \|z_1 - z_2\| \\
&= \frac{1}{2}\left\|f^{-1}(z_1) - f^{-1}(z_2)\right\| + \|z_1 - z_2\|
\end{aligned}$$

So, returning to the above equality, we get

$$\left\|f^{-1}(z_1) - f^{-1}(z_2)\right\| \leq 2\|z_1 - z_2\| \tag{2}$$

from which the continuity of f^{-1} follows.

Now, we will show that f^{-1} is differentiable on W – that $Df^{-1}(y)$ exists and that $Df^{-1}(y) = [Df(x)]^{-1}$. Fix any $y \in W$ and let $x = f^{-1}(y)$. Then, let $z \in W$ with $z \neq y$. We want to show that

$$\lim_{z \to y} \frac{\left\|f^{-1}(z) - f^{-1}(y) - [Df(x)]^{-1}(z-y)\right\|}{\left\|z - y\right\|} = 0.$$

Then, if we let $w = f^{-1}(z)$, we get that the quantity inside the norm in the numerator is

$$\begin{aligned}
f^{-1}(z) - f^{-1}(y) - [Df(x)]^{-1}(z-y) &= w - x - [Df(x)]^{-1}(f(w) - f(x)) \\
&= -[Df(x)]^{-1}(f(x) - f(x) - Df(x)(w-x)).
\end{aligned}$$

Then, since U was chosen such that $\left\|[Df(x)]^{-1}\right\| \leq 2, \forall x \in U$, we get

$$\left\|f^{-1}(z) - f^{-1}(y) - [Df(x)]^{-1}(z-y)\right\| \leq 2\left\|f(w) - f(x) - Df(x)(w-x)\right\|.$$

so instead we can show that

$$\lim_{z \to y} \frac{\left\|f(w) - f(x) - Df(x)(w-x)\right\|}{\left\|z - y\right\|} = 0.$$

And, since f is differentiable, we know that

$$\lim_{w \to x} \frac{\left\| f(w) - f(x) - Df(x)(w-x) \right\|}{\|w-x\|} = 0$$

but the continuity of f^{-1} implies that as $z \to y$, we also get $w \to x$, and our earlier proof of the continuity of f^{-1} (specifically inequality (2)), we see that

$$\frac{\|w-x\|}{\left\|z-y\right\|} \leq 2.$$

So we finally get

$$\lim_{z \to y} \frac{\left\| f(w) - f(x) - Df(x)(w-x) \right\|}{\left\|z-y\right\|} = \lim_{z \to y} \frac{\left\| f(w) - f(x) - Df(x)(w-x) \right\|}{\|w-x\|} \frac{\|w-x\|}{\left\|z-y\right\|} = 0$$

which now shows that f^{-1} is differentaible on W and $Df^{-1}(y) = [Df(x)]^{-1}$, where $x = f^{-1}(y)$.

Finally, remember that ϕ as defined earlier as $A \overset{\phi}{\longmapsto} A^{-1}$ is smooth, and since f is of class C^r on U, we know that Df is of class C^{r-1} on U, so Df^{-1} is of class C^{r-1} on W, which implies f^{-1} is of class C^r on W. $\square$

27 April 4

Suppose $f : \Omega : \mathbb{R}^n \to \mathbb{R}^n$ is of class C^r on the open set Ω and that $x_0 \in \Omega$ is a point at which $\det Df(x_0) \neq 0$. Then, per the inverse function theorem, $\exists$ open neighborhoods U of x_0 with $U \subseteq \Omega$ and $W = f(U)$ of $y_0 = f(x_0)$ such that $f|_U$ is a C^r-diffeomorphism of U onto W. Moreover, for each $y \in W$, $Df^{-1}(y) = [Df(x)]^{-1}$ where $x = f^{-1}(y)$.

Example 27.1. *Consider the mapping*

$$\begin{cases} u &= y + 3\cos\left(xy\right) \\ v &= \ln\left(x + 2y\right). \end{cases}$$

Can we solve (at least in principle) for $\begin{bmatrix} x \\ y \end{bmatrix}$ *as a function of* $\begin{bmatrix} u \\ v \end{bmatrix}$ *locally near* $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$*?*

We can consider this mapping to be the function $f : \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}^2$ *where*

$$f(x,y) = \begin{bmatrix} y + 3\cos\left(xy\right) \\ \ln\left(x + 2y\right) \end{bmatrix}.$$

Then, if we consider $\Omega = \left\{(x,y) \in \mathbb{R}^2 : x + 2y > 0\right\}$ *such that* $\ln\left(x+2y\right)$ *is defined on* Ω*, we find that*

$$Df(x,y) = \begin{bmatrix} -3y\sin\left(xy\right) & 1 - 3x\sin\left(xy\right) \\ \frac{1}{x+2y} & \frac{2}{x+2y} \end{bmatrix}$$

so then $Df(1,0) = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$*, which has determinant of* -1*, which is not equal to* 0*. Then, we can also tell visually that* Df *is continuous, so we can apply the inverse function theorem.*

As an extension, let us consider estimating near this point $(x,y) = (1,0)$*.*

$$\begin{aligned} \begin{bmatrix} u \\ v \end{bmatrix} &\approx f\begin{bmatrix} 1 \\ 0 \end{bmatrix} + Df(1,0)\begin{bmatrix} x-1 \\ y \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x-1 \\ y \end{bmatrix}. \end{aligned}$$

but we can do something else too, with the help of the inverse function theorem:

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix} &\approx f^{-1}\begin{bmatrix} 3 \\ 0 \end{bmatrix} + Df^{-1}(3,0)\begin{bmatrix} u-3 \\ v \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} u-3 \\ v \end{bmatrix}. \end{aligned}$$

27.1 Implicit Function Theorem

We will now formulate a generalization of the inverse function theorem that's good for exploring the solvability of nonlinear systems of equations.

Consider the following: when, if anywhere, does an equation $F(x,y) = 0$ implicitly define a function $y = f(x)$? (at least locally).

Example 27.2. *Consider* $F(x,y) = \frac{y^2}{2} - \frac{x^2}{2} + \frac{x^4}{4} = 0$*. It seems like we could locally find functions at each point on the curve except at the exceptional points* $(0,0)$ *and* $(\pm 1, 0)$*.*

Theorem 27.1. *(Implicit function theorem, special case). Suppose $F : \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$ is of class C^r ($r \geq 1$) on the open set Ω. Suppose $(x_0, y_0) \in \Omega$, $F(x_0, y_0) = 0$, and $\frac{\partial F}{\partial y} \neq 0$. Then, $\exists$ open neighborhoods $U \in \mathbb{R}$ of x_0 and $V \subseteq \mathbb{R}$ of y_0 and a* unique *function $f : U \to V$ such that $U \times V \subseteq \Omega$ and $F(x, f(x)) = 0$, $\forall x \in U$. Moreover, f is of class C^r on U.*

Proof. We know that Ω is open. We'll construct f and open sets U and V.

Assume, without loss of generality, that $\frac{\partial F}{\partial y} > 0$. If not, we can apply the following to $-F$. Since F is of class C^r on Ω, we know that $\frac{\partial F}{\partial y}$ is continuous on Ω. So, we can choose $\delta > 0$ such that the box $K = [x_0 - \delta, x_0 + \delta] \times [y_0 - \delta, y_0 + \delta]$ is contained in Ω and that $\frac{\partial F}{\partial y} > 0$ throughout K. Then, look at the real-valued function $g_{x_0}(y) = f(x_0, y)$ for $y \in [y_0 - \delta, y_0 + \delta]$. Then, we know that $g_{x_0}(y_0) = 0 = F(x_0, y_0)$. Additionally, we know that $g'_{x_0}(y) = \frac{\partial F}{\partial y} > 0$, so g_{x_0} is strictly increasing on the interval $[y_0 - \delta, y_0 + \delta]$, so $g_{x_0}(y_0 + \delta) = F(x_0, y + \delta_0) > 0$ and similarly $g_{x_0}(y_0 - \delta) = F(x_0, y_0 - \delta_0) < 0$.

Since F continuous, we can choose $\delta_1 > 0$ (with $\delta_1 \leq \delta$) such that $F(x, y_0 + \delta) > 0$ if $x \in [x_0 - \delta_1, x_0 + \delta_1]$. Then, choose $\delta_2 > 0$ (with $\delta_2 \leq \delta$) such that $F(x, y_0 - \delta) < 0$, if $x \in [x_0 - \delta_2, x_0 + \delta_2]$. Then, let $\delta_3 = \min\{\delta_1, \delta_2\}$.

Now, consider $U = (x_0 - \delta_3, x_0 + \delta_3)$. We know that $U \times [y_0 - \delta, y_0 + \delta] \subseteq K$. Again, $\frac{\partial F}{\partial y} > 0$ in K. Pick $x \in U$ and look at $g_x(y) = F(x, y)$, such that x is fixed (the notation does *not* refer to a partial derivative). Then, we know that g_x is strictly increasing on $[y_0 - \delta, y_0 + \delta]$. We then know that $g_x(y_0 - \delta) < 0$ and $g_x(y_0 + \delta) > 0$ and $g'_x(y) = \frac{\partial F}{\partial y} > 0$, so the intermediate value theorem and the mean value theorem allow us to say that $\exists$ unique $y \in (y_0 - \delta, y_0 + \delta)$ such that $g_x(y) = 0$. Then, define $f : U \to V = (y_0 - \delta, y_0 + \delta)$ defined as $x \xmapsto{f} y$ where $y \in V$ is the unique number such that $g_y(x) = 0$. Then, $F(x, f(x)) = 0, \forall x \in U$, which completes the proof apart from showing that f is of class C^r on U, which can be found in M&H. ☐

Note that, if f is differentaible on U, then we can compute $f'(x)$ by seeing that

$$0 = D_1F(x, f(x)) + f'(x)D_2F(x, f(x)) \implies f'(x) = -\frac{D_1F(x, f(x))}{D_2F(x, f(x))}.$$

Example 27.3. *Consider $\frac{\mathrm{d}x}{\mathrm{d}t} = \mu x - x^3$, where $\mu \in \mathbb{R}$ is constant. Then, equlibria satisfy $0 = \mu x - x^3 = F(u, x)$. Does this implicitly define $x = f(u)$?*

We see that $\frac{\partial F}{\partial x} = \mu - 3x^2$, which we can use to discover this model's bifurcating equilibrium.

Before we introduce the general implicit function theorem, a point about notation: for a point in $\mathbb{R}^n \times \mathbb{R}^m$, we write $(x, y) = (x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m)$ and $(x_0, y_0) = (x_1^{(0)}, \ldots, x_n^{(0)}, y_1^{(0)}, \ldots, y_m^{(0)})$.

Theorem 27.2. *(Implicit function theorem, not restricted). Let $F : \Omega \subseteq \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be of class C^r where $r \geq 1$ on the open set Ω. Then, let $(x_0, y_0) \in \Omega$ and $F(x_0, y_0) = 0 \in \mathbb{R}^m$. Then, if*

$$\det \begin{bmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & \vdots & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{bmatrix} \neq 0$$

then there exist open neighborhoods U of x_0 and V of y_0 with $U \times V \subseteq \Omega$ and a unique function $f : U \to V$ such that $F(x, f(x)) = 0$, $\forall x \in U$, and f is of class C^r on U.

28 April 6

Today is midterm review day, but we'll still do some practice examples using the implicit function theorem.

28.1 Implicit Function Theorem Example

Example 28.1. *Consider the bioswitch model (Gardner et al., 2000) defined as*

$$\begin{cases} \mathrm{d}u/\mathrm{d}t = \frac{\alpha_1}{1+v^\beta} - u \\ \mathrm{d}v/\mathrm{d}t = \frac{\alpha_2}{1+u^\gamma} - v \end{cases}$$

where $\alpha_1, \alpha_2, \beta, \gamma$ are real, positive parameters.

Given that $(u,v) = (1,2)$ is an equilibrium if $\alpha_1 = 5, \alpha_2 = 4, \beta = 2, \text{and} \gamma = 2$, is this the only equilibrium in this vicinity? Is it possible to solve for (u,v) uniquely in terms of $\alpha_1, \alpha_2, \beta, \gamma$ near this piont?

Consider $F : \mathbb{R}^4_+ \times \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$(\alpha_1, \beta, \alpha_2, \gamma, u, v) \mapsto \left(\frac{\alpha_1}{1+v^\beta} - u, \frac{\alpha_2}{1+u^\gamma} - v \right).$$

We already know from above that $F(5,2,4,2,1,2) = (0,0)$. Then, we look at

$$\begin{bmatrix} \frac{\partial F_1}{\partial u} & \frac{\partial F_1}{\partial v} \\ \frac{\partial F_2}{\partial u} & \frac{\partial F_2}{\partial v} \end{bmatrix} = \begin{bmatrix} -1 & -\frac{\alpha_1 \beta v^{\beta-1}}{(1+v^\beta)^2} \\ -\frac{\alpha_2 \gamma u^{\gamma-1}}{(1+u^\gamma)^2} & -1 \end{bmatrix}$$

and, if we evaluate this at our point of interest, we get the matrix

$$\begin{bmatrix} -1 & -4/5 \\ -2 & -1 \end{bmatrix}$$

which has non-zero determinant, so the implicit function theorem applies. Then, there exist open neighborhoods U of $(5,2,4,2)$ and V of $(1,2)$ and a unique function $f : U \to V$ such that $F(x, f(x)) = 0$, $\forall x \in U \subseteq \mathbb{R}^4_+$.

Example 28.2. *Continuing to work with the example above, what would happen if we looked at the equilibrium $(u,v) = (1,1)$ if $\alpha_1 = \beta = \alpha_2 = \gamma = 2$? If we look at the same "partial" Jacobian matrix as before, we get*

$$\begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}$$

whose determiant is 0, so we can't apply the implicit function theorem. This is because small perturbations of the parameters can spawn new equilibria.

29 April 8

29.1 Domain Straightening Theorem

Theorem 29.1. *(Domain straightening theorem). Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be of class C^r where $r \geq 1$ on the open set Ω, $x_0 \in \Omega$, $f(x_0) = 0$, $\nabla f(x_0) \neq 0$. Then, there exist open sets $U \subseteq \mathbb{R}^n$, $V \subseteq \Omega \subseteq \mathbb{R}^n$ with $x_0 \in V$ and a C^r-diffeomorphism $h : U \to V$ such that $f(h(x_1, x_2, \dots, x_n)) = x_n$ on U.*

The idea is that a change of variables can effectively straighten out level sets such that $f \circ h$ is simply a projection onto its last coordinate.

Example 29.1. *Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x,y) = \frac{y^2}{2} - \frac{x^2}{2} + \frac{x^4}{4}$. Then, $\nabla f(x,y) = (-x + x^3, y)$, so the critical points are $(0,0)$ and $(\pm 1, 0)$. The domain straightening theorem does not apply to the origin.*

The theorem is especially useful for proving "trapping theorems," and we'll use it to prove Lagrange multipliers. The implicit function theorem with the domain straightening theorem will help us prove a nice result about contrained optimization.

29.2 Lagrange Multipliers

Consider this motivating example: suppose $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(x,y) = xy$. Then, let $K \subseteq \mathbb{R}^2$ as

$$K = \left\{(x,y) : \mathbb{R}^2 : 4x^2 + y^2 \leq 72\right\}$$

. We know that absolute extrema are guaranteed to be achieved because f is continuous and K is compact, but how do we find them?

First, we look for critical points on the interior of K which we do by finding that $\nabla f(x,y) = (y,x)$ so the only critical point is $(0,0) \in \operatorname{int} K$. Then, we want to look at ∂K, the ellipse $4x^2 + y^2 = 72$, which we could do through parametrization, but we'll explore another method now.

This boundary is a level set of $g(x,y) = 4x^2 + y^2 - 72$, so on ∂K, the maximum or minimum of f is subject to the constraint $g(x,y) = 0$.

If we pick some (x_0, y_0) such that $g(x_0, y_0) = 0$ (i.e. a point on the constraint curve), then $\nabla g(x_0, y_0)$ is orthogonal to the constraint curve at (x_0, y_0). But (x_0, y_0) is on some level curve of f and $\nabla f(x_0, y_0)$ is orthogonal to its level curve, so we want $\nabla f(x_0, y_0)$ to be parallel to $\nabla g(x_0, y_0)$ for potential extrema of f restricted to the constraint curve. In other words, we want $\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)$ for some constant λ.

In this case, we have $g(x,y) = 4x^2 + y^2 - 72$ so we get $\nabla f(x,y) = (y,x)$ and $\nabla g(x,y) = (8x, 2y)$. Then, if $\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)$ and $g(x_0, y_0) = 0$, then we know that.

1. $y = 8\lambda x$
2. $x = 2\lambda y$
3. $4x^2 + y^2 = 72$

Then, combining the first two constraints, we get that either

1. $y = 0$, which would imply $x = 0$, which then would not satisfy the third condition.
2. $\lambda = \pm\frac{1}{4}$, which would then lead to the points $(-3,-6), (3,6), (3,-6), (-3,6)$.

Evaluating each of these points, along with $(0,0)$ tells us that $f|_K$ reaches an absolute maximium of 18 at $(3,6)$ and $(-3,-6)$ and an absolute minimum of -18 at $(3,-6)$ and $(-3,6)$.

Let's formulate this approach into a theorem.

Theorem 29.2. *(Lagrange multipliers). Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ and $g : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be of class C^1 on the open set Ω. Then, assume $x_0 \in \Omega$ and $g(x_0) = c$ and $\nabla g(x_0) \neq 0$. Let $S = g^{-1}(\{c\})$ such that the preimage of $\{c\}$ is the level set $g = c$. If $f|_S$ has an extremum at x_0, then $\exists \lambda \in \mathbb{R}$ such that $\nabla f(x_0) = \lambda \nabla g(x_0)$. (WLOG, $c = 0$. Otherwise apply to $\tilde{g} = g - c$.)*

29.3 Integration in More Dimensions

We will now generalize the Riemann integral to multivariate land and will introduce volume integrals, contour integrals, surface integrals, and more that measure various quantities.

For a review of partitions, lower/upper sums/integrals, and the definition of Riemann integrable, see Lecture 13: February 23.

We will generalize those definitions to functions $f : B \subseteq \mathbb{R}^n \to \mathbb{R}$ where B is bounded (not necessarily connected).

Definition 29.1. *Let $a_j < b_j$ for $j = 1, 2, \ldots, n$. The set $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ is called a rectangle or n-cell . The volume of this n-cell is then $\displaystyle\prod_{i=1}^{n}(b_i - a_i)$.*

Definition 29.2. *A partition of $[a_1, b_1] \times \cdots \times [a_n, b_n]$ is a collection of partitions $P = (P_1, P_2, \ldots, P_n)$ where each P_j partitions $[a_j, b_j]$. Each P_j is thus a list $\left\{a_j = c_j^{(0)}, c_j^{(1)}, \ldots, c_u^{(m_j)} = b_j\right\}, m_j \in \mathbb{N}$.*

If each p_j subdivides $[a_j, b_j]$ into m_j subintervals, then P subdivides R into $\prod_{i=1}^{n} m_j$ subrectangles. If R is a rectangle in $\mathbb{R}^n$ and S is a subrectangle induced by the partition P, then define

$$m_S = \inf\left\{f(x) : x \in S\right\} \qquad M_S = \sup\left\{f(x) : x \in S\right\}$$

and let $v(S)$ be the volume of S to get

Definition 29.3. *The upper sum and lower sum of f for P are, respectively*

$$U(f, P) = \sum_S M_S v(S) \qquad L(f, P) = \sum_S m_S v(S)$$

and similarly, if $\mathcal{P}$ is the set of all partitions of R, then the upper integral and lower integral of f over R are, respectively

$$\overline{\int_R} f = \inf_{P \in \mathcal{P}} U(f, P) \qquad \underline{\int_R} f = \sup_{P \in \mathcal{P}} L(f, P)$$

such that Riemann integral on R means that

$$\overline{\int_R} f = \underline{\int_R} f$$

in which case we simply write the integral as $\int_R f$

More generally, if $B \subseteq \mathbb{R}^n$ is a bounded domain, which is not necessarily a rectangle, we can still define $\int_B f$. Let R be any rectangle such that $B \subseteq R$. Define $f_{ext} : R \to \mathbb{R}$ as

$$f_{ext}(x) = \begin{cases} f(x) & \text{if } x \in B \\ 0 & \text{if } x \in R \setminus B. \end{cases}$$

Then, we get that $\int_B f = \int_R f_{ext}$.

30 April 11

30.1 Integrability and Calculations in More Dimensions

Suppose that P is a partition of some rectangle in $\mathbb{R}^n$. A refinement Q of P is a partition such that every subrectangle generated by Q is contained in some subrectangle generated by P.

We can similarly generate the following statements:

1. If Q is a refinement of P, then
$$L(f,P) \le L(f,Q) \le U(f,Q) \le U(f,P).$$
2. Given any partitions P and Q of a rectangle,
$$L(f,P) \le U(f,Q).$$

We previously discussed the Riemann criteria for integrability. That is,

Proposition 30.1. *Let f be bounded on a rectangle $R \subseteq \mathbb{R}^n$. Then, f is integrable on R iff given any $\epsilon > 0$, $\exists$ a partition P such that*
$$U(f,P) - L(f,P) < \epsilon.$$

There is another important criteria for integrability: the Darboux criteria.

Proposition 30.2. *(Darboux). If f is bounded on R, then f is integrable with the integral I iff $\forall \epsilon > 0$, $\exists \delta > 0$ such that every partition P of R into subrectangles of side lengths less than δ yields*
$$\left| \sum_{j=1}^{N} f(x_j)v(S_j) - I \right| < \epsilon$$
where $S_1, S_2, \ldots, S_N$ are subrectangles induced by P and x_j is any representatitve point from S_j.

How do we compute these integrals that we've been talking about? Let's consider rectangles in $\mathbb{R}^2$ first. Let $R = [a,b] \times [c,d]$ and let $f : R \to \mathbb{R}$ be continuous. Since f is continuous, for each fixed $y \in [c,d]$, the function $g_y : [a,b] \to \mathbb{R}$ defined by $g_y(x) = f(x,y)$ is continuous on $[a,b]$. Thus, g_y is integrable on $[a,b]$.

Then, we can partition $[a,b]$ and $[c,d]$ as
$$x_0 = a < x_1 < x_2 < \cdots < x_n = b \quad \text{and} \quad y_0 = c < y_1 < y_2 < \cdots < y_m = d.$$

By Darboux, we can ensure that $\int_a^b f(x,y_j)\,dx$ is as close as needed to $\sum_{i=1}^n f(x_i,y_j)(x_i - x_{i-1})$ by requiring that
$$\max_{1 \le i \le n}(x_i - x_{i-1})$$
is small enough. Then, we get that
$$\int_{[a,b]\times[c,d]} f \approx \sum_{j=1}^{m} \left(\int_a^b f(x,y_j)\,dx \right)(y_j - y_{j-1})$$
which again, by Darboux, is as close as needed to $\int_c^d \left(\int_a^b f(x,y)\,dx \right) dy$.

Formalizing this argument into a rigorous proof is not bad – look at M&H 9.2 for more details.

Theorem 30.1. *(Fubini). Let $R = [a,b]\times[c,d] \subseteq \mathbb{R}^2$ and $f : R \to \mathbb{R}$ is continuous. Then, f is integrable and*

$$\int_R f = \int_c^d \left(\int_a^b f(x,y)\, dx \right) dy = \int_a^b \left(\int_c^d f(x,y)\, dy \right) dx.$$

We can also generalize this to functions that are not continuous or that have $n \geq 2$ dimensions.

Example 30.1. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x,y) = xe^{xy}$ on $R = [1,2]\times[3,4]$. Then, we get*

$$\int_R f = \int_1^2 \int_3^4 xe^{xy}\, dy\, dx = \int_1^2 e^{xy}\Big|_{y=3}^{y=4} dx = \int_1^2 e^{4x} - e^{3x}\, dx.$$

We can also define $\int_B f$ as an iterated integral for certain bounded B that are not rectangles.

Example 30.2. *Let $f : K \subseteq \mathbb{R}^2 \to \mathbb{R}$ be continuous on the domain K as the triangle with vertices $(0,0),(2,4),(0,4)$. If f is continuous, we can evaluate*

$$\int_0^2 \int_{2x}^4 f(x,y)\, dy\, dx = \int_0^4 \int_0^{y/2} f(x,y)\, dx\, dy.$$

30.2 Change of Variables

We will now formalize the process of "u-substitution" to higher-dimensions.

Recall the fundamental theorem of calculus.

Theorem 30.2. *Let $f : [a,b] \to \mathbb{R}$ be integrable and let $F : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). If $F' = f$ on (a,b) then $\int_a^b f(x)\, dx = F(b) - F(a)$.*

Proof. Use Riemann criterion. Let $\epsilon > 0$ be given and choose a partition $P = \{x_0, x_1, \cdots, x_n\}$ such that $U(F',P) - L(F',P) < \epsilon$. Apply MVT to each $[x_{j-1}, x_j]$. Then, we see that

$$F(x_j) - F(x_{j-1}) = F'(c_j)(x_j - x_{j-1}).$$

Let $m_j' = \inf_{[x_{j-1},x_j]} F'(x)$ and $M_j' = \sup_{[x_{j-1},x_j]} F'(x)$. Then,

$$m_j'(x_j - x_{j-1}) \leq F(x_j) - F(x_{j-1}) \leq M_j'(x_j - x_{j-1}).$$

If we sum over all j, we get that $L(F',P) \leq F(b) - F(a) \leq U(F',P)$. We also know that $L(F',P) \leq \int_a^b F'(x)\, dx \leq U(F',P)$. Together, these imply that

$$\left| \int_a^b f(x)\, dx - [F(b) - F(a)] \right| < \epsilon$$

but $\epsilon > 0$ is arbitrary. □

We can also prove the following proposition.

Proposition 30.3. *Let $f : [a,b] \to \mathbb{R}$ and defined $F(x) = \int_a^x f(t)\, dt$. Then, F is differentiable on (a,b) and $F' = f$.*

Proposition 30.4. *Let $I = [a,b]$ and suppose $\varphi : I \to \mathbb{R}$ is of class C^1. If f is continuous on the interval J containing $\varphi(I)$,*

$$\int_a^b f(\varphi(t))\varphi'(t)\, dt = \int_{\varphi(a)}^{\varphi(b)} f(x)\, dx.$$

Proof. Let $c = \varphi(a)$ and $d = \varphi(b)$. Then, we know that J contains the interval with endpoints c and d, though we do not know the relation between c and d. Since f is continuous on J, define $F : J \to \mathbb{R}$ and $g : I \to \mathbb{R}$ according to

$$F(u) = \int_c^u f(u)\,du \quad \text{and} \quad g(t) = F(\varphi(t)).$$

By the chain rule and by FTC, we get that

$$g'(t) = f(\varphi(t))\varphi'(t), \forall t \in I.$$

So, then, we get

$$\int_a^b f(\varphi(t))\varphi'(t)\,dt = \int_a^b g'(t)\,dt = g(b) - g(a).$$

But note that $g(a) = F(\varphi(a)) = F(c) = 0$ by definition of F, so

$$g(b) = F(\varphi(b)) = F(d) = \int_c^d f(u)\,du.$$

$\square$

Proposition 30.5. *Let $I = [a,b]$ and $\varphi : I \to \mathbb{R}$ be of class C^1 and $\varphi'(t) \neq 0$, $\forall t \in I$. If f is continuous on $\varphi(I)$, then*

$$\int_a^b f(\varphi(t))\,dt = \int_{\varphi(a)}^{\varphi(b)} f(x)\left(\varphi^{-1}\right)'(x)\,dx.$$

31 April 13

Today we will generalize change of variables to functison of many variables. In what follows, we will identify $Df(x)$ with its standard matrix.

31.1 Change of Variables in More Dimensions

Theorem 31.1. *Let $\varphi : U \to W$ be a C^1 diffeomorphism between open subsets U and W of $\mathbb{R}^n$. Let $R \subseteq U$ be a rectangle and $f : W \to \mathbb{R}$ be bounded and integrable. Then, $(f \circ \varphi) \cdot \left|\det D\varphi\right|$ is integrable on R and*

$$\int_R (f \circ \varphi) \cdot \left|\det D\varphi\right| = \int_{\varphi(R)} f.$$

We won't prove this theorem, but as a non-rigorous explanation, consider the integrals' rough evaluations as

$$\int_S (f \circ \varphi) \cdot \left|\det D\varphi\right| \approx f(\varphi(x_0))\left|\det D\varphi(x_0)\right| \cdot \mathrm{vol}(S).$$

$$\int_{\varphi(S)} f \approx f(\varphi(x_0)) \cdot \mathrm{vol}(\varphi(S))$$

where $\left|\det D\varphi(x_0)\right| \cdot \mathrm{vol}(S)$ roughly takes the place of $\mathrm{vol}(\varphi(S))$.

Example 31.1. *Let $f(x,y) = x$ and let K be the parallelogram region with vertices $(0,0),(0,1),(2,2),(2,3)$. Then, to evaluate $\int_K f$, we can define $\varphi(e_1) = e_2$ and $\varphi(e_2) = 2e_1 + 2e_2$ such that φ maps the unit square onto K. Then, we see that*

$$\begin{bmatrix} x \\ y \end{bmatrix} \overset{\varphi}{\longmapsto} \begin{bmatrix} 0 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2y \\ x + 2y \end{bmatrix}$$

and that $\varphi(R) = K$. We can also notice that $\det D\varphi = \det \varphi = -2$, everywhere. So,

$$\int_K f = \int_R (f \circ \varphi) \cdot \left|\det D\varphi\right| = \int_R (2y)|-2| = \int_R 4y = \int_0^1 \int_0^1 4y \; dx \; dy = 2.$$

31.2 Polar, Cylindrical, and Spherical Coordinates

There are some nice change of variables that exploit geometry and radial symmetry.

For polar coordinates, we have the transformation

$$\begin{bmatrix} r \\ \theta \end{bmatrix} \overset{\varphi}{\longmapsto} \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix}$$

on the domain $U = \left\{(r,\theta) : r > 0 \text{ and } 0 < \theta < 2\pi\right\}$. We can't use a larger open set without ruining φ as a C^1 diffeomorphism.

If we extend the θ interval, φ will no longer have a differentiable inverse, violating injectivity.

However, If f is continuous, then the behavior of f above the positive x-axis won't affect the integral because there is "no 2D" area for that piece of the axis.

Example 31.2. *Evaluate $\int_K f$ where $f(x,y) = x$ and K is the region formed by the are above the line $y = x$ of the intersection of the circle centered at the origin of radius 2 and the complement of the circle centered at the origin of radius 1.*

Then, we can say that $K = \varphi(R)$ where $R = \left\{(r,\theta) : 1 \le r \le 2, \frac{\pi}{4} \le \theta \le \frac{5\pi}{4}\right\}$.

We can also calculate that, since φ is the same transformation as above,

$$D\varphi(r,\theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} \implies \left|\det D\varphi(r,\theta)\right| = r$$

so then

$$\int_{\varphi(R)=K} f = \int_R (f \circ \varphi)\left|\det D\varphi\right| = \int_{\pi/4}^{5\pi/4} \int_1^2 f(r\cos\theta, r\sin\theta) r \, dr \, d\theta = \int_{\pi/4}^{5\pi/4} \int_1^2 r^2 \cos\theta \, dr \, d\theta.$$

Now, we will discuss spherical coordinates. Recall the change of variables transformation defined as

$$\begin{bmatrix} \rho \\ \phi \\ \theta \end{bmatrix} \xmapsto{\psi} \begin{bmatrix} \rho\sin\phi\cos\theta \\ \rho\sin\phi\sin\theta \\ \rho\cos\phi \end{bmatrix}$$

on $U = \left\{(\rho,\phi,\theta) : \rho > 0, 0 < \phi < \pi, 0 < \theta < 2\pi\right\}$. Then, ψ is a C^1-diffeomorphism onto

$$\mathbb{R}^3 \setminus \left\{(x,y,z) : y = 0, x \ge 0\right\}.$$

But the excluded set has no volume. Attempting to include $\theta = 0$ in the definition of U, though, would cause ψ^{-1} to be discontinuous in the half plane.

Example 31.3. *Evaluate $\int_K f$ where $f(x,y,z) = z\sqrt{x^2+y^2+z^2}$ and*

$$K = \left\{(x,y,z) \in \mathbb{R}^3 : 1 \le x^2+y^2+z^2 \le 4 \text{ and } z \le 0\right\}.$$

Then, we see that if

$$R = \left\{(\rho,\phi,\theta) : 1 \le \rho \le 2, \frac{\pi}{2} \le \phi \le \pi, 0 \le \theta \le 2\pi\right\}$$

then $psi(R) = K$. We can also calculate that $\left|\det D\psi(\rho,\phi,\theta)\right| = \rho^2\sin\phi$. Together, we can now get that

$$\int_K f = \int_R (f \circ \psi)\left|\det D\psi\right| = \int_0^{2\pi} \int_{\pi/2}^{\pi} \int_1^2 (\rho\cos\theta)(\rho)(\rho^2\sin\phi) \, d\rho \, d\phi \, d\theta.$$

32 April 15

Today we will discuss line integrals and Green's theorem.

Let $\gamma : [a,b] \to \mathbb{R}^n$ be of class C^1 and $\gamma'(t) \neq 0$ on $[a,b]$. That is, let γ be regular. Such conditions then ensure that

1. $\int_a^b \left\|\gamma'(t)\right\| dt$ is finite.
2. We avoid non-regular curves like $\gamma(t) = (t^3, t^2)$ on $[-1,1]$, which will be bad if we wish to formulate integral theorems on regions whose boundaries are smooth curves.

32.1 Line Integrals for Scalar-Valued Functions

Definition 32.1. *Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be continuous on the open set Ω. Then, let C be a curve contained in Ω and have C admit a C^1 regular parameterization $\gamma : [a,b] \to \Omega$. Then, the line integral $\int_C f$ is defined as*

$$\int_a^b f(\gamma(t))\left\|\gamma'(t)\right\| dt.$$

We can interpret this quantity as follows: imagine that C is a wire of variable density with f representing the density at each point. Then, the line integral will give us the mass of the wire.

We arrive at the above definition, then, by estimating

$$\gamma(t_j + \delta t) \approx \gamma(t_j) + \gamma'(t_j)\delta t.$$

Then, the lengt hof the segment is about $\left\|\gamma'(t_j)\right\|\delta t$. If the density is about $f(\gamma(t_j))$, then the mass of the segment is approximately $f(\gamma(t_j))\left\|\gamma'(t_j)\right\|\delta t$. Sum over all the segments to yield the definition from above.

We show for homework that the value of $\int_C F$ is independent of the choice of a C^1 regular parameterization.

32.2 Line Integrals for Vector-Valued Functions

Definition 32.2. *Suppose $F : \Omega : \mathbb{R}^n \to \mathbb{R}^n$ be continuous on open set Ω. Then, let C be an oriented curve contained in Ω that admits a C^1 regular parameterization $\gamma : [a,b] \to \Omega$. The line integral $\int_C F$ is defined as*

$$\int_a^b \langle F(\gamma(t)),\ \gamma'(t)\rangle\, dt$$

with the usual inner product.

We can interpret this integral as the work done by the vector field F as a particle moves along C from $\gamma(a)$ to $\gamma(b)$.

Example 32.1. *Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by*

$$\begin{bmatrix} x \\ y \end{bmatrix} \overset{F}{\longmapsto} \begin{bmatrix} -y \\ x \end{bmatrix}$$

and let C be a line segment from $(0,0)$ to $(-3,4)$. Use the parameterization $\gamma(t) = (-3t, 4t), t \in [0,1]$.

To evaluate $\int_C F$, we first calculate that $F(\gamma(t)) = (-4t, -3t)$ and that $\gamma'(t) = (-3,4), \forall t \in [0,1]$. So,

$$\int_C F = \int_0^1 \langle(-4t,-3t),(-3,4)\rangle\, dt = 0.$$

This result makes sense if we realize that $F(\gamma(t))$ is always orthogonal to $\gamma'(t)$, so F does no work as a particle travels along C.

There are, however, other paths we can take from $(0,0)$ to $(-3,4)$ that allow F to do some work. Consider $\gamma_{new}(t) = (3\cos t - 3, 4\sin t), 0 \le t \le \pi/2$. Then, $\gamma_{new}(0) = (0,0)$ and $\gamma_{new}\left(\frac{\pi}{2}\right) = (-3,4)$. We also calculate that $\gamma_{new}'(t) = (-3\sin t, 4\cos t)$. Then, we get

$$\int_{C_{new}} F = \int_0^{\pi/2} \langle(-4\sin t, 3\cos t - 3), (-3\sin t, 4\cos t)\rangle\, dt = 12((\pi/2)-1) > 0.$$

Notice that

1. *The integral is the same regardless of the C^1, regular parameterization we choose, as long as we respect orientation.*
2. *Sometimes it is okay to allow $\gamma'(t) = 0$ at a single point.*
3. *If $-C$ denotes C but with the reverse orientation, then*

$$\int_{-C} F = -\int_C F.$$

32.3 Conservative Vector Fields

In the last example, $\int_C F$ and $\int_{C_new} F$ differed, so the line integral was path dependent. Are there conditions for path independence?

Definition 32.3. *A vector field $F : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is conservative if $\exists$ a C^1 function $f : \Omega \to \mathbb{R}$ such that $F = \nabla f$ on Ω. Then, we call f a potential function for F.*

Proposition 32.1. *If $F : \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}^2$ is of class C^1 and F is conservative then $\frac{\partial F_1}{\partial y} = \frac{\partial F_2}{\partial x}$.*

Proof. If F is conservative, then $F = \nabla f$, where f must be C^2 since F is C^1. Also, we know that $F = \left(\partial f \big/ \partial x, \partial f \big/ \partial y\right)$. But, we know that $f_{xy} = f_{yx}$ by Clairaut-Schwarz, so $(F_1)_y = (F_2)_x$. □

In the earlier example, for instance, $\frac{\partial F_1}{\partial y} = -1 \neq 1 = \frac{\partial F_2}{\partial x}$, so F was not conservative.

Proposition 32.2. *(FTC for line integrals). Let $F : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^n$ have a C^1 potential function, f, $f : \Omega \to \mathbb{R}$. Then, given points $A, B \in \Omega$ and any C^1 regular oriented curve C in Ω from A to B, we have*

$$\int_C F = f(B) - f(A).$$

Proof. Let $\gamma : [a,b] \to \gamma$ be a C^1 regular parameterization with $\gamma(a) = A$ and $\gamma(b) = B$. Assume C in Ω. Then,

$$\begin{aligned}\int_C F &= \int_a^b \langle F(\gamma(t)), \gamma'(t)\rangle\, dt \\ &= \int_a^b \langle \nabla f(\gamma(t)), \gamma'(t)\rangle\, dt \\ &= \int_a^b \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\, dt \\ &= f(\gamma(b)) - f(\gamma(a)) \\ &= f(B) - f(A)\end{aligned}$$

□

Note that:

1. If F has a potential function on an open set, the above result holds if C is piecewise, C^1 regular (useful if C is a rectangle).
2. The FTC tells us that, if F is conservative, then $\int_C F$ is independent of path (respecting orientation) from A to B.
3. If C is piecewise C^1 regular, simple, closed curve, then for conservative F, $\oint_C F = 0$.

33 April 18

33.1 Green's Theorem

Recall that we previously defined a Jordan curve as a simple, continuous, closed curve in the plane $\mathbb{R}^2$. We can think of a Jordan curve as the image of a continuous function $\gamma : [0,1] \to \mathbb{R}^2$ such that $\gamma(0) = \gamma(1)$ and $\gamma|_[0,1)$ is injective.

Theorem 33.1. *(Jordan). If C is a Jordan curve in $\mathbb{R}^2$, then $\mathbb{R}^2 \setminus C$ consists of precisely two connected components: a bounded "interior" component and an unbounded "exterior" component.*

Definition 33.1. *Let C be an oriented Jordan curve with parameterization $\gamma : [0,1] \to \mathbb{R}^2$. We say that C is positively oriented if the bounded interior component $\mathbb{R}^2 \setminus C$ is always to the "left" of the curve if we follow γ in the direction of increasing t.*

If the curve is C^1, regular, then it's not too bad to explain what "to the left" means. For instance, if $\gamma(t) = (x(t), y(t))$ is a C^1, regular parameterization of C, then $\gamma'(t) = (x'(t), y'(t)) \neq 0$ is a velocity vector. Then, $(-y'(t), x'(t))$ points to the "left."

Theorem 33.2. *(Green). Let $F : \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$, where Ω is open, be of class C^1. Then, let Γ be a piecewise C^1 regular Jordan curve such that Γ and its bounded "interior" are in Ω. If Γ is positively oriented, then*

$$\oint_\Gamma F = \iint_D \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}$$

where D is the region consisting of Γ and its interior.

Example 33.1. *Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the function defined as*

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x+y \\ -x+y \end{bmatrix}.$$

Let Γ be the unit circle, positively oriented, which we can parameterize as $\gamma(t) = (\cos t, \sin t), t \in [0, 2\pi]$. Then, we can calculate the left side of Green's theorem as described above as

$$\begin{aligned} \oint_\Gamma F &= \int_0^{2\pi} \langle F(\gamma(t)), \gamma'(t) \rangle \, dt \\ &= \int_0^{2\pi} \left\langle \begin{bmatrix} \cos t + \sin t \\ -\cos t + \sin t \end{bmatrix}, \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix} \right\rangle dt \\ &= -2\pi \end{aligned}$$

and the right side of Green's theorem as described as

$$\begin{aligned} \iint_D \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} &= \iint_{\overline{B(0,1)}} -1 - 1 \\ &= -2(\text{area of } D) \\ &= -2\pi \end{aligned}$$

How do we interpret the term $\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}$ as included in Green's theorem?

Somehow, this term is associated with rotation. If $\frac{\partial F_1}{\partial y}$ is negative, for instance, if we placed a pinwheel in the vector field, we would expect the field to spin the pinwheel counterclockwise. Similarly, if $\frac{\partial F_2}{\partial x}$ is positive, we would also expect the field to spin the pinwheel counterclockwise.

33.2 Surface Integrals for Scalar-Valued Functions

The following, like much of our discussion earlier, will be relatively informal. We need a way to generalize some of the notions from curves and line integrals up to surfaces.

Informally, an n-manifold is an object for which each point has a neighborhood (restricted to the object itself) that is homeomorphic to $\mathbb{R}^n$.

Example 33.2. *Consider* $S^1 = \left\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\right\}$*. Pick* $(x_0, y_0) \in S$ *and assume that* $|x_0| < 1$*. Pick* $\epsilon > 0$ *such that* $(x_0 - \epsilon, x_0 + \epsilon) \subseteq (-1, 1)$*. Project this interval onto the x-axis and translate the projection to* $(-\epsilon, \epsilon)$*. Then, let* $f : (-\epsilon, \epsilon) \to \mathbb{R}$ *be*

$$f(x) = \frac{x}{\epsilon^2 - x^2}$$

such that $f \circ T \circ P$ *is a homeomorphism of the section of* S^1 *onto* $\mathbb{R}$*. Thus,* S^1 *is a* 1*-manifold.*

We will, in the following discussion, primarily focus on 2-manifolds (surfaces) in $\mathbb{R}^3$.

If $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation of rank 2, then the image of T is a 2-manifold. Consider a surface S in $\mathbb{R}^3$ parameterized by $r : A \subseteq \mathbb{R}^2 \to \mathbb{R}^3$ where A is open and connected. We can extend the concept of C^1,regular parameterizations to such surfaces.

Pick some $(s_0, t_0) \in A$ and consider the corresponding curves $C_{s_0}(t) = r(s_0, t)$ and $C_{t_0}(s) = r(s, t_0)$ such that we vary t in the former and vary s in the latter. These are the grid curves of r. Near (s_0, t_0), these curves make sense.

We want assurance that the "velocity" curves $C'_{s_0}(t_0)$ and $C'_{t_0}(s_0)$ are nonzero and linearly independent. Then, we will get a a well-defined tangent plane at (s_0, t_0). We want r to map small, open neighborhoods of (s_0, t_0) to 2-manifolds.

Definition 33.2. (s_0, t_0) *as described above is called a regular point . That is, the grid curves are* C^1 *regula, and the tangent vectors are linearly independent.*

How can we now use the above groundwork to define some notion of "intensity" over some surface or to measure surface area?

We need to be able to calculate areas of image sof small rectangles in A. To do so, we want to estimate the area of the image of $[s_0, s_0 + \Delta s] \times [t_0, t_0 + \Delta t]$ udner r.

Definition 33.3. *Given vectors* $u, v \in \mathbb{R}^3$*, their cross product is the vector*

$$u \times v = \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix}.$$

One trick for remembering how to compute the cross product is to find the "determinant" of the matrix

$$\begin{bmatrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix}.$$

Also note three important properties of the cross product:

1. $u \times v$ is orthogonal to both u and v.
2. $u \times v = -(v \times u)$.
3. $\|u \times v\| = \|u\|\|v\| \sin\theta$, where $\theta \in [0, \pi]$ is the angle between u and v. Also, $\|u \times v\|$ is equal to the area of the parallelogram formed by the two vectors.

Now, instead of writing $C_s'(t)$ or $C_t'(s)$, write

$$r_s(s,t) = \begin{bmatrix} \partial r_1 \big/ \partial s \\ \partial r_2 \big/ \partial s \\ \partial r_3 \big/ \partial s \end{bmatrix} \quad \text{and} \quad r_t(s,t) = \begin{bmatrix} \partial r_1 \big/ \partial t \\ \partial r_2 \big/ \partial t \\ \partial r_3 \big/ \partial t \end{bmatrix}$$

Notice that these are the columsn of $Dr(s,t)$.

For small Δs and Δt, we have that

$$\begin{aligned} r(s_0 + \Delta s, t_0) &\approx r(s_0, t_0) + \Delta s r_s(s_0, t_0) \\ r(s_0, t_0 + \Delta t) &\approx r(s_0, t_0) + \Delta t r_t(s_0, t_0) \end{aligned}$$

which implies that, if $r_s(s_0, t_0)$ and $r_t(s_0, t_0)$ are linearly independent, then the area of the rectangle is approximately $\left\| r_s(s_0, t_0) \times r_t(s_0, t_0) \right\| \Delta s \Delta t$.

Definition 33.4. *Let $\Omega \subseteq \mathbb{R}^2$ be open, connected, and bounded. Let S be the image of Ω under some C^1, regular $r : \Omega \to \mathbb{R}^3$. If f is bounded, continuous, and real-valued on some open subset of $\mathbb{R}^3$ that contains $r(\Omega)$, then the surface integral*

$$\int_S f = \iint_\Omega f(r(s,t)) \| r_s \times r_t \|.$$

Note that, if $t \equiv 1$, the integral is equal to the surface area of S.

34 April 20

Last time, we defined the scalar surface integral $\int_S f$ where $f : \mathbb{R}^3 \to \mathbb{R}$ is ctontinuous and S is a surface in $\mathbb{R}^3$ that admits a C^1 regular parameterization. Now, we will define the surface integral $\int_S F$ where $F : \Omega \subseteq \mathbb{R}^3 \to \mathbb{R}^3$ is a continuous vector field.

34.1 Surface Integrals for Vector-Valued Functions

For motivation, suppose that a fluid in $\mathbb{R}^3$ has constant unit density and let $F : \mathbb{R}^3 \to \mathbb{R}$ be a velocity field for fluid particles. Given a patch of surface S, we want to compute the mass flux across S; that is, the rate of mass flow per unit time per unit area. But we need to specify in which direction the fluid is moving across the surface – we need to orient S.

Pick a unit vector N normal to S to orient S. Pick a parameterization of S, $r : A \subseteq \mathbb{R}^2 \to \mathbb{R}^3$, and a point $(s_0, t_0) \in A$. Now, assume the parameterization is C^1, regular. Then, $r_s(s_0, t_0)$ and $r_t(s_0, t_0)$ are linearly independent. For small Δs, Δt, the vectors $r_s(s_0, t_0)\Delta s$ and $r_t(s_0, t_0)\Delta t$ approximately span a small patch of S with total area approximately equal to $\left\|r_s(s_0, t_0) \times r_t(s_0, t_0)\right\| \Delta s \Delta t$.

But $r_s(s_0, t_0) \times r_t(s_0, t_0)$ is orthogonal to the surface at (s_0, t_0), so we get that

$$N = \pm \frac{r_s(s_0, t_0) \times r_t(s_0, t_0)}{\left\|r_s(s_0, t_0) \times r_t(s_0, t_0)\right\|}.$$

We also now that $F(r(s_0, t_0))$ is the fluid velocity at $r(s_0, t_0)$. So, the mass that crosses a small patch per unit time is approximately $(F \cdot N)\|r_s \times r_t\|\Delta s \Delta t$.

Definition 34.1. *Suppose that S is a bounded surface that admits a C^1 regular parameterization $r : A \subseteq \mathbb{R}^2 \to \mathbb{R}^3$ where A is connected. Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be continuous. If S is oriented such that $r_s \times r_t$ has the same direction as N, the unit normal vector, then the surface flux integral is defined as*

$$\int_S F = \int_S F \cdot n = \iint_A F(r(s,t)) \cdot \left(\frac{\partial r}{\partial s} \times \frac{\partial r}{\partial t}\right)$$

which is now an area integral.

Example 34.1. *Let $S = \partial B(0,1) \subseteq \mathbb{R}^3$, with normal vectors pointing inward. Then, if we hope to evaluate $\int_S F$ where F is defined as*

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \xmapsto{F} \begin{bmatrix} -x/\rho^2 \\ -y/\rho^2 \\ -z/\rho^2 \end{bmatrix}$$

where $\rho^2 = x^2 + y^2 + z^2$. Then, if $A = [0, \pi] \times [0, 2\pi]$, we can parameterize S with $r : A \to \mathbb{R}^3$ defined as

$$\begin{bmatrix} \phi \\ \theta \end{bmatrix} \xmapsto{r} \begin{bmatrix} \sin\phi\cos\theta \\ \sin\phi\sin\theta \\ \cos\phi \end{bmatrix}.$$

We then get that

$$r_\phi = \begin{bmatrix} \cos\phi\cos\theta \\ \cos\phi\cos\theta \\ -\sin\phi \end{bmatrix} \quad \text{and} \quad r_\theta = \begin{bmatrix} -\sin\phi\sin\theta \\ \sin\phi\cos\theta \\ 0 \end{bmatrix} \implies r_\phi \times r_\theta = \begin{bmatrix} \sin^2\phi\cos\theta \\ \sin^2\phi\sin\theta \\ \cos\phi\sin\phi \end{bmatrix}.$$

To check for orientation, let us test the point on the equation with $\phi = \frac{\pi}{2}$ *and* $\theta = 0$. *We find that, here,* $r_\phi, r_\theta = (1,0,0)$, *which suggests that we have the wrong orientation. Instead, set* N *to be*

$$N = -\frac{(r_\phi \times r_\theta)}{\|r_\phi \times r_\theta\|}.$$

Finally, we can calculate

$$\int_{\partial B(0,1)} F = \int_0^\pi \int_0^{2\pi} F(r(\phi,\theta)) \cdot (-(r_\phi \times r_\theta))\, d\theta\, d\phi.$$

34.2 Divergence Theorem

In the pset, we described the divergence of a C^1 vector field $F : \mathbb{R}^2 \to \mathbb{R}^2$ as

$$(\text{div } F)(x_0, y_0) = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y}$$

which came from

$$\lim_{\epsilon \to 0^+} \frac{1}{\pi \epsilon^2} \oint_{C_\epsilon} F \cdot N$$

where N is the unit outward unit and C_ϵ is the circle of radius ϵ with the center (x_0, y_0).

Example 34.2. *Consider* $F : \mathbb{R}^2 \to \mathbb{R}^2$ *defined as*

$$\begin{bmatrix} x \\ y \end{bmatrix} \overset{F}{\longmapsto} \begin{bmatrix} x \\ y \end{bmatrix}.$$

We find that div $F = 1 + 1 = 2$, *which makes sense considering how the vector field disperses from the origin.*

Example 34.3. *Consider* $F : \mathbb{R}^2 \to \mathbb{R}^2$ *defined as*

$$\begin{bmatrix} x \\ y \end{bmatrix} \overset{F}{\longmapsto} \begin{bmatrix} -x \\ -y \end{bmatrix}.$$

Then, div $F = -2$.

Example 34.4. *Consider* $F : \mathbb{R}^2 \to \mathbb{R}^2$ *defined as*

$$\begin{bmatrix} x \\ y \end{bmatrix} \overset{F}{\longmapsto} \begin{bmatrix} -y \\ x \end{bmatrix}.$$

Then, div $F = 0$.

In $\mathbb{R}^3$, let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be C^1 and let $(x_0, y_0, z_0) \in \mathbb{R}^3$. Then, the divergence of F at this point is

$$\lim_{\epsilon \to 0^+} \frac{1}{\frac{4}{3}\pi\epsilon^3} \int_{\partial B_\epsilon} F \cdot N = \left(\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} \right)\Bigg|_{(x_0, y_0, z_0)}$$

where N is again the outward unit normal vector.

Theorem 34.1. *(2D Divergence theorem). Let* $\overline{\Omega} \in \mathbb{R}^2$ *be compact and connected, whose bondary* $\partial\overline{\Omega}$ *is a Jordan curve that admits a piecewise* C^1 *regular parameterization, and let* F *be* C^1 *on an open set containing* $\overline{\Omega}$. *Then,*

$$\oint_{\partial\overline{\Omega}} F \cdot N = \iint_{\overline{\Omega}} \text{div } F$$

where N *is the outward unit normal.*

Proof. We will provide a proof for the rectangle $\overline{\Omega} = [a,b] \times [c,d]$. An easy and effective way of parameterizing the edges for this proof is as follows:

$$\gamma_R(t) = \begin{bmatrix} b \\ t \end{bmatrix}, c \le t \le d, \ \gamma_L(t) = \begin{bmatrix} a \\ c+d-t \end{bmatrix}, c \le t \le d,$$
$$\gamma_T(t) = \begin{bmatrix} a+b-t \\ d \end{bmatrix}, a \le t \le b, \ \gamma_B(t) = \begin{bmatrix} t \\ c \end{bmatrix}, a \le t \le b$$

from which we get

$$\int_R F \cdot N = \int_c^d F_1(b,t)\, dt, \ \int_B F \cdot N = -\int_a^b F_2(t,c)\, dt.$$

Evaluating the line integrals for the left side, we get, with the help of a change of variables,

$$\begin{aligned}\int_L F \cdot N &= \int_c^d -F_1(a, c+d-t)\, dt \\ &= \int_d^c F_1(a,u)\, du \\ &= -\int_c^d F_1(a,u)\, du\end{aligned}$$

and similarly

$$\int_T F \cdot N = \int_a^b F_2(u,d)\, du.$$

Then, combining sides, we get that

$$\begin{aligned}\int_{\partial\overline{\Omega}} F \cdot N &= \int_R F \cdot N + \int_L F \cdot N + \int_B F \cdot N + \int_T F \cdot N \\ &= \int_c^d F_1(b,t)\, dt - \int_c^d F_1(a,t)\, dt - \int_a^b F_2(s,c)\, ds + \int_a^b F_2(s,d)\, ds \\ &= \int_c^d F_1(b,t) - F_1(a,t)\, dt + \int_a^b F_2(s,d) - F_2(s,c)\, ds \\ &= \int_c^d \int_a^b \frac{\partial F_1}{\partial s}\, ds\, dt + \int_a^b \int_c^d \frac{\partial F_2}{\partial t}\, dt\, ds \\ &= \int_c^d \int_a^b \frac{\partial F_1}{\partial s}\, ds\, dt + \int_c^d \int_a^b \frac{\partial F_2}{\partial t}\, ds\, dt \\ &= \int_c^d \int_a^b \frac{\partial F_1}{\partial s} + \frac{\partial F_2}{\partial t}\, ds\, dt \\ &= \iint_{\overline{\Omega}} \operatorname{div} F.\end{aligned}$$

□

Note that, if $F : \mathbb{R}^3 \to \mathbb{R}^3$ is C^1, we often write $\nabla \cdot F$ instead of div F. This notation becomes more intuitive if we consider $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$ and $\nabla \cdot F$ is computed like a dot product.

Theorem 34.2. *Let $M \subseteq \mathbb{R}^3$ be a compact 3-manifold with boundary ∂M. Let N be the outward unit normal on ∂M. Let $F : \Omega \subseteq \mathbb{R}^3 \to \mathbb{R}$ be of class C^1 on some open Ω containing M. Then,*

$$\int_M \nabla \cdot F = \oint_{\partial M} F \cdot N.$$

Example 34.5. *Let $M \subseteq \mathbb{R}^3$ be the surface defined by $x^2 + y^2 \leq z \leq 4$ and $F : \mathbb{R}^3 \to \mathbb{R}^3$ defined as*

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \overset{F}{\longmapsto} \begin{bmatrix} x^2 \\ -z \\ y \end{bmatrix}.$$

Then, we find that $\nabla \cdot F(x,y,z) = 2x$, so with ∂M oriented outward, we get

$$\int_{\partial M} F = \int_{\partial M} F \cdot N = \int_M \nabla \cdot F = \int_M 2x = \int_0^{2\pi} \int_0^2 \int_{r^2}^4 2r^2 \cos\theta \; dz \; dr \; d\theta = 0$$

35 April 22

We learned earlier that, under Green's theorem, if C is a piecewise C^1 regular Jordan curve, Ω is the interior of C, and $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a C^1 vector field, then

$$\oint_C F = \iint_R \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}$$

assuming C positively oriented.

35.1 Stokes' Theorem

Now we will generalize this result to three dimensions. Assume $F : \mathbb{R}^3 \to \mathbb{R}^3$ is C^1.

Definition 35.1. *The curl of F at (x_0, y_0, z_0) is the vector*

$$\begin{bmatrix} (F_3)_y - (F_2)_z \\ (F_1)_z - (F_3)_x \\ (F_2)_x - (F_1)_y \end{bmatrix}$$

with all partials evaluated at (x_0, y_0, z_0).

We will notate this as $(\text{curl } F)(x_0, y_0, z_0)$ or $\nabla \times F(x_0, y_0, z_0)$. The second notation makes sense if we consider the cross product of the gradient operator (taken as a vector) and the function (also taken as a vector).

To interpret this value, imagine putting a pinwheel in a fluid with a velocity field F. Let N be a fixed axis of rotation. Then, if N is a unit vector, then $(\nabla \times F) \cdot N$ measures the tendency of the wheel to rotate.

If $(\nabla \times F) \cdot N > 0$, then we'll see counterclockwise rotation if we look down the axis N.

The following theorem statement is almost extraordinarily non-technical.

Theorem 35.1. *(Stokes) Assume that*

1. *Let $S \subseteq \mathbb{R}^3$ be a compact, connected, 2-manifold that admits a C^1-regular parameterization.*
2. *The "boundary" of S is a closed curve that is piecewise C^1-regular. By "boundary," imagine that we paint one side of S blue and the other side red. If an ant crawls on S, the "boundary" of C is where the ant passes from one color to the other.*
3. *S is oriented, and the orientation of S induces an orientation of C (and vice versa) as follows: a person walks along the boundary curve C with their head pointing in the direction N. The surface S is on their left.*
4. *$F : \mathbb{R}^3 \to \mathbb{R}^3$ is C^1.*

Then, we have that

$$\oint_C F = \int_S \nabla \times F.$$

Note that Stokes' theorem generalizes Green's theorem. If S is in the xy-plane, for instance, we can use $N = (0, 0, 1)$ such that

$$(\nabla \times F) \cdot N = \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}$$

as in Green.

35.2 Function Spaces

Remember that, if V is a real vector space, then an inner product $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ is a function such that

1. $\langle u,u\rangle \geq 0, \forall u \in V$, with equality if and only if $u = 0$.
2. $\langle u,v\rangle = \langle v,u\rangle$.
3. $\langle \lambda u,v\rangle = \lambda\langle u,v\rangle$ if $\lambda \in \mathbb{R}$.
4. $\langle u+w,v\rangle = \langle u,v\rangle + \langle w,v\rangle$.

With these properties, the inner product is a symmetric, positive, definite, bilinear form.

Recall that a complete inner product space is called a Hilbert space. In Hilbert spaces, Cauchy sequences converge with respect to the norm induced by the inner product $\|v\| = \sqrt{\langle v,v\rangle}$.

Remember one such inner product for functions that we defined in 25a. If V is the space of real-valued continuous functions on $[a,b]$, then V is an inner product space with

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

Note that $\left\|f\right\|$ induced by this inner product is not the same sup norm that we have traditionally used in this course on $C([a,b],\mathbb{R})$.

However, this inner product space that we mentioned earlier is not complete. Consider the sequence of functions $\left\{f_n\right\}$ where $f_n : [a,b] \to \mathbb{R}$ is defined as

$$f_n(x) = \begin{cases} 1 & \text{if } a \leq x < \frac{b+a}{2} \\ -\frac{2n}{b-a}x + \frac{n(b+a)}{b-a} + 1 & \text{if } \frac{b+a}{2} \leq x < x_n \\ 0 & \text{if } x_n \leq x \leq b. \end{cases}$$

Then, for each $n \in \mathbb{N}$, f_n is continuous on $[a,b]$, and $\left\{f_n\right\}_{n=1}^{\infty}$ is Cauchy. But there is no continuous f such that $\left\|f_n - f\right\| \to 0$ as $n \to \infty$, thus demonstrating that the inner product space described earlier is not complete.

If we use $\left\|f\right\|_\infty = \sup_{x\in[a,b]}|f(x)|$, then the space is complete. But the norm does not arise from an inner product, a fact we can check using the real polarization identity.

We'll want the structure of this inner product space in support of the following goal: we wish to represent any $f \in V$ as $\sum_{k=0}^{\infty} c_k\varphi_k$ where $\varphi_0, \varphi_1, \varphi_2, \ldots$ forms an orthogonal family.

This, however, raises some natural questions:

1. What space V of functions should we use?
2. What inner product should we use?
3. What do we mean by "represent?"
4. How do we approximate functions $f \in V$ using finite linear combinations?

We'll begin with this not-so-much-a-definition definition:

Definition 35.2. *$L^2[a,b]$ is the space of "functions" $f : [a,b] \to \mathbb{R}$ such that f^2 is "integrable."*

The fine print is that

1. We actually mean the Lebesgue integral, not the Riemann integral that we're used to.
2. Technically, elements of $L^2[a,b]$ are equivalence classes of functions. Two functions of the same class differ on a set of measure zero.

Take Math 114 if this interests you!

In what follows, it's okay if we consider elements of $L^2[a,b]$ to be piecewise continuous functions that are square integrable.

We claim that $L^2[a,b]$ is a real inner product space with the inner product

$$\langle f, g\rangle = \int_a^b fg.$$

This claim would be easy if the space consisted only of continuous functions. But elements of $L^2[a,b]$ need not be bounded. Consider, for instance,

$$f(x) = \begin{cases} 0 & \text{if } x = 0 \\ x^{-1/4} & \text{otherwise.} \end{cases}$$

Then, $\int_0^1 f^2$ exists, so $f \in L^2[0,1]$.

36 April 25

36.1 Orthonormal Sequences and Fourier Series

Last time we defined $L^2[a,b]$, which uses the inner product

$$\langle f, g\rangle = \int_a^b fg$$

which induces a norm

$$\|f\| = \left(\int_a^b f^2\right)^{1/2}.$$

Recall from 25a the following inequality:

Theorem 36.1. *(Cauchy-Schwarz). Let V be a real inner product space and $\|\cdot\|$ the norm induced by the inner product. Then,*

$$\left|\langle f, g\rangle\right| \le \|f\|\|g\|, \forall f, g \in V$$

Proof. It is clear that the above holds if $g = 0$, so assume $g \neq 0$. Then, for all $c \in \mathbb{R}$, we see that $0 \le \|f - cg\|^2$. In particular, it is true if

$$c = \frac{\langle f, g\rangle}{\langle g, g\rangle}$$

from which Cauchy-Schwarz easily follows. □

As a consequence, note that if $f, g \in L^2[a,b]$, then, by definition, $\int_a^b f^2$ and $\int_a^b g^2$ are finite, and Cauchy-Schwarz implies that

$$\left|\int_a^b f(x)g(x)\,dx\right| \le \left(\int_a^b f^2\right)^{1/2}\left(\int_a^b g^2\right)^{1/2}.$$

Note also that, per the triangle inequality, we get Minkowski's inequality, which states that

$$\left(\int_a^b (f+g)^2\right)^{1/2} \le \left(\int_a^b f^2\right)^{1/2} + \left(\int_a^b g^2\right)^{1/2}.$$

Definition 36.1. *If V is a real inner product space, we say that f and g are orthonormal if $\langle f, g\rangle = 0$.*

Definition 36.2. *An orthonormal sequence $\varphi_0, \varphi_1, \varphi_2, \dots$ in an inner product spaces V is a sequence of elements satisfying*

$$\langle \varphi_i, \varphi_j\rangle = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, (i, j \ge 0)$$

Example 36.1. *The following is an orthonormal sequence in $L^2[-\pi, \pi]$:*

$$\frac{1}{\sqrt{2\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin(2x)}{\sqrt{\pi}}, \frac{\cos(2x)}{\sqrt{\pi}}, \cdots$$

Mathematica is my best friend, but we can also use the following properties to help us evaluate complicated integrals for inner products:

1. $\cos\left(\alpha+\beta\right) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$
2. $\cos\left(\alpha-\beta\right) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$
3. $e^{i\theta} = \cos\theta + i\sin\theta$

Q: In the real finite dimensional vector space $\mathbb{R}^n$, we can represent any $v \in \mathbb{R}^n$ as a linear combination $v = \sum_{j=1}^n c_j u_j$ if $u_1, u_2, \cdots, u_n$ are an orthonormal basis.

But can we generalize this to infinite dimensional function spaces?

Recall that, in a finite dimensional vector space,

$$\|v\|^2 = \sum_{i=1}^n \langle v, u_j\rangle^2.$$

Definition 36.3. *If $\varphi_0, \varphi_1, \varphi_2, \ldots$ is an orthonormal sequence in an inner product space V and $f \in V$, we can write $\sum_{k=0}^\infty c_k\varphi_k$ to mean*

$$\left\| f - \sum_{k=0}^n c_k\varphi_k \right\| \to 0 \text{ as } n \to \infty$$

where $\|\cdot\|$ is norm induced by the inner product.

Proposition 36.1. *Suppose $f \in V$ and $\varphi_0, \varphi_1, \varphi_2, \ldots$ is an orthonormal sequence in a real inner product space V. If $f = \sum_{k=0}^\infty c_k\varphi_k$ then $c_k = \langle f, \varphi_k\rangle$.*

Proof. Let $\epsilon > 0$ be given. Choose $N \in \mathbb{N}$ such that

$$\left\| f - \sum_{k=0}^n c_k\varphi_k \right\| < \epsilon, \forall n \geq N.$$

Freeze some $k \in \{0\} \cup \mathbb{N}$. We want to find c_k. Choose $n \geq \max\{k, N\}$.

By linearity, we know that

$$\langle f, \varphi_k\rangle = \langle f - S_n, \varphi_k\rangle + \langle S_n, \varphi_k\rangle = \langle f - S_n, \varphi_k\rangle + c_k.$$

Then, by Cauchy-Schwarz, we know that

$$\left|\langle f - S_n, \varphi_k\rangle\right| \leq \left\|f - S_n\right\|\left\|\varphi_k\right\| < \epsilon.$$

But ϵ was arbitrary, so $\langle f, \varphi_k\rangle = c_k$. □

So, if f has such a series representation, then each coefficient c_k can be found as described above.

Definition 36.4. *Let $\varphi_0, \varphi_1, \varphi_2, \ldots$ be an orthonormal sequence in a real inner product space V. The sequence is complete if every $f \in V$ can be written as a sum*

$$f = \sum_{k=0}^\infty c_k\varphi_k.$$

We call this the Fourier series for f.

Note the following property:

Example 36.2. *We have previously stated that*

$$\frac{1}{\sqrt{2\pi}}, \frac{\sin x}{\sqrt{pi}}, \frac{\cos x}{\sqrt{\pi}}, \dots$$

is an orthonormal sequence in $L^2[-\pi,\pi]$. *The functions* $f(x) = -2 + 7\cos(2x)$ *and the function* g *defined as*

$$g(x) = \begin{cases} f(x) & \text{if } x \in (\mathbb{R} \setminus \mathbb{Q}) \cap [-\pi,\pi] \\ 0 & \text{if } x \in \mathbb{Q} \cap [-\pi,\pi]. \end{cases}$$

Then, f *and* g *have the same Fourier series, since they differ on a set of measure* 0. *In this case, we calculate that* $c_0 = -2\sqrt{2\pi}, c_4 = 7\sqrt{\pi}$, *and all other coefficients are* 0.

Proposition 36.2. *(Bessel). If* $\varphi_0, \varphi_1, \dots, \varphi_n, \dots$ *is an orthonormal sequence in a real inner product space* V, *then, for each* $f \in V$, *the series*

$$\sum_{k=0}^{\infty} \left|\langle f, \varphi_k \rangle\right|^2$$

converges and is bounded above by $\left\|f\right\|^2$.

Proposition 36.3. *(Parseval). Let* $\varphi_0, \varphi_1, \dots, \varphi_n, \dots$ *be an orthonormal sequence in a real inner product space* V. *The orthonormal sequence is complete if and only if,* $\forall f \in V$, *we have equality in Bessel's inequality. That is,*

$$\sum_{k=0}^{\infty} \left|\langle f, \varphi_k \rangle\right|^2 = \left\|f\right\|^2.$$

Theorem 36.2. *Let* V *be a real inner product space, and let* $\varphi_0, \varphi_1, \dots, \varphi_n$ *be a set of orthonormal vectors. Then, for every list of real numbers* $\alpha_0, \alpha_1, \cdots, \alpha_n$, *we have*

$$\left\| f - \sum_{k=0}^{n} \alpha_k \varphi_k \right\| \geq \left\| f - \sum_{k=0}^{n} \langle f, \varphi_k \rangle \varphi_k \right\|.$$

37 April 27

Example 37.1. *Consider the function $f(x) = x - x^2$ on the interval $[-\pi, \pi]$, and consider the orthoormal set*

$$\varphi_0 = \frac{1}{\sqrt{2\pi}},\ \varphi_1 = \frac{\sin x}{\sqrt{\pi}},\ \varphi_2 = \frac{\cos x}{\sqrt{\pi}}.$$

Then, to best approximate f in the L^2 sense using these functions, we want to use

$$\langle f, \varphi_0\rangle\varphi_0 + \langle f, \varphi_1\rangle\varphi_1 + \langle f, \varphi_2\rangle\varphi_2$$

per the last theorem covered last class.

Theorem 37.1. *(Mean Completeness). In the space $L^2[-\pi, \pi]$, the orthonormal sequence*

$$\frac{1}{\sqrt{2\pi}},\ \frac{1}{\sqrt{\pi}}\sin x,\ \frac{1}{\sqrt{\pi}}\cos x,\ \frac{1}{\sqrt{\pi}}\sin(2x),\ \frac{1}{\sqrt{\pi}}\cos(2x),\dots$$

is complete. That is, every $f \in L^2[-\pi, \pi]$ has a unique representation of the for $\sum_{k=0}^{\infty} c_k\varphi_k$ where $c_k = \langle f, \varphi_k\rangle$.

37.1 Convergence of Fourier Series

This leaves us with a question: under what circumstances does the Fourier series for $f \in L^2[a, b]$ converge to f uniformly?

Note that we would need f to be continuous at the very least because the uniform limit of continuous functions is continuous. Note also that, at π and $-\pi$, the functions have to be equal, since each function in the orthonormal list is the same at these two points.

Theorem 37.2. *Let $f \in C^1([-\pi, \pi], \mathbb{R})$ and $f(-\pi) = f(\pi)$. If $\varphi_0, \varphi_1, \varphi_2, \dots$ are the complete orthonormal sequence from the Mean Completeness Theorem, then the Fourier series $\sum_{k=0}^{\infty}\langle f, \varphi_k\rangle\varphi_k$ converges uniformly to f on $[-\pi, \pi]$.*

Proof. Let $f \in C^1([-\pi, \pi], \mathbb{R})$. Then f and f' are both continuous on $[-\pi, \pi]$, so f^2 and $(f')^2$ are also both continuous on the interval. Thus, f^2 and $(f')^2$ are integrable on $[-\pi, \pi]$, so $f, f' \in L^2[-\pi, \pi]$. By the Mean Completeness theorem, we can write

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k\cos(kx) + b_k\sin(kx))$$

and similarly

$$f'(x) = \frac{a_0'}{2} + \sum_{k=1}^{\infty}\left(a_k'\cos(kx) + b_k'\sin(kx)\right).$$

First, we will relate the a_k' and b_k' to a_k and b_k. Note that, if $k \geq 1$, then $a_k' = \frac{\sqrt{\pi}a_k'}{\sqrt{\pi}}$. So,

$$\sqrt{\pi}a_k' = \int_{-\pi}^{\pi} f'(x)\frac{\cos(kx)}{\sqrt{\pi}}\,dx$$

which then leads to

$$a_k' = \frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)\cos(kx)\,dx = \frac{1}{\pi}\left[f(x)\cos(kx)|_{-\pi}^{\pi} + k\int_{-\pi}^{\pi} f(x)\sin(kx)\,dx\right]$$

but since $f(-\pi) = f(\pi)$ and $\cos(kx) = \cos(-kx)$, we get that $a_k' = kb_k, k \geq 1$, and we can similarly get that $b_k' = -ka_k, k \geq 1$. Additionally, we find that

$$a_0' = \frac{1}{\pi} \int_{-\pi}^{\pi} f'(x)\, dx = 0$$

since $f(\pi) = f(-\pi)$.

Now, we will use the Weierstrass-M test to show that the Fourier series for $f(x)$ converges uniformly. We will bound the terms. For each $k \geq 1$, we know that

$$\left|a_k \cos(kx) + b_k \sin(kx)\right| \leq \left|a_k\right| + \left|b_k\right| = M_k$$

and we now wish to show that $\sum_{k=1}^{\infty} M_k$ converges. Look at the partial sums of the form

$$S_n = \sum_{k=1}^{n} \left|a_k\right| = \sum_{k=1}^{n} \frac{1}{k} \left|b_k'\right| \leq \left(\sum_{k=1}^{n} \frac{1}{k^2}\right)^{1/2} \left(\sum_{k=1}^{n} \left|b_k'\right|^2\right)^{1/2}$$

where the left quantity is bounded above by the p-series rule and the right side is bounded above by applying Bessel's inequality to f'. The same bounds work for each n. We have similar considerations also for $\sum_{k=1}^{n} \left|b_k\right|$.

Then, since the sequence of partial sums $\sum_{k=1}^{n} M_k$ is monotone increasing and bounded above, $\sum_{k=1}^{\infty} M_k$ converges by the monotone convergence theorem. By the Weierstrass-M test, we're done. □

Note that:

1. If $f(-\pi) \neq f(\pi)$, then for each $\epsilon \in (0, \pi)$, convergence is still uniform on the interval $[-\pi + \epsilon, \pi - \epsilon]$.
2. Proofs of convergence for Fourier series are way more work than proofs for Taylor series. In the latter, the ratio test is usually useless.

We'll wrap up with a brief discussion on when we can get pointwise convergence, rather than the uniform convergence we just discussed.

Consider our usual complete orthonormal sequence on $L^2[-\pi, \pi]$. If $f \in L^2[-\pi, \pi]$, we have

$$\left\| f - \sum_{k=0}^{n} \langle f, \varphi_k \rangle \varphi_k \right\|_{L^2} \to 0 \text{ as } n \to \infty.$$

We also gave sufficient conditions for

$$\left\| f - \sum_{k=0}^{n} \langle f, \varphi_k \rangle \varphi_k \right\|_{\infty} \to 0 \text{ as } n \to \infty$$

which is equivalent to uniform convergence.

Recall that $f : [a, b] \to \mathbb{R}$ is piecewise continuous if there exists a finite partition ($a = x_0 < x_1 < x_2 < \cdots < x_n = b$) such that f is bounded and continuous on each (x_{k-1}, x_k), $k = 1, 2, \ldots, n$. Then, we write

$$f(x_0^+) = \lim_{x \to x_0^+} f(x) \quad \text{and} \quad f(x_0^-) = \lim_{x \to x_0^-} f(x)$$

if the limits exist. We have said that, if $f(x_0^+)$ and $f(x_0^-)$ exist and are unequal, there is a jump discontinuity at x_0. If both exist and are not equal to $f(x)$, then there is a jump discontinuity at x_0. Finally, if they both exist and are both equal to $f(x)$, then f is continuous at x_0.

Finally, let

$$f'(x_0^+) = \lim_{h \to 0^+} \frac{f(x_0 + h) - f(x_0^+)}{h} \quad \text{and} \quad f'(x_0^-) = \lim_{h \to 0^+} \frac{f(x_0^+) - f(x_0 - h)}{h}.$$

Theorem 37.3. *(Pointwise convergence). Let $\varphi_0, \varphi_1, \varphi_2, \ldots$ be our complete orthonormal sequence from the mean completeness theorem. Let $f \in L^2[-\pi, \pi]$ be piecewise C^1 and assume that any discontinuities are jump or removable. Then, the Fourier series for f converges pointwise to*

$$\frac{f(x^+) + f(x^-)}{2}$$

which is just $f(x)$ is f is continuous at x.

Note that we can also do Fourier series on other intervals. For the interval $[0, L]$, for instance, we can use

$$\frac{1}{L}, \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right), \sqrt{\frac{2}{L}} \cos\left(\frac{n\pi x}{L}\right), n \in \mathbb{N}.$$

Conclusion

That's it for Math 25b – and for Math 25! Thank you to Wes and to everyone else for making this a special math class!