Math Journal

Isabelle Mills

February 2, 2026

8/31/2024

My goal for today is to work through the appendix to chapter 1 in Baby Rudin. This appendix focuses on constructing the real numbers using Dedikind cuts.

We define a cut to be a set $\alpha \subset \mathbb{Q}$ such that:

1. $\alpha \neq \emptyset$
2. If $p \in \alpha, \quad q \in \mathbb{Q}$, and $q < p$, then $q \in \alpha$.
3. If $p \in \alpha$, then $p < r$ for some $r \in \alpha$

Point 3 tells us that α doesn't have a max element. Also, point 2 directly implies the following facts:

a. If $p \in \alpha, \quad q \in \mathbb{Q}$, and $q \notin \alpha$, then $q > p$.
b. If $r \notin \alpha, \quad r, s \in \mathbb{Q}$, and $r < s$, then $s \notin \alpha$.

As a shorthand, I shall refer to the set of all cuts as R.

An example of a cut would be the set of rational numbers less than 2.

Firstly, we shall assign an ordering to R. Specifically, given any $\alpha, \beta \in R$, we say that $\alpha < \beta$ if $\alpha \subset \beta$ (a proper subset).

Here we prove that $<$ satisfies the definition of an ordering.

I. It's obvious from the definition of a proper subset that at most one of the following three things can be true: $\alpha < \beta, \quad \alpha = \beta$, and $\beta < \alpha$.

Now let's assume that $a \not< \beta$ and $\alpha \neq \beta$. Then $\exists p \in \alpha$ such that $p \notin \beta$. But then for any $q \in \beta$, we must have by fact b. above that $q < p$. Hence $q \in \alpha$, meaning that $\beta \subset \alpha$. This proves that at least one of the following has to be true: $\alpha < \beta, \quad \alpha = \beta$, and $\beta < \alpha$.

II. If for $\alpha, \beta, \gamma \in R$ we have that $\alpha < \beta$ and $\beta < \gamma$, then clearly $\alpha < \gamma$ becuase $\alpha \subset \beta \subset \gamma$.

Now we claim that R equipped with $<$ has the least-upper-bound property.

Proof:

Let $A \subset R$ be nonempty and $\beta \in R$ be an upper bound of A. Then set $\gamma = \bigcup\limits_{\alpha \in A} \alpha$. Firstly, we want to show that $\gamma \in R$

Since $A \neq \emptyset$, there exists $\alpha_0 \in A$. And as $\alpha_0 \neq \emptyset$ and $\alpha_0 \subseteq \gamma$ by definition, we know that $\gamma \neq \emptyset$. At the same time, we know that $\gamma \subset \beta$ since $\forall \alpha \in A$, $\alpha \subset \beta$. Hence, $\gamma \neq \mathbb{Q}$, meaning that γ satisfies property 1. of cuts.

Next, let $p \in \gamma$ and $q \in \mathbb{Q}$ such that $q < p$. We know that for some $\alpha_1 \in A$, we have that $p \in \alpha_1$. Hence by property 2. of cuts, we know that $q \in \alpha_1 \subset \gamma$, thus showing that γ satisfies property 2. of cuts.

Thirdly, by property 3. we can pick $r \in \alpha_1$ such that $p < r$ and $r \in \alpha_1 \subset \gamma$. So, γ satisfies property 3. of cuts.

With that, we've now shown that $\gamma \in R$. Clearly, γ is an upper bound of A since $\alpha \subset \gamma$ for all $\alpha \in A$. Meanwhile, consider any $\delta < \gamma$. Then $\exists s \in \gamma$ such that $s \notin \delta$. And since $s \in \gamma$, we know that $s \in \alpha$ for some $\alpha \in A$. Hence, $\delta < \alpha$, meaning that δ is not an upper bound of A. This shows that $\gamma = \sup A$.

Secondly, we want to assign $+$ and $\cdot$ operations to R so that R is an ordered field.

To start, given any $\alpha, \beta \in R$, we shall define $\alpha + \beta$ to be the set of all sums $r + s$ such that $r \in \alpha$ and $s \in \beta$.

Here we show that $\alpha + \beta \in R$.

1. Clearly, $\alpha + \beta \neq \emptyset$. Also, take $r' \notin \alpha$ and $s' \notin \beta$. Then $r' + s' > r + s$ for all $r \in \alpha$ and $s \in \beta$. Hence, $r' + s' \notin \alpha + \beta$, meaning that $\alpha + \beta \neq \mathbb{Q}$.

Now let $p \in \alpha + \beta$. Thus there exists $r \in \alpha$ and $s \in \beta$ such that $p = r + s$.

2. Suppose $q < p$. Then $q - s < r$, meaning that $q - s \in \alpha$. Hence, $q = (q - s) + s \in \alpha + \beta$.

3. Let $t \in \alpha$ so that $t > r$. Then $p = r + s < t + s$ and $t + s \in \alpha + \beta$.

Also, we shall define 0^* to be the set of all negative rational numbers. Clearly, 0^* is a cut. Furthermore, we claim that $+$ satisfies the addition requirements of a field with 0^* as its 0 element.

Commutativity and associativity of $+$ on R follows directly from the commutativity and associativity of addition on the rational numbers.

Also, for any $\alpha \in R$, $\quad \alpha + 0^* = \alpha$.

If $r \in \alpha$ and $s \in 0^*$, then $r + s < r$. Hence $r + s \in \alpha$, meaning that $\alpha + 0^* \subseteq \alpha$. Meanwhile, if $p \in \alpha$, then we can pick $r \in \alpha$ such that $r > p$. Then, $p - r \in 0^*$ and $p = r + (p - r) \in \alpha + 0^*$. So, $\alpha \subseteq \alpha + 0^*$.

Finally, given any $\alpha \in R$, let $\beta = \{p \in \mathbb{Q} \mid \exists r \in \mathbb{Q}^+ \ s.t. \ -p - r \notin \alpha\}$.

To give some intuition on this definition, firstly we want to guarentee that for all $p \in \beta$, $-p$ is greater than all elements of α. Secondly, we add the $-r$ term to guarentee that β doesn't have a maximum element.

We claim that $\beta \in R$ and $\beta + \alpha = 0^*$. Hence, we can define $-\alpha = \beta$.

To start, we'll show that $\beta \in R$:

1. For $s \notin \alpha$ and $p = -s - 1$, we have that $-p - 1 \notin \alpha$. Hence, $p \in \beta$, meaning that $\beta \neq \emptyset$. Meanwhile, if $q \in \alpha$, then $-q \notin \beta$ because there does not exist $r > 0$ such that $-(-q) - r = q - r \notin \alpha$. So $\beta \neq \mathbb{Q}$.

Now let $p \in \beta$ and pick $r > 0$ such that $-p - r \notin \alpha$.

2. Suppose $q < p$. Then $-q - r > -p - r$, meaning that $-q - r \notin \alpha$. Hence, $q \in \beta$.

3. Let $t = p + \frac{r}{2}$. Then $t > p$ and $-t - \frac{r}{2} = -p - r \notin \alpha$, meaning $t \in \beta$.

Now that we've proved $\beta \in R$, we next prove that β is the additive inverse of α. To start, suppose $r \in \alpha$ and $s \in \beta$. Then $-s \notin \alpha$, meaning that $r < -s$. So $r + s < 0$, thus showing that $\alpha + \beta \subseteq 0^*$.

As for the other inclusion, pick any $v \in 0^*$ and set $w = -\frac{v}{2}$. Then because $w > 0$, we can use the archimedean property of $\mathbb{Q}$ to say that there exists $n \in \mathbb{Z}$ such that $nw \in \alpha$ but $(n+1)w \notin \alpha$. Put $p = -(n+2)w$. Then $p \in \beta$ because $-p-w = (n+1)w \notin \alpha$. And finally, $v = nw + p \in \alpha + \beta$. Thus, $0^* \subseteq \alpha + \beta$.

9/1/2024

Based on the definition of $+$, it's also hopefully clear that for any $\alpha, \beta, \gamma \in R$ such that $\alpha < \beta$, we have that $\alpha + \gamma < \beta + \gamma$.

Next, we shall define multiplication on R. Except, first we're going to limit ourselves to the set R^+ of all cuts greater than 0^*. So, given any $\alpha, \beta \in R^+$, we shall define $\alpha\beta$ to be the set of all $p \in \mathbb{Q}$ such that $p \leq rs$ where $r \in \alpha$, $s \in \beta$, $r > 0$, and $s > 0$.

Here we show that $\alpha\beta \in R^+$.

1. Clearly $\alpha\beta \neq \emptyset$. Also, take any $r' \notin \alpha$ and $s' \notin \beta$. Then $r's' > rs$ for all $r \in \alpha \cap \mathbb{Q}^+$ and $s \in \beta \cap \mathbb{Q}^+$ since all four rational numbers are positive. By extension, $r's'$ is greater than all the elements (both positive and negative) of $\alpha\beta$. So, $r's' \notin \alpha\beta$, meaning that $\alpha\beta \neq \mathbb{Q}$.

Now let $p \in \alpha\beta$. Based on our definition of $\alpha\beta$, we know that the conditions of a cut trivially hold for any negative p. So, we'll assume from now on that $p > 0$. (Also note that a positive choice of p must exist because both α and β by assumption have positive elements.)

Since $p \in \alpha\beta \cap \mathbb{Q}^+$, we know there exists $r \in \alpha$ and $s \in \beta$ such that $p = rs$ and $r, s > 0$.

2. Suppose $0 < q < p$ (the case where $q \leq 0$ is trivial). Then $\frac{q}{s} < r$, meaning that $\frac{q}{s} \in \alpha$. So, $q = \frac{q}{s} \cdot s \in \alpha\beta$.

3. Let $t \in \alpha$ so that $t > r$. Then $p = rs < ts$ and $ts \in \alpha\beta$.

Also, we shall define 1^* to be the set of all rational numbers less than 1. Clearly, 1^* is a cut. And we claim that $\cdot$ satisfies the multiplication requirements of a field with 1^* as its 1 element.

As before, commutativity and associativity of $\cdot$ on R^+ follows directly from commutativity and associativity of multiplication on the rational numbers.

Next, for any $\alpha \in R^+$, we have that $\alpha 1^* = \alpha$.

It's clear that for any rational number $r \leq 0$, we have that $r \in \alpha 1^*$ and $r \in \alpha$. So we can exclusively focus on positive rational numbers.

Now suppose $r \in \alpha \cap \mathbb{Q}^+$ and $s \in 1^*$. Then $rs < r$, meaning that $rs \in \alpha$. So $\alpha 1^* \subseteq \alpha$. Meanwhile, if $p \in \alpha \cap \mathbb{Q}^+$, then we can pick $r \in \alpha$ such that $r > p$. Then $\frac{p}{r} \in 1^*$ and $p = \frac{p}{r} \cdot r \in \alpha 1^*$. So, $\alpha \subseteq \alpha 1^*$.

Thirdly, given any $\alpha \in R^+$, define:

$$\beta = \{p \in \mathbb{Q} \mid p \leq 0\} \cup \{p \in \mathbb{Q}^+ \mid \exists r \in \mathbb{Q}^+ \; s.t. \; \tfrac{1}{q} - r \notin \alpha\}$$

Here we show that $\beta \in R^+$.

1. Clearly $\beta \neq \emptyset$. Also, if $q \in \alpha$, then $\frac{1}{q} \notin \beta$. Hence, $\beta \neq \mathbb{Q}$.

Now let $p \in \beta$ and pick $r > 0$ such that $\frac{1}{p} - r \notin \alpha$. Also, assume $p > 0$ because the proof is trivial if $p \leq 0$. (The fact that $p > 0$ in β exists is trivial to show.)

2. If $q \leq 0 < p$, then trivially $q \in \beta$. Meanwhile, if $0 < q < p$, then $\frac{1}{q} - r > \frac{1}{p} - r$, meaning that $\frac{1}{q} - r \notin \alpha$. Hence, $q \notin \beta$.

3. Let $t = \frac{1}{\frac{1}{p} - \frac{r}{2}}$. Then since $\frac{1}{p} - r \notin \alpha$, we know that $\frac{1}{p} - \frac{r}{2} > 0$. Also since $\frac{1}{t} = \frac{1}{p} - \frac{r}{2} < \frac{1}{p}$, we have that $t > p$. But note that $\frac{1}{t} - \frac{r}{2} = \frac{1}{p} - r \notin \alpha$. Hence $t \notin \beta$.

We claim that $\beta\alpha = 1^*$. Hence, we can define $\frac{1}{\alpha} = \beta$.

To start, suppose $r \in \alpha \cap \mathbb{Q}^+$ and $s \in \beta \cap \mathbb{Q}^+$. Then $\frac{1}{s} \notin \alpha$, meaning that $r < \frac{1}{s}$. So $rs < 1$, thus showing that $\alpha\beta \subseteq 1^*$.

The other inclusion has a more complicated proof. Firstly, take any $v \in 1^* \cap \mathbb{Q}^+$ (the proof is trivial if $v \leq 0$). Then set $w = \frac{1}{v}$, meaning that $w > 1$. Now since $\alpha \in R^+$, we know there exists $n \in \mathbb{Z}$ such that $w^n \in \alpha$ but $w^{n+1} \notin \alpha$. Then as $w^{n+2} > w^{n+1}$, we know that $\frac{1}{w^{n+2}} \in \beta$. Hence, $v^2 = w^n \frac{1}{w^{n+2}} \in \alpha\beta$.

Now that we've shown that the square of every $v \in 1^* \cap \mathbb{Q}^+$ is also in $\alpha\beta$, we next show that there exists $z \in 1^* \cap \mathbb{Q}^+$ such that $z^2 > v$. Suppose $v = \frac{p}{q}$ where $p, q \in \mathbb{Z}^+$. Then set $z = \frac{p+q}{2q}$. Importantly, since $p < q$, we still have that $z \in 1^*$. But also note that:

$$z^2 - v = \frac{p^2+2pq+q^2}{4q^2} - \frac{4pq}{4q^2} = \frac{p^2-2pq+q^2}{4q^2} = \left(\frac{p-q}{2q}\right)^2 \geq 0$$

Thus as $v \leq z^2$ and $z^2 \in \alpha\beta$, we have that $v \in \alpha\beta$ as well. So $1^* \subseteq \alpha\beta$.

Finally, so long as $\alpha, \beta, \gamma \in R^+$, we have that $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$ because the rational numbers satisfy the distributive property.

Notably, in proving that $\alpha\beta \in R^+$ before, we also guarenteed that for $\alpha, \beta > 0$, we have that $\alpha\beta > 0$.

9/7/2024

Now we still need to extend our definition of multiplication from R^+ to all of R. To do this, set $\alpha 0^* = 0^* \alpha = 0^*$ and define:

$$\alpha\beta = \begin{cases} (-\alpha)(-\beta) & \text{if } \alpha < 0^*, \beta < 0^* \\ -((-\alpha)\beta) & \text{if } \alpha < 0^*, \beta > 0^* \\ -(\alpha(-\beta)) & \text{if } \alpha > 0^*, \beta < 0^* \end{cases}$$

Having done that, reproving those properties of multiplication on all of R just becomes a matter of addressing many cases and using the identity that $(-(-\alpha)) = \alpha$.

Note that that identity can be proven just from the addition properties of a field.

Because I'm bored with this construction at this point, I'm going to skip reproving those properties.

So now that we've established that R is a field, all we have left to do is to show that all numbers $r, s \in \mathbb{Q}$ are represented by cuts $r^*, s^* \in R$ such that:

- $(r + s)^* = r^* + s^*$
- $(rs)^* = r^* s^*$
- $r < s \Longleftrightarrow r^* < s^*$

Again, I'm super bored and demotivated at this point. So, I'm going to skip showing this.

With all that done, we've now shown that R satisfies all of the properties of real numbers. That concludes the proof of the existence theorem of the real numbers.

9/9/2024

Today I'm just looking at James Munkres' book *Topology*. Now while I'm done with the era of my life of taking exhaustive notes on a textbook, I still want to write down some interesting proofs. I also hope to do some exercises.

Theorem 7.8: Let A be a nonempty set. There is no injective map $f : \mathcal{P}(A) \longrightarrow A$ and there is no surjective map $g : A \longrightarrow \mathcal{P}(A)$.

In other words, the power set of a set has strictly greater cardinality.

Proof:
If such an injective f existed, then that would imply a surjective g exists. So, we just need to show that any function $g : A \longrightarrow \mathcal{P}(A)$ isn't surjective.

Let $g : A \longrightarrow \mathcal{P}(A)$ be any function and define $B = \{a \in A \mid a \in A - g(a)\}$. Clearly, $B \subseteq A$. However, B cannot be in the image of g. After all, suppose there exists $a_0 \in A$ such that $g(a_0) = B$. Then we get a contradiction because:

$$a_0 \in B \Longleftrightarrow a_0 \in A - g(a_0) \Longleftrightarrow a_0 \in A - B$$

Hence, $g(A) \neq \mathcal{P}(A)$ and we conclude that g can't be surjective. ■

Exercise 7.3: Let $X = \{0, 1\}$. Show there is a bijective correspondence between the set $\mathcal{P}(\mathbb{Z}_+)$ and the Cartesian product X^ω.

For any set $A \in \mathcal{P}(\mathbb{Z}_+)$, define $f(A)$ to be the ω-tuple $\mathbf{x}$ such that for all $i \in \mathbb{Z}^+$, $\mathbf{x}_i = 1$ if $i \in A$ and $\mathbf{x}_i = 0$ if $i \notin A$. Then clearly f is injective as $\forall A, B \in \mathcal{P}(\mathbb{Z}_+)$, $f(A) = f(B) \Longrightarrow A = B$. Also, given any $\mathbf{x} \in X^\omega$, we know that the set $A = \{i \in \mathbb{Z}_+ \mid \mathbf{x}_i = 1\}$ satisfies that $f(A) = \mathbf{x}$, meaning f is surjective.

Hence, f is a bijective function between $\mathcal{P}(\mathbb{Z}_+)$ and X^ω.

Note that this construction still works if $\mathbb{Z}_+$ is replaced with any countably infinite set.

Exercise 7.5: Determine whether the following sets are countable or not.

(f) The set F of all functions $f : \mathbb{Z}_+ \longrightarrow \{0, 1\}$ that are "eventually zero", meaning there is a positive integer N such that $f(n) = 0$ for all $n \geq N$.

F is countable. To see why, let:

$$A_n = \{f : \mathbb{Z}_+ \longrightarrow \{0, 1\} \mid \forall i \geq n, \;\; f(i) = 0\}$$

Thus each A_n is finite (with 2^n elements) and $F = \bigcup\limits_{n=1}^{\infty} A_n$.

(g) The set G of all functions $f : \mathbb{Z}_+ \longrightarrow \mathbb{Z}_+$ that are eventually 1.

G is countable. To see why, let:

$$A_n = \{f : \mathbb{Z}_+ \longrightarrow \mathbb{Z}_+ \mid \forall i \geq n, \ \ f(i) = 1\}$$

Then each A_n has a bijective correspondence with $(\mathbb{Z}_+)^n$, meaning each A_n is countable, and $G = \bigcup_{n=1}^{\infty} A_n$.

The same argument applies to all functions $f : \mathbb{Z}_+ \longrightarrow \mathbb{Z}_+$ that are eventually any constant.

(h) The set H of all functions $f : \mathbb{Z}_+ \longrightarrow \mathbb{Z}_+$ that are eventually constant.

H is countable. To see why, let A_n be the set of all functions $f : \mathbb{Z}_+ \longrightarrow \mathbb{Z}_+$ that are eventually n. Because of part g of this exercise, we know that each A_n is countable. Also, $H = \bigcup_{n=1}^{\infty} A_n$.

(i) The set I of all two-element subsets of $\mathbb{Z}_+$

(j) The set J of all finite subsets of $\mathbb{Z}_+$.

Both I and J are countably infinite. We know this because we can define surjections from $(\mathbb{Z}_+)^2$ to I and $\bigcup_{n=1}^{\infty} (\mathbb{Z}_+)^n$ to J.

(Finite cartesian products of countable sets and unions of countably many countable sets are countable.)

Exercise 7.6.a: Show that if $B \subset A$ and there is an injection $f : A \longrightarrow B$, then $|A| = |B|$.

According to the hint, we set $A_1 = A$ and $A_n = f(A_{n-1})$ for all $n > 1$. Similarly, we set $B_1 = B$ and $B_n = f(B_{n-1})$ for all $n > 1$.

We can assume A_2 is a proper subset of B_1 because if $A_2 = B_1$, then we already have that f is a bijection. Also, as f is an injection, we know that $B_2 \subset A_2$. Thus by induction, we can conclude that:

$$A_1 \supset B_1 \supset A_2 \supset B_2 \supset A_3 \supset B_3 \supset \cdots$$

Now, the textbook recommends defining $h : A \longrightarrow B$ by:

$$h(x) = \begin{cases} f(x) & \text{if } x \in A_n - B_n \text{ for any } n \in \mathbb{Z}_+ \\ x & \text{otherwise} \end{cases}$$

I want to ask a professor about this definition because it urks me. My issue with this definition of h is that I feel like it should be possible for:

$$\bigcap_{n=1}^{\infty}(A_n \cap B_n) \neq \emptyset.$$

However, we wouldn't be able to know that some x is in that intersection and thus falls into case 2 until after an infinite number of steps.

On the other hand, $S_1 = \bigcup_{n=1}^{\infty}(A_n - B_n)$ is a valid definition for a set, as is $S_2 = A - S_1$. So the definition h is valid because it's saying that $h(x) = f(x)$ if $x \in S_1$ and $h(x) = x$ if $x \in S_2$.

Maybe my issue is just that I have trouble trusting the validity of a function definition if I can't actually evaluate that function myself. Although, there are lots of functions like that that I don't have any problem with. For example, given $g(x) = 0$ if x is rational and $g(x) = 1$ if x is irrational, what is $g(\pi^2)$?

Hopefully it is clear that h is in fact a valid function from A to B. Now firstly, we shall show that h is injective.

Let $x, y \in A$ such that $x \neq y$. If there are integers n and m such that $x \in A_n - B_n$ and $y \in A_m - B_m$, then $h(x) \neq h(y)$ because f is injective. Meanwhile, if no such n or m exists, then $h(x) \neq h(y)$ because $x \neq y$.

This leaves the case that there exists $n \in \mathbb{Z}_+$ such that $x \in A_n - B_n$ but for all $m \in \mathbb{Z}_+,\ \ y \notin A_m - B_m$. Then, note that $f(x) \in f(A_n - B_n)$. And since f is injective, we thus have that $f(x) \in f(A_n) - f(B_n) = A_{n+1} - B_{n+1}$. Therefore, as $y \notin A_{n+1} - B_{n+1}$, we know that $h(x) \neq y = h(y)$.

Next, we show h is surjective.

Let $y \in B$.

Suppose there exists $n \in \mathbb{Z}_+$ such that $y \in A_n - B_n$. We know that $n \neq 1$ since $y \in B$. Thus, there must exist $x \in A_{n-1}$ such that $y = f(x) \in f(A_{n-1}) = A_n$. Furthermore, this x can't be in B_{n-1} because otherwise y would be in B_n which we know isn't true. So, $x \in A_{n-1} - B_{n-1}$, meaning that $h(x) = f(x) = y$.

Meanwhile, if no such n exists, then we simply have that $h(y) = y$. Hence, $h(A) = B$.

Thus, we've shown that h is a bijection, meaning that $|A| = |B|$.

Exercise 7.7: Show that $|\{0,1\}^\omega| = |(\mathbb{Z}_+)^\omega|$.

Firstly, obviously a bijection exists between $\{0,1\}^\omega$ and $\{1,2\}^\omega$. Also, $\{1,2\}^\omega \subset (\mathbb{Z}_+)^\omega$. So, if we can construct an injective function from $(\mathbb{Z}_+)^\omega$ to $\{1,2\}^\omega$, then we can apply the result of exercise 7.6.a to prove this exercise's claim.

We shall create this injection using a diagonalization argument. Let $x \in (\mathbb{Z}_+)^\omega$. Then we define $f(x) = y \in \{1,2\}^\omega$ as follows:

$y(1) = 2$ if $x(1) = 1$. Otherwise $y(1) = 1$.

$y(2) = 2$ if $x(1) = 2$. Otherwise $y(2) = 1$.
$y(3) = 2$ if $x(2) = 1$. Otherwise $y(3) = 1$.

$y(4) = 2$ if $x(1) = 3$. Otherwise $y(4) = 1$.
$y(5) = 2$ if $x(2) = 2$. Otherwise $y(5) = 1$.
$y(6) = 2$ if $x(3) = 1$. Otherwise $y(6) = 1$.

$y(7) = 2$ if $x(1) = 4$. Otherwise $y(7) = 1$.

$$\vdots$$

Clearly f is an injection since $f(x_1) = f(x_2)$ implies that x_1 and x_2 have the same integers at all indices.

Exercise 7.6.b: (Schroeder-Bernstein theorem) If there are injections $f : A \longrightarrow C$ and $g : C \longrightarrow A$, then A and C have the same cardinality.

I did my work on paper and now it's late and I don't want to write more tonight.

9/11/2024

Since today's my day off, I'm gonna work through Munkres' textbook *Topology* some more.

Theorem 8.4 (Principle of recursive definition): Let A be a set and let a_0 be an element of A. Suppose ρ is a function assigning an element of A to each function f mapping a nonempty section of the positive integers onto A. Then there exists a unique function $h : \mathbb{Z}_+ \longrightarrow A$ such that:

$$(*) \qquad \begin{aligned} h(1) &= a_0 \\ h(i) &= \rho(h|_{\{1,\ldots,i-1\}}) \quad \text{for } i > 1. \end{aligned}$$

Proof outline:

Part 1: Given any $n \in \mathbb{Z}_+$, there exists a function $f : \{1, \ldots, n\} \longrightarrow A$ that satisfies $(*)$.

This is obvious from induction.

Part 2: Suppose that $f : \{1, \ldots, n\} \longrightarrow A$ and $g : \{1, \ldots, m\} \longrightarrow A$ both satisfy $(*)$ for all i in their respective domains. Then $f(i) = g(i)$ for all i in both domains.

Proof:
Suppose not. Let i be the smallest integer for which $f(i) \neq g(i)$.

We know $i \neq 1$ because $f(1) = a_0 = g(1)$. But then note that $f|_{\{1,\ldots,i-1\}} = g|_{\{1,\ldots,i-1\}}$. Hence:

$$f(i) = \rho(f|_{\{1,\ldots,i-1\}}) = \rho(g|_{\{1,\ldots,i-1\}}) = g(i).$$

This contradicts that i is the smallest integer for which $f(i) \neq g(i)$.

Part 3: Let $f_n : \{1, \ldots, n\} \longrightarrow A$ be the unique function satisfying $(*)$ (uniqueness was proven in part 2). Then we define:

$$h = \bigcup_{i=1}^{\infty} f_n$$

Because of part 2, we can fairly easily show that for each $i \in \mathbb{Z}_+$, there is exactly one element in h with i as it's first coordinate. Hence, the set h defines a functions from $\mathbb{Z}_+$ to A.

Also, hopefully it's clear that h satisfies $(*)$.

Axiom of choice: Given a collection $\mathcal{A}$ of disjoint nonempty sets, there exists a set C consisting of exactly one element from each element of $\mathcal{A}$.

A few notes:

1. If we restrict $\mathcal{A}$ to being a finite collection, then there is nothing controversial about this axiom. It only becomes controversial when $\mathcal{A}$ is allowed to be infinite.
2. There are multiple instances in baby Rudin where we made an infinite number of arbitrary choices. Looking at a lot of those proofs closer, I think many of them could avoid using the axiom of choice by specifying that we had to pick rational numbers in a set. However, being able to pick elements without worrying about a preexisting choice function is way easier.

   My take away from this is that not only does it make proofs cleaner to not worry about using constructed choice functions, but it's also perfectly acceptable now-a-days to use this axiom.

Plus, some really commonly used theorems require the axiom of choice to prove them. For example, the union of countably many countable sets being countable. This makes it really easy to accidentally use the axiom of choice in a proof.

Lemma 9.2: (Existence of a choice function) Given a collection $\mathcal{B}$ of nonempty sets (not necessarily disjoint), there exists a function

$$c : \mathcal{B} \longrightarrow \bigcup_{B \in \mathcal{B}} B$$

such that $c(B)$ is an element of B for each $B \in \mathcal{B}$.

Proof:
Given any set $B \in \mathcal{B}$, we define $B' = \{(B, b) \mid b \in B\}$. Because $B \neq \emptyset$, we know that $B' \neq \emptyset$ as well. Furthermore, given $B_1, B_2 \in \mathcal{B}$ if $B_1 \neq B_2$, then we have that the first element of all the pairs in B_1' are different from that of B_2'. So B_1' and B_2' are disjoint.

Now form the collection $\mathcal{C} = \{B' \mid B \in \mathcal{B}\}$. From before, we know that $\mathcal{C}$ is a collection of disjoint sets. So by the axiom of choice, there exists a set c consisting of exactly one element from each element of $\mathcal{C}$.

This set c is a subset of $\mathcal{B} \times \bigcup_{B \in \mathcal{B}} B$ which satisfies our definition of a choice function.

Hopefully it's obvious enough why c satisfies those properties.

A set A with an order relation $<$ is said to be well-ordered if every nonempty subset of A has a smallest element.

Tangent: inductiveness of $\mathbb{Z}_+$ is equivalent to the well-orderedness of $\mathbb{Z}_+$

This proof is taken from https://math.libretexts.org/ on their page for the well-ordering principle.

($\Longrightarrow$)
Suppose S is a nonempty subset of $\mathbb{Z}_+$ with no least element. Then let R be the set of lower bounds of S. Since 1 is the least element of $\mathbb{Z}_+$, we know that $1 \in R$.

Now given any $k \geq 1$, if $k \in R$, we know that $\{1, \ldots, k\}$ must be a subset of R. Also note that $R \cap S = \emptyset$ because if that wasn't true, we'd know that S has a least element. Therefore, $\{1, \ldots, k\} \cap S = \emptyset$. But then that shows that $k+1 \notin S$ since otherwise $k+1$ would be the least element of S. Furthermore, since no element of $\{1, \ldots, k\}$ is in S, we automatically have that $k+1 \in R$.

By induction, this means that $R = \mathbb{Z}_+$. Hence, we get a contradiction as S must be empty.

($\Longleftarrow$)
Let S be a subset of $\mathbb{Z}_+$ such that $1 \in S$ and $k \in S \Longrightarrow k+1 \in S$. Then suppose that $S \neq \mathbb{Z}_+$. In that case, we know that $S^{\mathsf{C}} \neq \emptyset$, and since $\mathbb{Z}_+$ is well-ordered, we know there is a least element α of S^{C}.

Because $1 \in S$, we know that $\alpha \geq 2$. But then consider that $1 \leq \alpha - 1 < \alpha$. Therefore, $\alpha - 1 \in S$, thus implying that $\alpha \in S$. This contradicts that $\alpha \in S^{\mathsf{C}}$.

From what I've heard, when defining the positive integers, one usally takes one of the two above properties as an axiom and then proves the other as a theorem. In Munkres' book, he starts with induction and proves well-orderedness.

Facts:

- If A with the order relation $<$ is well-ordered, then any subset of A is well-ordered as well with $<$ restricted to that subset.
- If A has the order relation $<_1$ and B has the order relation $<_2$ and both are well-ordered, then $A \times B$ is well-ordered with the dictionary order.
- Given any countable set A, we know there exists a bijection f from A to $\mathbb{Z}_+$. Hence, given $a, b \in A$, we can say that $a < b \Longleftrightarrow f(a) < f(b)$. Then, A is well-ordered by $<$ with the least element of any subset S of A being the element $\alpha \in A$ such that $f(\alpha)$ is the least element in $f(S)$.
- If a set A is well-ordered, then we can make a choice function $c : \mathcal{P}(A) \longrightarrow A$ using that well-ordering.

  Specifically, given any $B \subseteq A$, assign $c(B) = \beta$ where β is the least element of B.

  This is why we can pick elements of $\mathbb{Q}$ without worrying about the axiom of choice.

An important theorem (which I will hopefully prove soon) is:

The Well Ordering Theorem: If A is a set, there exists an order relation on A that is well-ordering.

Note: this theorem requires the axiom of choice to prove.

Exercise 10.5: Show that the well-ordering theorem implies the (infinite) axiom of choice.

Let $\mathcal{A}$ be a collection of disjoint sets. By the well-ordering theorem, we can pick an order relation on $\bigcup\limits_{A \in \mathcal{A}} A$ that is well-ordering.

Note that the previous sentence is carefully worded to only make use of the finite axiom of choice. Specifically, the order relation we are picking is an element of some subset of $\bigcup_{A \in \mathcal{A}} A \times \bigcup_{A \in \mathcal{A}} A$.

If we had instead picked a well-ordering for each $A \in \mathcal{A}$, then that would require the axiom of choice as we would be making potentially infinitely many arbitrary choices of order relations.

Now let $C = \{a \in \bigcup_{A \in \mathcal{A}} A \mid \exists A \in \mathcal{A} \ s.t. \ a \in A$ and $\forall b \in A, \ \ a \leq b\}$.

Then C fulfils the properties of the set that the axiom of choice would guarentee exists.

9/14/2024

Exercise 10.1: Show that every well-ordered set has the least-upper-bound property.

Let the set A with the order relation $<$ be well-ordered. Then consider any nonempty $B \subseteq A$ and suppose there exists $\alpha \in A$ such that $b < \alpha$ for all $b \in B$.

Let $U = \{a \in A \mid \forall b \in B, \ \ b \leq a\}$. Since $\alpha \in U$, we know that $U \neq \emptyset$. So, because A is well-ordered, we know that U has a least element β. This β is by definition the least upper bound of B. So $\sup B = \beta$.

Let X be a well-ordered set. Given $\alpha \in X$, let S_α denote the set $\{x \in X \mid x < \alpha\}$. We call S_α the section of X by α.

Lemma 10.2: There exists a well-ordered set A having a largest element Ω such that S_Ω is uncountable but every other section of A is countable.

Proof:
Starting off, let B be an uncountable well-ordered set. Then let C be the well-ordered set $\{1, 2\} \times B$ with the dictionary order. Clearly, given any $b \in B$, we have that $S_{(2,b)}$ is uncountable. So the set of $c \in C$ such that S_c is uncountable is not empty.

Let Ω be the least element of C such that S_Ω is uncountable. Then define $A = S_\Omega \cup \{\Omega\}$. This is called a minimal uncountable well-ordered set.

The reason we are considering $\{1, 2\} \times B$ instead of just B is that if we were just considering B, then we wouldn't be able to guarentee that there exists $b \in B$ such that S_b is uncountable.

User MJD on https://math.stackexchange.com wrote some good intuition for why this is.

While the set $\mathbb{Z}_+$ is countably infinite, all sections S_x of $\mathbb{Z}_+$ are finite. However, when considering $\{1, 2\} \times \mathbb{Z}_+$ with the dictionary order, we have that $S_{(2,1)}$ is countably infinite. Furthermore, all sections of $S_{(2,1)}$ are finite. Thus, $S_{(2,1)}$ would be a minimal *countable* well-ordered set.

We call a set described by lemma 10.2 $\overline{S}_\Omega = S_\Omega \cup \{\Omega\}$.

Theorem 10.3: If A is a countable subset of S_Ω, then A has an upper bound in S_Ω.

Proof:
Let A be a countable subset of S_Ω. For all $a \in A$, we know that S_a is countable. Therefore, $B = \bigcup\limits_{a \in A} S_a$ is also countable, meaning that $S_\Omega - B \neq \emptyset$.

If we pick $x \in S_\Omega - B$, we must have that x is an upper bound to A because if $x < a$ for some $a \in A$, we would have that $x \in S_a \subseteq B$.

If you combine this with exercise 10.1, we know that A has a least upper bound.

Exercise 10.6: Let S_Ω be a minimal uncountable well-ordered set.

(a) Show that S_Ω has no largest element.

Suppose $\alpha \in S_\Omega$ is the largest element of S_Ω. In that case, we'd have that $S_\alpha = S_\Omega - \{\alpha\}$. However, by theorem 10.3, we know that S_α is countable. This implies that $S_\Omega = S_\alpha \cup \{\alpha\}$ must also be countable, which is a contradiction.

(b) Show that for every $\alpha \in S_\Omega$, the subset $\{x \in S_\Omega \mid \alpha < x\}$ is uncountable.

Let $\alpha \in S_\Omega$. By the law of trichotomy, we know that:

$$S_\Omega = \{x \in S_\Omega \mid x < \alpha\} \cup \{\alpha\} \cup \{x \in S_\Omega \mid \alpha < x\}.$$

Now suppose $\{x \in S_\Omega \mid \alpha < x\}$ is countable. Then as both $\{x \in S_\Omega \mid x < \alpha\}$ and $\{\alpha\}$ are countable, we have a contradiction as the three's union must also be countable. But we know S_Ω isn't.

Some definitions I've been lacking:

1. Let A be a set and suppose x, y, z are any three different elements of A.

Simple [Default] Order Relation: ($<$)	Strict Partial Order Relation: ($\prec$)
Nonreflexivity: $x \not< x$ Transitivity: $x < y$ and $y < z \Rightarrow x < z$ Comparability: $x < y$ or $y < x$ is true	Nonreflexivity: $x \not\prec x$ Transitivity: $x \prec y$ and $y \prec z \Rightarrow x \prec z$

Basically, a partial order relation is allowed to not give an order for some pairings of elements. If someone just says a set is ordered, they mean the set is simply ordered.

2. Let A and B be sets ordered by $<_A$ and $<_B$ respectively. We say that A and B have the same order type if there exists an order-preserving bijection $f : A \longrightarrow B$, meaning that $\forall a_1, a_2 \in A,\ \ a_1 <_A a_2 \Longrightarrow f(a_1) <_B f(a_2)$.

   It is trivial to show that if f is an order-preserving bijection, then f^{-1} is also an order-preserving bijection.

3. If A is an ordered set and a and b are two different elements, then consider the set $S = \{x \in A \mid a < x < b\}$. If $S = \emptyset$ we say that b is the successor of a and a is the predecessor of b.

Exercise 10.2:

(a) Show that in a well-ordered set, every element except the largest (if one exists) has an immediate successor

Let A be a well-ordered set and let α be any element in A such that there exists $\beta \in A$ for which $\alpha < \beta$. Then consider the set $S = \{x \in A \mid \alpha < x < \beta\}$. If $S = \emptyset$, then we know α has β as its successor. Meanwhile, if $S \neq \emptyset$, then since A is well-ordered, we know that A has a least element γ. Thus, the set $\{x \in A \mid \alpha < x < \gamma\} = \emptyset$ and we know that γ is the successor of α.

(b) Find a set in which every element has an immediate successor that is not well-ordered.

Consider the set $\mathbb{Z}$ of all integers using the standard ordering. Then for any $n \in \mathbb{Z}$, we know that its successor is $n + 1$. At the same time though, the set of all negative integers has no least element. So $\mathbb{Z}$ is not well-ordered by $<$.

Exercise 10.6:

(c) Let X_0 be the subset of S_Ω consisting of all elements x such that x has no immediate predecessor. Show that X_0 is uncountable.

Suppose X_0 is bounded above by some $\alpha \in S_\Omega$. Thus, there is a predecessor $x \in S_\Omega$ for any y in the set $T = \{z \in S_\Omega \mid z > \alpha\}$.

Now define a function $f : \mathbb{Z}_+ \longrightarrow T$ such that $f(1) =$ the least element of T and $f(n) =$ the successor of $f(n-1)$ for all $n > 1$. We know this function is well-defined because S_Ω has no largest element according to exercise 10.6.a. So, all elements of S_Ω and thus T have a successor by exercises 10.2.a, meaning our formula for $f(n)$ is always defined no matter what $f(n-1)$ is. Hence, the principle of recursive definition guarentees a unique f exists.

Now it's easy to show that f is injective. For suppose that given some $x, n \in \mathbb{Z}_+$ we had that $f(x) = f(x+n)$. Then that would mean that:

$$f(x) < f(x+1) < \cdots < f(x+n-1) < f(x+n) = f(x)$$

Hence we have a contradiction as $f(x) < f(x)$.

Next, we show that f is surjective. Suppose the set $R = T - f(\mathbb{Z}_+) \neq \emptyset$. Then since S_Ω and hence T is well-ordered, we know that R has a least element β. But note that β has a predecessor γ which isn't in R. More specifically, since we know that the least element of T is in $f(\mathbb{Z}_+)$, we know that γ is at least the least of element of T. So $\gamma \in T$.

Thus we conclude that $\gamma \in T - (T - f(\mathbb{Z}_+)) = f(\mathbb{Z}_+)$, meaning there exists N such that $f(N) = \gamma$. But this means that $f(N+1) = \beta$, which contradicts that β is the least element of R.

With that, we've now shown that $f : \mathbb{Z}_+ \longrightarrow T$ is a bijection, meaning that T is countable. However, this contradicts exercise 10.6.b. which asserts that T is uncountable.

Therefore, we conclude that X_0 cannot be bounded above. And by theorem 10.3, that means that X_0 can't be a countable subset of S_Ω.

Exercise 10.4:

(a) Let $\mathbb{Z}_-$ be the set of negative integers in the usual order. Show that a simply ordered set A fails to be well-ordered if and only if it contains a subset having the same order type as $\mathbb{Z}_-$.

($\Longleftarrow$)
If for some $B \subseteq A$, we have that $f : \mathbb{Z}_- \longrightarrow B$ is an order preserving bijection, then we must have that B has no least element. Hence, not all subsets of A have a least element, meaning that A is not well-ordered.

($\Longrightarrow$)
If A is not well ordered, then we know there is a set $B \subseteq A$ with no least element. Now using the axiom of choice, choose any $\beta_1 \in B$. Then for all $n > 1$, choose $\beta_n \in B_{\beta_{n-1}}$. In other words, choose $\beta_n \in B$ such that $\beta_n < \beta_{n-1}$.

Finally, define $f : \mathbb{Z}_- \longrightarrow \{\beta_n \mid n \in \mathbb{Z}_+\}$ by the rule: $f(n) = \beta_{-n}$. This f is an order preserving bijection. Thus, the set $\{\beta_n \mid n \in \mathbb{Z}_+\} \subseteq A$ has the same order type as $\mathbb{Z}_-$.

(b) Show that if A is simply ordered and every countable subset of A is well-ordered, then A is well-ordered.

It's easy to show the contrapositive of this statement.

If A is not well-ordered, then by part a. we know there exists a set $B \subseteq A$ and a function $f : \mathbb{Z}_- \longrightarrow B$ that is an order-preserving bijection. Clearly, B has no least element. Also, the function $g(n) = f(-n)$ gives a bijection from $\mathbb{Z}_+$ to B, meaning that B is countable. Hence, we have shown that B is a countable subset of A that is not well-ordered.

Let J be a well-ordered set. A subset J_0 of J is said to be inductive if for every $\alpha \in J$, we have that $(S_\alpha \subseteq J_0) \implies \alpha \in J_0$.

Exercise 10.7: (The principle of transfinite induction) If J is a well-ordered set and J_0 is an inductive subset of J, then $J_0 = J$.

Proof:
Suppose $J_0 \neq J$. That would mean the set $J - J_0$ is nonempty. So let α be the least element of $J - J_0$. We know that S_α must be disjoint to $J - J_0$, meaning that $S_\alpha \in J_0$. But then by the inductiveness of J_0, we must have that $\alpha \in J_0$. This contradicts that α is the least element of $J - J_0$.

Exercise 10.10: (Theorem) Let J and C be well-ordered sets; assume that there is no surjective function mapping a section of J onto C. Then there exists a unique function $h : J \longrightarrow C$ satisfying for each $x \in J$ the equation:

$$(*) \qquad h(x) = \text{smallest element of } C - h(S_x).$$

Proof:

(a) If h and k map sections of J or all of J into C and satisfy $(*)$ for all x in their domains, then $h(x) = k(x)$ for all x in both domains.

Proof:
Suppose not. Let y be the smallest element of the domains of h and k for which $h(y) \neq k(y)$. Then note that $\forall z \in S_y$, we must have that $h(z) = k(z)$. Thus, we get a contradiction since:

$$h(y) = \text{smallest}(C - h(S_y)) = \text{smallest}(C - k(S_y)) = k(y).$$

(b) If there exists a function $h : S_\alpha \longrightarrow C$ satisfying $(*)$, then there exists a function $k : S_\alpha \cup \{\alpha\} \longrightarrow C$ satisfying $(*)$.

Proof:
Since there is no surjective function mapping a section of J onto C, we know that $C - h(S_\alpha) \neq \emptyset$. Hence, we can define $k(x) = h(x)$ for $x < \alpha$ and $k(\alpha) = \text{smallest}(C - h(S_\alpha))$.

(c) If $K \subseteq J$ and for all $\alpha \in K$ there exists $h_\alpha : S_\alpha \longrightarrow C$ satisfying $(*)$, then there exists a function $k : \bigcup\limits_{\alpha \in K} S_\alpha \longrightarrow C$ satisfying $(*)$.

Proof:
Define $k = \bigcup\limits_{\alpha \in K} h_\alpha$.

We know k is a valid function definition because part (a) guarentees that for all $\alpha_1, \alpha_2 \in K$ greater than x, we have that $h_{\alpha_1}(x) = h_{\alpha_2}(x)$. Plus, given any $x \in \bigcup\limits_{\alpha \in K} S_\alpha$, we know that there is $\alpha \in K$ such that $\forall y \in S_x,\ \ k(y) = h_\alpha(y)$.
This shows that k satisfies $(*)$ at any x due to the relevant h_α satisfying $(*)$.

(d) For all $\beta \in J$, there exists a function $h_\beta : S_\beta \longrightarrow C$ satisfying $(*)$.

Proof:
Let J_0 be the set of all $\beta \in J$ for which there exists a function $h_\beta : S_\beta \longrightarrow C$ satisfying $(*)$. Our goal is to show that J_0 is inductive. That way, we can conclude by transfinite induction (exercise 10.7) that $J_0 = J$.

Pick any $\beta \in J$ and suppose $S_\beta \in J_0$.

Case 1: β has an immediate predecessor α.

Then $S_\beta = S_\alpha \cup \{\alpha\}$. So, knowing that h_α satisfying $(*)$ exists, we can use part (b) to define h_β satisfying $(*)$.

Case 2: β has no immediate predecessor.

Then $S_\beta = \bigcup\limits_{\alpha \in S_\beta} S_\alpha$.

And since we assumed that there exists $h_\alpha : S_\alpha \longrightarrow C$ satisfying $(*)$ for all $\alpha \in S_\beta$, we thus know by part (c) that there exists a function from $\bigcup\limits_{\alpha \in S_\beta} S_\alpha = S_\beta$ to C satisfying $(*)$.

Thus in both cases, we have shown that $S_\beta \in J_0$ implies that $h_\beta : S_\beta \longrightarrow C$ satisfying $(*)$ exists. Or in other words, $S_\beta \in J_0 \Longrightarrow \beta \in J_0$.

(e) Finally, we now finish proving this theorem.

Case 1: J has a max element β.

Then since we know there exists $h_\beta : S_\beta \longrightarrow C$ satisfying $(*)$, we can apply part (b) to get a function h from $J = S_\beta \cup \{\beta\}$ to C satisfying $(*)$.

Case 2: J has no max element.

Then $J = \bigcup_{\beta \in J} S_\beta$.

And since there exists $h_\beta : S_\beta \longrightarrow C$ satisfying $(*)$ for all $\beta \in J$, we can thus apply part (c) to get a function h from $J = \bigcup_{\beta \in J} S_\beta$ to C satisfying $(*)$.

9/17/2024

Theorem (The Hausdorff maximum principle): Let A be a set and let $\prec$ be a strict partial order on A. Then there exists a maximal simply ordered subset B of A.

In other words, there exists a subset B of A such that B is simply ordered by $\prec$ and no subset of A that properly contains B is simply ordered by $\prec$.

Proof:
To start out, let J be a set well-ordered by $<$ such that the elements of A are indexed in a bijective fashion by the elements of J. In other words, $A = \{a_\alpha \in A \mid \alpha \in J\}$.

Assuming the well-ordering theorem, we know that J exists. Specifically let J refer to the same set as A but equip J with the well-ordering $<$ that we know exists instead of the partial ordering $\prec$ which we equipped A.

Now our goal is to construct a function $h : J \longrightarrow \{0, 1\}$ such that $h(\alpha) = 1$ if a_α is in our maximal simply ordered subset of A and $h(\alpha) = 0$ otherwise. To do this, we rely on the **general principle of recursive definition**.

Theorem: (General principle of recursive definition):
Let J be a well-ordered set and C be any set. Given a function $\rho : \mathcal{F} \longrightarrow C$ where $\mathcal{F}$ is the set of all functions mapping sections of J into C, we have that there exists a unique functon $h : J \longrightarrow C$ satisfying that $h(\alpha) = \rho(h|_{S_\alpha})$ for all $\alpha \in J$.

The proof for this is supplementary exercise 1. of this chapter. But I'm not going to do it because it's mostly identical to exercise 10.10.

Given any $\alpha \in J$ and $f : S_\alpha \longrightarrow \{0, 1\}$, define $\rho(\alpha) = 1$ if $a_\alpha \in A$ is comparable to all $a_\beta \in A$ such that $\beta \in f^{-1}(1)$ (the preimage of 1).

Note that a_α is comparable to a_β if either $a_\alpha \prec a_\beta$ or $a_\beta \prec a_\alpha$.

Then by the general principle of recursive definition, we know a unique function $h : J \longrightarrow \{0,1\}$ exists such that for all $\alpha \in J$, we have that $h(\alpha) = 1$ only when a_α is comparable to all $a_\beta \in A$ such that $\beta \in S_\alpha$ and $h(\beta) = 1$.

Let $B = \{a_\alpha \in A \mid \alpha \in J \text{ and } h(\alpha) = 1\}$. Then given any $a_\alpha, a_\beta \in B$ such that $\alpha < \beta$, we know that either $a_\alpha \prec a_\beta$ or $a_\beta \prec a_\alpha$. Hence, B is simply ordered by $\prec$. At the same time, if $a_\gamma \notin B$, then we know $h(\gamma) = 0$, meaning there exists $a_\alpha \in B$ such that $\alpha < \gamma$ and a_γ is not comparable to a_α. This shows that any set properly containing B is not simply ordered by $\prec$.

Note that the maximal simply ordered subset B is not unique. In fact, choosing a different well-ordering of J is likely to give a completely different maximal simply ordered subset.

Also, B is not empty because any set with one element is simply ordered by $\prec$.

Let A be a set and let $\prec$ be a strict partial order on A. If B is a subset of A, we say an upper bound on B is an element c of A such that for every $b \in B$, either $b = c$ or $b \prec c$. A maximal element of A is an element m of A such that for no element a of A does the relation $m \prec a$ hold.

Zorn's Lemma: Let A be a set that is strictly partially ordered. If every simply ordered subset of A has an upper bound in A, then A has a maximal element.

Proof:
By the Hausdorff maximum principle, there exists a maximal simply ordered subset B of A. Let c be an element of A that is an upperbound to B. We claim that c is a maximal element of A. For suppose there exists $d \in A$ such that $c \prec d$. We know $d \notin B$ since that would imply $d \prec c$. But by the transitivity of $\prec$, we know that $b \preceq c \prec d \Longrightarrow b \prec d$ for all $b \in B$. Hence, $B \cup \{d\}$ is simply ordered by $\prec$. This contradicts that B is a maximal simply ordered subset of A.

Exercise 11.1: If a and b are real numbers, define $a \prec b$ if $b - a$ is positive and rational.

- It's easy to show that $\prec$ is a strict partial order. After all, for all $a \in \mathbb{R}$, we have that $a - a$ is not positive. Also, if $a \prec b$ and $b \prec c$, then we know that $b - a = p$ and $c - b = q$ where $p, q \in \mathbb{Q}_+$. But then $c - a = c - b + b - a = p + q \in \mathbb{Q}_+$. So $a \prec c$.

- Clearly, given any $x \in \mathbb{R}$, the maximal simply ordered set containing x is the set $\{x + p \mid p \in \mathbb{Q}\}$.

Tangent: I never got around to writing this down last quarter. So here's a proof that assuming the axiom of choice, non-Lebesgue measurable sets exist.

Let $\mathcal{B}$ be the collection of sets of the form $S_x = [0,1] \cap \{x + p \mid p \in \mathbb{Q}\}$ where x is any real number. Obviously, all the sets in $\mathcal{B}$ are nonempty. We also claim that all the sets in $\mathcal{B}$ are disjoint. For suppose $S_x, S_y \in \mathcal{B}$ and $S_x \cap S_y \neq \emptyset$. Then fix $c \in S_x \cap S_y$ and consider any $a \in S_x$ and $b \in S_y$.

We know $c - x = p_1$, $a - x = p_2$, $c - y = q_1$, and $b - y = q_2$ where $p_1, p_2, q_1, q_2 \in \mathbb{Q}$. Thus, we have that $a - y = (a - x) + (x - c) + (c - y) = p_2 - p_1 + q_1 \in \mathbb{Q}$. Similarly, we have that $b - x = (b - y) + (y - c) + (c - x) = q_2 - q_1 + p_1 \in \mathbb{Q}$. This tells us that $a \in S_y$ and $b \in S_x$. And since this works for all $a \in S_x$ and $b \in S_y$, we thus must have that $S_x = S_y$.

Now using the axiom of choice, let V be a set containing one element from each set in $\mathcal{B}$.

To show that V is nonmeasurable, we'll reach a contradiction by supposing V is measurable. Let $q_1, q_2, \ldots$ be an enumeration of all the rational numbers in the set $[-1, 1]$. Then having defined $V + q_n = \{v + q_n \mid v \in V\}$, consider the set: $\bigcup\limits_{n \in \mathbb{Z}_+} (V + q_n)$.

Obviously, since $V \subseteq [0,1]$, we know that $\bigcup\limits_{n \in \mathbb{Z}_+} (V + q_n) \subseteq [-1, 2]$.

Also, consider any $x \in [0,1]$ and let v be the element of V which was chosen from the set $S_x \in \mathcal{B}$. Then $v - x = p$ where p is some rational number in $[-1, 1]$. So, we also know that $[0,1] \subseteq \bigcup\limits_{n \in \mathbb{Z}_+} (V + q_n)$. This means that $1 \leq \mu(\bigcup\limits_{n \in \mathbb{Z}_+} (V + q_n)) \leq 3$.

But now note that for any $n, m \in \mathbb{Z}_+$, we have that $n \neq m \Longrightarrow V + q_n \cap V + q_m = \emptyset$. To prove this, assume $V + q_n \cap V + q_m \neq \emptyset$. Thus, there would exist $v, u \in V$ such that $v + q_n = u + q_m$. In turn, we'd have that $v - u = q_m - q_n \in \mathbb{Q}$, which means that $v \in S_u$. However, this contradicts that V has only one element of S_u.

Now since μ is countably additive, we have that $\mu(\bigcup\limits_{n \in \mathbb{Z}_+} (V + q_n)) = \sum\limits_{n=1}^{\infty} \mu(V + q_n)$.

Finally, note that $\mu(V) = \mu(V + q_n)$ for all n. Thus $\sum\limits_{n=1}^{\infty} \mu(V + q_n) = \sum\limits_{n=1}^{\infty} \mu(V)$ is either 0 or ∞.

But this contradicts our earlier finding that the measure was between 1 and 3. So, we conclude that $V \notin \mathcal{M}(\mu)$. ■

Exercise 11.2:

(a) Let $\prec$ be a strict partial order on the set A. Define a (non-strict partial) relation $\preceq$ on A by letting $a \preceq b$ if either $a \prec b$ or $a = b$. Show that this relation has the following properties which are called the *partial order axioms*:

(i) $a \preceq a$ for all $a \in A$

This is true because $a = a$ for all $x \in A$.

(ii) $a \preceq b$ and $b \preceq a \Longrightarrow a = b$.

Given any $a, b \in A$ such that $a \preceq b$ and $b \preceq a$, if $a \neq b$, then we'd have that $a \prec b$ and $b \prec a$. This gives a contradiction since $a \prec b \prec a \Longrightarrow a \prec a$ which is not allowed.

(iii) $a \preceq b$ and $b \preceq c \Longrightarrow a \preceq c$

Proving this is a matter of considering six rather trivial cases.

(b) Let P be a relation on A satisfying the three axioms above. Define a relation S on A by letting $a\,S\,b$ if $a\,P\,b$ and $a \neq b$. Show that S is a strict partial order on A.

Obviously, $a \not{S}\, a$ for all $a \in A$ since $a = a$ for all $a \in A$. Meanwhile, suppose $a\,S\,b$ and $b\,S\,c$. Then we know that $a\,P\,b$ and $b\,P\,c$, meaning that $a\,P\,c$. So we just need to show that $a \neq c$ and then we will have proven that $a\,S\,c$.

Suppose $a = c$. Then we know that $c\,P\,a$ and $a\,P\,b$, meaning that $c\,P\,b$. But then since $b\,P\,c$, we know that $b = c$. This contradicts that $b\,S\,c$.

In the next exercises we will explore some equivalent theorems to the Hausdorff maximum principle and Zorn's lemma.

Exercise 11.5: Show that Zorn's lemma implies the following:

Kuratowski's Lemma: Let $\mathcal{A}$ be a collection of sets. Suppose that for every subcollection $\mathcal{B}$ of $\mathcal{A}$ that is simply ordered by proper inclusion, the union of the elements of $\mathcal{B}$ belongs to $\mathcal{A}$. Then $\mathcal{A}$ has an element that is properly contained in no other element of $\mathcal{A}$.

To be clear, given any $A, B \in \mathcal{A}$, we defined above that $A \prec B$ if $A \subset B$. Importantly, our assumption about $\mathcal{A}$ means that every subcollection $\mathcal{B}$ of $\mathcal{A}$ that is simply ordered by $\prec$ has an upper bound in $\mathcal{A}$: $\bigcup\limits_{B \in \mathcal{B}} B$.

Thus by Zorn's lemma, we know that $\mathcal{A}$ has a maximal element C. And since there is no element $D \in \mathcal{A}$ such that $C \prec D$, we know that C is properly contained by no sets in $\mathcal{A}$.

Exercise 11.6: A collection $\mathcal{A}$ of subsets of a set X is said to be of *finite type* provided that a subset B of X belongs to $\mathcal{A}$ if and only if every finite subset of B belongs to $\mathcal{A}$. Show that the Kuratowski lemma implies the following:

Tukey's Lemma: Let $\mathcal{A}$ be a collection of sets. If $\mathcal{A}$ is of finite type, then $\mathcal{A}$ has an element that is properly contained in no other element of $\mathcal{A}$.

To start off I want to clarify that $\mathcal{A}$ being of finite types means both that:

1. For each $A \in \mathcal{A}$, every finite subset of A belongs to $\mathcal{A}$.
2. If every finite subset of a given set A belongs to $\mathcal{A}$, then A belongs to $\mathcal{A}$.

Now let $\mathcal{B}$ be any subcollection of $\mathcal{A}$ that is simply ordered by proper inclusion. Next, consider the set $S = \bigcup_{B \in \mathcal{B}} B$. We want to show that any finite subset of S is in $\mathcal{A}$.

To do this, let $n \in \mathbb{Z}_+$ and consider any subset $\{b_1, b_2, \ldots, b_n\}$ of S with n elements. Note that for each $1 \leq i \leq n$, there exists $B_i \in \mathcal{B}$ such that $b_i \in B_i$. Then since $\{B_1, B_2, \ldots B_n\}$ is a simply ordered finite set, we know that it has a maximum element B_m such that $B_i \subseteq B_m$ for all i. Hence, we have that $\{b_1, b_2, \ldots b_n\}$ is contained by some B_m in $\{B_1, B_2, \ldots, B_n\} \subseteq \mathcal{B}$. Because $\mathcal{A}$ is of finite type, this tells us that $\{b_1, b_2, \ldots b_n\} \in \mathcal{A}$.

Since we showed above that any finite subset of S is in $\mathcal{A}$, we can thus conclude because $\mathcal{A}$ is of finite type that $S \in \mathcal{A}$. And so, we have now proven the hypothesis of Kuratowski's lemma, meaning that $\mathcal{A}$ must have a set that is properly contained in other element of $\mathcal{A}$.

Exercise 11.7: Show that the Tukey lemma implies the Hausdorff maximum principle.

Let A be a set with the strict partial order $\prec$. Then let $\mathcal{A}$ be the collection of all subsets of A that are simply ordered by $\prec$. We shall show below that $\mathcal{A}$ is of finite type.

1. Suppose $B \in \mathcal{A}$. Then given any subset C of B (finite or not), we know that C is also simply ordered by $\prec$. So $C \in \mathcal{A}$.
2. Let $B \subseteq A$ and suppose every finite subset of B is in $\mathcal{A}$. Then given any two different elements $b_1, b_2 \in B$, we know that $\{b_1, b_2\} \in \mathcal{A}$, meaning that either $b_1 \prec b_2$ or $b_2 \prec b_1$. In other words, B is simply ordered by $\prec$, meaning that $B \in \mathcal{A}$.

Because $\mathcal{A}$ is of finite type, we know that $\mathcal{A}$ has an element that is properly contained in no other element of $\mathcal{A}$. Or in other words, there exists a subset of A which is simply ordered by $\prec$ and not properly contained in any other subset of A that is simply ordered by $\prec$.

9/19/2024

In the past 14 pages, we've learned a lot about the axiom of choice. All the blue arrows in the diagram to the right represent proofs we've already done. Meanwhile, the red arrows represent proofs that Munkres left to the supplementary exercises of section 1 of his book. We're gonna do those proofs now.

$$\begin{tikzcd}
 & & \text{Non-measurable Sets} \\
\text{Axiom of Choice} \arrow[urr, Rightarrow] \arrow[d, Rightarrow, shift left] & & \\
\text{Well Ordering Theorem} \arrow[u, Rightarrow, shift left] \arrow[d, Rightarrow, shift left] & & \\
\text{Maximum Principle} \arrow[u, Rightarrow, shift left] \arrow[d, Rightarrow] & & \text{Tukey's Lemma} \arrow[ll, Rightarrow] \\
\text{Zorn's Lemma} \arrow[rr, Rightarrow] & & \text{Kuratowski's Lemma} \arrow[u, Rightarrow]
\end{tikzcd}$$

Exercise 1: (General principle of recursive definition)

We already addressed this before. I'm skipping proving this because the proof is mostly identical to exercise 10.10. In fact, exercise 10.10 is just this exercise but with a specific $\rho : \mathcal{F} \longrightarrow C$.

Exercise 2:

(a) Let J and E be well-ordered sets and let $h : J \longrightarrow E$. Show that the following two statement are equivalent:

(i) h is order preserving and its image is E or a section of E.

(ii) $h(\alpha) = \text{smallest}(E - h(S_\alpha))$ for all $\alpha \in J$.

(i) $\Longrightarrow$ (ii):
Given any $\alpha \in J$, we know that $h(\alpha)$ must be an upper bound to $h(S_\alpha)$. Now suppose $\exists \beta \in S_{h(\alpha)}$ such that $\beta \notin h(S_\alpha)$. Because of our assumption about the image of h, we know that $\beta \in h(J)$, meaning there exists $\gamma \in J$ such that $h(\gamma) = \beta$. But because h is order-preserving, we must have that $\beta < f(\alpha) \Longrightarrow \gamma < \alpha$. This contradicts that $\beta \notin h(S_\alpha)$.

With that, we've now shown that $h(S_\alpha) = S_{h(\alpha)}$. In turn, this shows that $h(\alpha)$ is the smallest element in $E - h(S_\alpha)$.

(ii) $\Longrightarrow$ (i):
It's easy to show h is order preserving. Let $\alpha, \beta \in J$ such that $\alpha < \beta$. Then $h(S_\alpha) \subset h(S_\beta)$, meaning that $E - h(S_\beta) \subset E - h(S_\alpha)$. And since the least element of $E - h(S_\alpha)$ is not in $E - h(S_\beta)$, that means that $h(\alpha) = \text{smallest}(E - h(S_\alpha)) < \text{smallest}(E - h(S_\beta)) = h(\beta)$.

As for showing the other property of h, let $J_0 = \{\alpha \in J \mid h(S_\alpha) = S_{h(\alpha)}\}$. Now suppose that for some $\alpha \in J$, we have that $S_\alpha \subseteq J_0$. Then we can show that $\alpha \in J_0$.

Case 1: α has an immediate predecessor β.

Then $S_\alpha = S_\beta \cup \{\beta\}$, meaning that:
$$h(S_\alpha) = h(S_\beta) \cup \{h(\beta)\} = S_{h(\beta)} \cup \{h(\beta)\}.$$

Since $h(\alpha)$ is the least element of E not in $h(S_\alpha)$. We can thus say that $S_{h(\beta)} \cup \{h(\beta)\} = S_{h(\alpha)}$.

Case 2: α has no immediate predecessor.

Then we have that $h(S_\alpha) = h(\bigcup\limits_{\beta \in S_\alpha} S_\beta) = \bigcup\limits_{\beta \in S_\alpha} h(S_\beta) = \bigcup\limits_{\beta \in S_\alpha} S_{h(\beta)}$.

Hence, $h(S_\alpha)$ is a section of E, and since $h(\alpha)$ is the least element not in that section, we can conclude that $h(S_\alpha) = S_{h(\alpha)}$.

By transfinite induction, we thus know that $J_0 = J$. So finally, we consider two cases.

Case 1: J has a max element α.

Then $h(J) = h(S_\alpha) \cup \{h(\alpha)\} = S_{h(\alpha)} \cup \{h(\alpha)\}$. And since $h(\alpha)$ is the least element not in $S_{h(\alpha)}$, we thus know that $h(J)$ is either a section of or the whole of E.

Case 2: J has no max element.

Then $h(J) = h(\bigcup\limits_{\alpha \in J} S_\alpha) = \bigcup\limits_{\alpha \in J} h(S_\alpha) = \bigcup\limits_{\alpha \in J} S_{h(\alpha)}$.

So, $h(J)$ is either a section of or the whole of E.

(b) If E is a well-ordered set, show that no section of E has the same order type as E, nor do any two different sections of E have the same order type.

Let J be any well-ordered set. By combining part (a) of this exercise with exercise 10.10 (which is a special case of the general principle of recursive definition), we know that there is at most one order preserving map from J to E whose image is either E or a section of E. Hence, J can only have the same order type as one of either the entirety of E or one section of E.

Based on that fact, we can get an easy contradiction if we assume that the claim of part (b) is false.

9/21/2024

Unfortunately I tested positive for Covid on the two days ago. So I've been really delirious. However, right now I'm in an airport in the process of moving back out to California (great idea). And since my flight just got delayed, I feel like I might as well kill time and try to do some math.

Exercise 3: Let J and E be well-ordered sets, and suppose there is an order-preserving map $k : J \longrightarrow E$. Using exercises 1 and 2, show that J has the order type of one of either E or one section of E.

Pick any $e_0 \in E$. Then define $h : J \longrightarrow E$ by the rule:

$$h(\alpha) = \begin{cases} \text{smallest}(E - h(S_\alpha)) & \text{if } h(S_\alpha) \neq E \\ e_0 & \text{otherwise} \end{cases}$$

Note that the second case of our definition of h is just included to ensure that h is well-defined before we begin the proof in earnest. I mention that because our goal now is to show that the second case will never apply.

Let $J_0 = \{\alpha \in J \mid h(\alpha) \leq k(\alpha)\}$. Then suppose that for some $\alpha \in J$, we have that $S_\alpha \subseteq J_0$. Because k is order preserving, we know that $k(\alpha) > k(\beta) \geq h(\beta)$ for all $\beta \in S_\alpha$. Hence, $k(\alpha) \notin h(S_\alpha)$, meaning that $h(S_\alpha) \neq E$. So, we conclude that $h(\alpha) = \text{smallest}(E - h(S_\alpha))$. And since $k(\alpha) \in E - h(S_\alpha)$, we thus know that $h(\alpha) \leq k(\alpha)$

Therefore, $\alpha \in J_0$. By transfinite induction, this proves that $J = J_0$. The reason this is relevant is that we can now say that $k(\alpha)$ is never in $h(S_\alpha)$, meaning that $E - h(S_\alpha) \neq \emptyset$. So $h(\alpha)$, will never be determined by the second case of our definition above.

By exercise 2, we know that $h : J \longrightarrow E$ is the unique order-preserving map whose image is either E or a section of E. Thus, J has the same order type as exactly one of either the entirety of E or one section of E.

Exercise 4: Use exercises 1-3 to prove the following:

(a) If A and B are well-ordered sets, then exactly one of the following three conditions holds: A and B have the same order type, A has the order type of a section of B, or B has the order type of a section of A.

To start off, it's relatively easy to show that at most one of the above three cases is true. After all, A having the same order type as B as well as a section of B contradicts exercise 2. Similarly B having the same order type as A as well as a section of A contradicts exercise 2.

Meanwhile, to find a contradiction if A has the order type of S_β and B has the order type of S_α where $\alpha \in A$ and $\beta \in B$, let $h : A \longrightarrow S_\alpha$ be the function defined by the rule $h(a) = g(f(a))$ where f is the order-preserving bijection from A to S_β and g is the order-preserving bijection from B to S_α.

Then given any $a, b \in A$, we know that:

$$a < b \Rightarrow f(a) < f(b) \Rightarrow h(a) = g(f(a)) < g(f(b)) = h(b).$$

Hence, h is an order preserving map from A to S_α. This gives us a contradiction since exercise 3 would then imply that A has the same order type as either S_α or a section of S_α (which would still be a section of A).

Now, what's left to show is that at least one of the three above cases must be true. Unfortunately, the hinted route for showing this uses an exercise I didn't do. And right now I really don't want to do that exercise. So I'm just going to write out the thing I was supposed to have proven earlier.

Exercise 10.8.a:

Let A_1 and A_2 be disjoint sets well-ordered by $<_1$ and $<_2$ respectively. Then define an order relation on $A_1 \cup A_2$ by letting $a < b$ either if $a, b \in A_1$ and $a <_1 b$, or if $a, b \in A_2$ and $a <_2 b$, or if $a \in A_1$ and $b \in A_2$. This is a well-ordering of $A_1 \cup A_2$.

Let $A' = \{A\} \times A$ and let $B' = \{B\} \times B$. That way, so long as $A \neq B$, we know that A' and B' are disjoint. (The case where $A = B$ is trivial.)

It's hopefully obvious that the well-orderings of A and B can be used to well-order A' and B'. For A', define $(A, a_1) <_{A'} (A, a_2)$ if $a_1 <_A a_2$. Similarly, define the analogous ordering for B'. Clearly, A and A' have the same order type, as do B and B'. Also, given any $\alpha \in A$ and $\beta \in B$, S_α and $S_{(A,\alpha)}$ have the same order type, as do S_β and $S_{(B,\beta)}$

Next, define a well-ordering on $A' \cup B'$ by letting $a' < b'$ if either $a', b' \in A'$ and $a' <_{A'} b'$, or if $a', b' \in B'$ and $a' <_{B'} b'$, or if $a' \in A'$ and $b' \in B'$.

Note that the inclusion function from B' to $A' \cup B'$ is an order-preserving map. Thus, by exercise 3, we know that B' has the order type of one of either $A' \cup B'$ or one section of $A' \cup B'$.

Case 1: B' has the order type of a section S_α of $A' \cup B'$.

If $\alpha \in A'$, then B' has the order type of a section of A', meaning B has the order type of a section of A.

If α is the first element of B, then B' has the same order type as A', meaning B has the same order type as A.

If $\alpha \in B'$, then there exists an order preserving bijection from B' to $A' \cup \{b \in B' \mid b <_{B'} \alpha\}$. So let f be the inverse of that bijection but with it's domain restricted to just A'. Since f is also an order-preserving map, we know by exercise 3 that A' has the order type of either B' or a section of B'. This would mean that A has the order type of either B or a section of B.

Case 2: B' has the order type of $A' \cup B'$.

Let f be the inverse of the order preserving bijection from B' to A', except with it's inverse restricted to just A'. Since f is also an order-preserving map, we know by exercise 3 that A' has the order type of either B' or a section of B'. This would mean that A has the order type of either B or a section of B.

With that, we've now shown that at least one of the three cases posed by the exercise will always be true.

(b) Suppose that A and B are well-ordered sets that are uncountable such that every section of A and of B is countable. Show that A and B have the same order type.

If A did not have the same order type as B, then by part (a) of this exercise we would know that either A has the order type of a section of B or B has the order type of a section of A. However, that would suggest the existence of a bijection between a countable set and an uncountable set, which by definition is not possible.

9/23/2024

Exercise 5: Let X be any set and let $\mathcal{A}$ be the collection of all pairs $(A, <)$ where A is a subset of X and $<$ is a well-ordering of A. Define:

$$(A, <) \prec (A', <')$$

if $(A, <)$ equals a section of $(A', <')$.

In other words, $A = S_\alpha = \{a \in A' \mid a <' \alpha\}$ where $\alpha \in A'$, and $<$ is the order relation $<'$ restricted to A.

(a) Show that $\prec$ is a strict partial order on $\mathcal{A}$.

Clearly no A is a section of itself. So $(A, <) \not\prec (A, <)$.

Also if $(A, <_A) \prec (B, <_B) \prec (C, <_C)$, then we know that A is a section of a section of C (which is still a section). Plus, $<_A$ is just $<_C$ restricted to $<_A$. Hence, $(A, <) \prec (C, <_C)$.

(b) Let $\mathcal{B}$ be a subcollection of $\mathcal{A}$ that is simply ordered by $\prec$. Define B' to be the union of the sets B for all $(B, <) \in \mathcal{B}$, and define $<'$ to be the union of the relations $<$ for all $(B, <) \in \mathcal{B}$. Show that $(B', <')$ is a well-ordered set.

To start, let's quickly double check that $<'$ is a valid order relation on B'.

(i) Given any $b \in B'$, if $b \in B$ for any $(B, <) \in \mathcal{B}$, then we know that $(b, b) \notin <$. So $(b, b) \notin <'$.

(ii) Suppose $a, b \in B'$ such that $(a, b) \notin <'$. Then for all $(B, <) \in \mathcal{B}$ such that $a, b \in B$, we know that $(a, b) \notin <$, meaning that $(b, a) \in <$. So $(b, a) \in <'$.

(iii) Given $a, b, c \in B'$, suppose $a <' b <' c$. Then there exists $(B_1, <_1)$ and $(B_2, <_2)$ in $\mathcal{B}$ such that $(a, b) \in <_1$ and $(b, c) \in <_2$. Now by how we defined $\mathcal{B}$, we know that either $<_1 \subset <_2$ or $<_2 \subset <_1$. Thus, we know $(a, b), (b, c) \in \{<_i\}$ for some $i \in \{1, 2\}$. Hence, $(a, c) \in <_i$, meaning that $(a, c) \in <'$.

Next, we show that B' is well-ordered by $<'$.

Let $S \subseteq B'$ be nonempty and pick any element β in S. Then we know there exists $(B_1, <_1) \in \mathcal{B}$ such that $\beta \in B_1$. Also, B_1 is well-ordered by $<$. So let α be the least element (using $<_1$) of $B_1 \cap S$.

We claim that α is the least element (using $<'$) of S. To prove this, suppose there exists $c \in S$ such that $c <' a$. Then we know $(c, \alpha) \in <_2$ for some $(B_2, <_2) \in \mathcal{B}$. Importantly, $(B_1, <_1) \neq (B_2, <_2)$ since otherwise we'd have chosen α differently. So one must be a section of the other.

- If $(B_2, <_2)$ is a section of $(B_1, <_1)$, then we know that $<_2 \subset <_1$ and $c \in B_1 \cap S$. But this contradicts how we chose α.
- If $(B_1, <_1)$ is a section of $(B_2, <_2)$, then we know there exists $\gamma \in B_2$ such that $B_1 = S_\gamma \subseteq B_2$. If $c <_2 \gamma$, then we know that $c \in B_1$ and thus $B_1 \cap S$. This contradicts how we chose α. So we must have that $\gamma <_2 c$. But then this also gives us a contradiction as $\alpha <_2 \gamma <_2 c \Longrightarrow \alpha <_2 c$, meaning that $\alpha <' c$.

(c) [Not in the book...] Given any $\mathcal{B}$ from part (b) of this problem and defining $(B', <')$ as before, we have that $(B, <) \preceq (B', <')$ for all $(B, <) \in \mathcal{B}$.

Consider any $(B_1, <_1) \in \mathcal{B}$. If $B_1 \neq B'$, then we know there exists $\alpha \in B' - B_1$, thus meaning there exists $(B_2, <_2) \in \mathcal{B}$ such that $\alpha \in B_2$. Since $B_2 \not\prec B_1$, we know that $B_1 \prec B_2$, meaning that $B_1 = S_\beta \subseteq B_2$ for some $\beta \in B_2$.

Now we know that $\{b \in B' \mid b <' \beta\} \subseteq \{b \in B_2 \mid b <_2 \beta\}$. For suppose there exists a in the former set but not the latter set. Then there must exist $(B_3, <_3) \in \mathcal{B}$ such that $(a, b) \in <_3$.

If $a \in B_2$, then we'd have that $(b, a) \in <_2$. But that would imply that (a, b) and (b, a) are in $<'$ which we know isn't possible. So we know that $B_3 \not\subseteq B_2$.

Since $\mathcal{B}$ is simply ordered by $\prec$ and we can't have that $B_3 \prec B_2$, we know that $B_2 \prec B_3$. So $B_2 = S_\gamma$ where $\gamma \in S_3$. Now $a <_3 \gamma$ would contradict that $a \notin B_2$. So we must have that $\gamma <_3 a$. However, we also must have that $b <_3 \gamma$, which contradicts that $a <_3 b$.

Hence, we've shown that $(B_1, <_1) \neq (B', <')$ implies that $(B_1, <_1) \prec (B', <')$.

Exercise 6: Use exercise 5 to prove that the maximum principle implies the well-ordering theorem.

Let X be any set and construct $\mathcal{A}$ and $\prec$ as before in exercise 5. By the maximal principle, we know there exists $\mathcal{B} \subseteq \mathcal{A}$ such that $\mathcal{B}$ is simply ordered by $\prec$ and no proper superset of $\mathcal{B}$ is simply ordered by $\prec$.

Next, construct B' and $<'$ as in exercise 5.b. We claim that $B' = X$. To see this, suppose there exists $c \in X - B'$. Then $B' \cup \{c\}$ is well-ordered by the order relation: $<' \cup \{(b, c) \mid b \in B'\}$. Hence $(B' \cup \{c\}, <' \cup \{(b, c) \mid b \in B'\}) \in \mathcal{A}$.

At the same time, note that $(B', <') \prec (B' \cup \{c\}, <' \cup \{(b, c) \mid b \in B'\})$. And since we have that $(B, <) \preceq (B', <')$ for all $(B, <) \in \mathcal{B}$, we thus know that for any $(B, <) \in \mathcal{B}$:

$$(B, <) \prec (B' \cup \{c\}, <' \cup \{(b, c) \mid b \in B'\}).$$

This tells us both that $(B' \cup \{c\}, <' \cup \{(b, c) \mid b \in B'\}) \notin \mathcal{B}$ and that $(B' \cup \{c\}, <' \cup \{(b, c) \mid b \in B'\})$ is comparable with all elements of $\mathcal{B}$. But that contradicts that $\mathcal{B}$ is a maximal simply ordered subset of $\mathcal{A}$.

So we must have that $B' = X$. And thus by exercise 5.b, we know that a well-ordering of X exists.

9/24/2024

Exercise 7: Use exercises 1-5 to prove that the choice axiom implies the well-ordering theorem.

Let X be a set and c be a fixed choice function for the nonempty subsets of X. If T is a subset of X and $<$ is a relation on T, we say that $(T, <)$ is a tower in X if $<$ is a well-ordering of T and if for each $x \in T$, $x = c(X - S_x(T))$ where $S_x(T)$ is the section of T by x.

Well, shit. I wish I was given that notation for specifying which set I was taking a section of before I did exercise 2. $h(S_x(J)) = S_{h(x)}(E)$ is a lot clearer notation than just $h(S_x) = S_{h(x)}$

(a) Let $(T_1, <_1)$ and $(T_2, <_2)$ be two towers in X. Show that either these two ordered sets are the same or one equals a section of the other.

By applying exercise 4 and switching indices if necessary, we know that T_1 has the order type of one of either T_2 or one section of T_2. In other words, there exists an order preserving map $h : T_1 \longrightarrow T_2$ such that $h(T_1)$ equals T_2 or a section of T_2.

Now we assert that given any $x \in T_1$, $h(x) = x$. To prove this, first note that because of transfinite induction, we can assume that $h(x) = x$ for all x in $S_x(T_1)$. This means that we can assume $h(S_x(T_1)) = S_x(T_1)$. Also, as part of doing exercise 2, we proved that h must satisfy that $h(S_x(T_1)) = S_{h(x)}(T_2)$. Hence, $S_x(T_1) = S_{h(x)}(T_2)$. This let's us conclude that:

$$x = c(X - S_x(T_1)) = c(X - S_{h(x)}(T_2)) = h(x).$$

With that we now know that $h(T_1) = T_1$. So T_1 equals either T_2 or a section of T_2.

(b) If $(T, <)$ is a tower in X and $T \neq X$, then there is a tower in X of which $(T, <)$ is a section.

Since $T \neq X$, let $y = c(X - T)$. Then define $T' = T \cup \{y\}$ and $<' = < \cup \{(x, y) \mid x \in T\}$. Clearly, $(T', <')$ is a tower which contains $(T, <)$ as a section.

Clearly T' is well-ordered by $<'$.

Also, if $x \in T' - \{y\}$, then we have that $c(X - S_x(T')) = c(X - S_x(T)) = x$. Plus, we know that $c(X - S_y(T')) = c(X - T) = y$.

(c) Let $\{(T_k, <_k) \mid k \in K\}$ be the collection of all towers in X. Then define:

$$T = \bigcup_{k \in K} T_k \text{ and } < = \bigcup_{k \in K} <_k.$$

Show that $(T, <)$ is a tower in X. Conclude that $T = X$.

If we define $\mathcal{A}$ and $\prec$ from X as we did in exercise 5, we can see from part (a) of this problem that $\{(T_k, <_k) \mid k \in K\}$ is a subset of $\mathcal{A}$ that is simply ordered by $\prec$. Thus, from part (b) of exercise 5, we know that T is well-ordered by $<$.

To prove that T is a tower, consider any $y \in T$. Then we know there exists $k \in K$ such that $y \in T_k$. Furthermore, we know that $y = c(X - S_y(T_k))$. By, part (c) of exercise 5, we know that T_k is either a section of T or all of T. Hence, $S_y(T) = S_y(T_k)$. And thus we have that $y = c(X - S_y(T))$.

Now that we have shown $(T, <)$ is a tower in X, we get an easy contradiction if $T \neq X$. This is because T must contain all towers, but T not equalling X would imply the existence of a tower not contained by T due to part (b) of this exercise.

And since $T = X$, we thus have that $<$ is a well-ordering of X. ■

I'm gonna skip doing exercise 8 of the supplementary exercise. Basically it shows that you can construct a well-ordered set with higher cardinality than an arbitrary well-ordered set, all without using the axiom of choice. Also, while that does mean we can construct a minimal uncountable well-ordered set without using the axiom of choice, theorem 10.3 requires the axiom of choice to prove. So almost nothing we discovered about a minimal uncountable well-ordered set can be proven without the axiom of choice.

9/25/2024

I'm gonna try to cram as much topology as I can today before class starts tomorrow. After all, I suspect and fear that a bunch of this will be necessary at some point in 240. As before, I'm shamelessly ripping off James Munkres' book.

A Topology on a set X is a collection $\mathcal{T}$ of subsets of X having the properties:

1. $\emptyset$ and X are in $\mathcal{T}$.
2. The union of the elements of any subcollection of $\mathcal{T}$ is in $\mathcal{T}$.
3. The intersection of the elements of any finite subcollection of $\mathcal{T}$ is in $\mathcal{T}$.

Technically, a topological space is an ordered pair $(X, \mathcal{T})$ consisting of a set X and a topology $\mathcal{T}$ on X. But when no confusion will arise, we usually omit mentioning $\mathcal{T}$ and just call X a topological space.

Given a topological space $(X, \mathcal{T})$, we say that a subset U of X is an open set if $U \in \mathcal{T}$.

Suppose $\mathcal{T}$ and $\mathcal{T}'$ are topologies on X Such that $\mathcal{T} \subseteq \mathcal{T}'$. Then we say $\mathcal{T}'$ is finer or larger than $\mathcal{T}$ Also, we say $\mathcal{T}$ is coarser or smaller than $\mathcal{T}'$. And we say both are comparable with each other.

If $\mathcal{T}$ is properly contained by $\mathcal{T}'$, then we add the word *strictly* before those adjectives.

If X is a set, a basis for a topology on X is a collection $\mathcal{B}$ of subsets of X (called basis elements) such that:

1. For each $x \in X$, there is at least one basis element B containing x.
2. If x belongs to the intersection of two basis elements B_1 and B_2, then there is a basis element B_3 containing x such that $B_3 \subseteq B_1 \cap B_2$.

If $\mathcal{B}$ satisfies these two conditions, then we define the topology $\mathcal{T}$ generated by $\mathcal{B}$ as follows:

$U \subseteq X$ is open if for each $x \in U$, there is a basis element $B \in \mathcal{B}$ such that $x \in B$ and $B \subseteq U$.

Proof that the $\mathcal{T}$ generated by $\mathcal{B}$ is a topology:

We fairly trivially have that $\emptyset$ and X are included in $\mathcal{T}$.

Let $\{U_\alpha\}_{\alpha \in J}$ be an indexed family of elements of $\mathcal{T}$ and define $U = \bigcup\limits_{\alpha \in J} U_\alpha$.
Given any $x \in U$, we know there exists $\alpha \in J$ such that $x \in U_\alpha$.
And since U_α is open, there exists $B \in \mathcal{B}$ such that $x \in B$ and $B \subseteq U_\alpha \subseteq U$. So, we conclude that U is also open.

Finally, we shall prove by induction that given $U_1, \ldots U_n \in \mathcal{T}$, we have that $U_1 \cap \ldots \cap U_n \in \mathcal{T}$.

Firstly, consider any $U_1, U_2 \in \mathcal{T}$. Then, given any $x \in U_1 \cap U_2$, choose basis elements $B_1, B_2 \in \mathcal{B}$ such that $x \in B_1 \subseteq U_1$ and $x \in B_2 \subseteq U_2$.
Since $x \in B_1 \cap B_2$, we know there is a basis element $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$. Then $x \in B_3 \subseteq U$.

With that, we've now shown that the intersection of any two elements of $\mathcal{T}$ is also in $\mathcal{T}$. So, we can proceed by induction.

Suppose for $i < n$ that $(U_1 \cap \ldots \cap U_i) \in \mathcal{T}$. Then we know that $(U_1 \cap \ldots \cap U_i) \cap U_{i+1} \in \mathcal{T}$.

Lemma 13.1: Let X be a set and $\mathcal{B}$ be a basis for a topology $\mathcal{T}$ on X. Then $\mathcal{T}$ equals the collection of all unions of elements of $\mathcal{B}$.

Proof:
Let $\mathcal{T}'$ be the collection of all unions of elements of $\mathcal{B}$.

Since every $B \in \mathcal{B}$ is an element of $\mathcal{T}$, we trivially have that $\mathcal{T}' \subseteq \mathcal{T}$. Meanwhile, given any $U \in \mathcal{T}$, choose for each $x \in U$ an element B_x of $\mathcal{B}$ such that $x \in B_x \subseteq U$. Then $U = \bigcup_{x \in U} B_x$, meaning $U \in \mathcal{T}'$.

(Axiom of Choice usage alert!!)

Lemma 13.2: Let X be a topological space. Suppose that $\mathcal{C}$ is a collection of open sets of X such that for each open set U of X and each $x \in U$, there is an element $C \in \mathcal{C}$ such that $x \in C \subseteq U$. Then $\mathcal{C}$ is a basis for the topology of X.

Proof:
Firstly, we need to show that $\mathcal{C}$ is a basis.

Since X is an open set, we know by hypothesis that for all $x \in X$, there is $C \in \mathcal{C}$ such that $x \in C$. As for the second condition of a basis, suppose $x \in C_1 \cap C_2$ where $C_1, C_2 \in \mathcal{C}$. Since C_1 and C_2 are open, we know that $C_1 \cap C_2$ is open. So by hypothesis, there is $C_3 \in \mathcal{C}$ such that $x \in C_3 \subseteq (C_1 \cap C_2)$.

Secondly, we need to show that $\mathcal{C}$ is a basis for the topology of X.

Let $\mathcal{T}$ be the collection of open sets of X, and let $\mathcal{T}'$ be the topology generated by $\mathcal{C}$. Firstly, if $U \in \mathcal{T}$ and $x \in U$, there is by hypothesis $C \in \mathcal{C}$ such that $x \in C$ and $C \subseteq U$. So $U \subseteq \mathcal{T}'$. Meanwhile, if $W \in \mathcal{T}'$, then W equals a union of elements of $\mathcal{C}$ by lemma 13.1. Since each element of $\mathcal{C}$ is in $\mathcal{T}$, we know W is the union of elements of $\mathcal{T}$, meaning $W \in \mathcal{T}$. So, we've shown that $\mathcal{T} \subseteq \mathcal{T}' \subseteq \mathcal{T}$.

Lemma 13.3: Let $\mathcal{B}$ and $\mathcal{B}'$ be bases for the topologies $\mathcal{T}$ and $\mathcal{T}'$ respectively on X. Then $\mathcal{T}'$ is finer than $\mathcal{T}$ if and only if for each $x \in X$ and each basis element $B \in \mathcal{B}$ containing x, there is a basis element $B' \in \mathcal{B}'$ such that $x \in B' \subseteq B$.

Proof:
($\Longrightarrow$) Let $x \in X$ and $B \in \mathcal{B}$ such that $x \in B$. Since $B \in \mathcal{T}$ and we are assuming $\mathcal{T} \subseteq \mathcal{T}'$, we know that $B \in \mathcal{T}'$. Then since $\mathcal{B}'$ generated $\mathcal{T}'$, we know there is $B' \in \mathcal{B}'$ such that $x \in B' \subseteq B$.

($\Longleftarrow$)
Given an element U of $\mathcal{T}$, we need to show that $U \in \mathcal{T}'$. To do this, consider any $x \in U$. Since $\mathcal{B}$ generates $\mathcal{T}$, there is an element $B \in \mathcal{B}$ such that $x \in B \subseteq U$. Now by hypothesis, there exists $B' \in \mathcal{B}'$ such that $x \in B' \subseteq B$. So $x \in B' \subseteq U$. Hence, $U \in \mathcal{T}'$.

If $\mathcal{B}$ is the collection of all open intervals $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$ in the real line, then we call the topology generated by $\mathcal{B}$ the standard topology on the real line.

We assume $\mathbb{R}$ has this topology unless stated otherwise.

If $\mathcal{B}'$ is the collection of all intervals $[a, b)$ of the real line, we call the topology generated by $\mathcal{B}'$ the lower limit topology.

When $\mathbb{R}$ has this topology, we denote it $\mathbb{R}_l$.

Letting $K = \{\frac{1}{n} \mid n \in \mathbb{Z}_+\}$, if $\mathcal{B}''$ is the collection of all intervals (a, b) of the real line along with all sets of the form $(a, b) - K$, then we call the topology generated by $\mathcal{B}''$ the K-topology on the real line.

When $\mathbb{R}$ has this topology, we denote it $\mathbb{R}_K$.

Lemma 13.4: The topologies of $\mathbb{R}_l$ and $\mathbb{R}_K$ are strictly finer than the standard topology on $\mathbb{R}$. But, they aren't comparable with one another.

Proof:
Let $\mathcal{T}, \mathcal{T}', \mathcal{T}''$ be the topologies of $\mathbb{R}, \mathbb{R}_l, \mathbb{R}_K$ respectively.

Given any $(a, b) \in \mathcal{B}$ and $x \in (a, b)$, we know that $[x, b) \in \mathcal{B}'$ and that $x \in [x, b) \subseteq (a, b)$. So by lemma 13.3, $\mathcal{T} \subseteq \mathcal{T}'$. On the other hand, for any $[x, b) \in \mathcal{B}'$, there is no set $(a, b) \in \mathbb{B}$ such that $x \in (a, b) \subseteq [x, b)$. So $\mathcal{T}' \not\subseteq \mathcal{T}$. Hence, $\mathcal{T}'$ is strictly finer than $\mathcal{T}$.

Also, given any $(a, b) \in \mathcal{B}$, we also know that $(a, b) \in \mathcal{B}''$. So $\mathcal{T} \subseteq \mathcal{T}''$. On the other hand, given $(-1, 1) - K \in \mathcal{B}''$, we know there is no interval $(a, b) \in \mathcal{B}$ such that $0 \in (a, b) \subseteq (-1, 1) - K$. So by lemma 13.3, we know that $\mathcal{T}'' \not\subseteq \mathcal{T}$. Hence, $\mathcal{T}''$ is strictly finer than $\mathcal{T}$.

Finally, we show $\mathcal{T}'$ and $\mathcal{T}''$ aren't comparable. Firstly, given the set $(-1, 1) - K$ in $\mathcal{B}''$, there is no set $[a, b) \in \mathcal{B}'$ such that $0 \in [a, b) \subseteq (-1, 1) - K$. After all, for any $b > 0$, we can use the archimedean property to find $\frac{1}{n} < b$. Secondly, given the set $[0, 1) \in \mathcal{B}'$, no set of the form (a, b) can satisfy that $0 \in (a, b) \subseteq [0, 1)$. Similarly, no set of the form $(a, b) - K$ can satisfy that $0 \in (a - b) - K \subseteq [0, 1)$. So neither $\mathcal{T}' \subseteq \mathcal{T}''$ nor $\mathcal{T}'' \subseteq \mathcal{T}'$.

A subbasis $\mathcal{S}$ for a topology on X is a collection of subsets of X whose union equals X. The topology generated by the subbasis $\mathcal{S}$ is defined to be the collection $\mathcal{T}$ of all unions of finite intersections of elements of $\mathcal{S}$.

Proof that the $\mathcal{T}$ generated by $\mathcal{S}$ is a topology:

It suffices to show that the collection $\mathcal{B}$ of all finite intersections of elements of $\mathcal{S}$ is a basis. The first condition of a basis is trivially true for $\mathcal{B}$ since the union of the elements of $\mathcal{S}$ is all of X and $\mathcal{S} \subseteq \mathcal{B}$.

As for the second condition of a basis, given any $(S_1 \cap \ldots \cap S_n), (S_1' \cap \ldots \cap S_m') \in \mathcal{B}$, we know that $(S_1 \cap \ldots \cap S_n) \cap (S_1' \cap \ldots \cap S_m')$ is a finite intersection of elements of $\mathcal{S}$ and thus an element in $\mathcal{B}$. Thus, the condition easily follows.

9/26/2024

Well, it looks like I'll be able to survive 240A with the topology information I've learned so far. However, it doesn't look like I'll be able to survive 240B with what I know right now. So, I've got to study more of this. But if needed for 188 this quarter, I'll take a break to study algebra.

Exercise 13.3 Show that $\mathcal{T} = \{U \subseteq X \mid X - U \text{ is countable or all of } X\}$ is a topology on X.

Clearly $\emptyset, X \in \mathcal{T}$ since $|X - X| = 0$ and $X - \emptyset = X$.

Suppose $\{U_\alpha\}_{\alpha \in A}$ is a collection of sets in $\mathcal{T}$. Then $X - \bigcup_{\alpha \in A} U_\alpha = \bigcap_{\alpha \in A} (X - U_\alpha)$ is countable since it's a subset of a countable set.

Hence, $\bigcup_{\alpha \in A} U_\alpha \in \mathcal{T}$.

Finally, consider any $\{U_1, \ldots, U_n\}$ in $\mathcal{T}$. Then $X - \bigcap_{k=1}^{n} U_k = \bigcup_{k=1}^{n} (X - U_k)$ is countable since it's a union of finitely many countable sets.

Hence, $\bigcap_{k=1}^{n} U_k \in \mathcal{T}$.

Exercise 13.4:

(a) If $\{\mathcal{T}_\alpha\}_{\alpha \in A}$ is a family of topologies on X, show that $\bigcap \mathcal{T}_\alpha$ is a topology on X. Is $\bigcup \mathcal{T}_\alpha$ a topology on X?

Let $\mathcal{T} = \bigcap_{\alpha \in A} \mathcal{T}_\alpha$.

Since $\emptyset$ and X belong to all $\mathcal{T}_\alpha$, we know that $\emptyset, X \in \mathcal{T}$.

Next, suppose $\{U_\beta\}_{\beta \in B}$ is a collection of sets in $\mathcal{T}$. Since $\{U_\beta\}_{\beta \in B} \subseteq \mathcal{T}_\alpha$ for all α, we know that $\bigcup_{\beta \in B} U_\beta \in \mathcal{T}_\alpha$ for all α. Hence, $\bigcup_{\beta \in B} U_\beta \in \bigcap_{\alpha \in A} \mathcal{T}_\alpha = \mathcal{T}$.

The same argument as used for arbitrary unions also shows that any finite intersection of sets in $\mathcal{T}$ is also in $\mathcal{T}$.

We've now shown that $\mathcal{T}$ is a topology. As for the other question asked, no we don't necessarily have that $\bigcup_{\alpha \in A} \mathcal{T}_\alpha$ is a topology.

To see this, consider the set $X = \{a, b, c\}$ with the topologies $\mathcal{T}_1 = \{\emptyset, \{a\}, \{a, b\}, \{a, b, c\}\}$ and $\mathcal{T}_2 = \{\emptyset, \{c\}, \{b, c\}, \{a, b, c\}\}$. Then $\mathcal{T}_1 \cup \mathcal{T}_2$ is not a topology because $\{a\}, \{c\} \in \mathcal{T}_1 \cup \mathcal{T}_2$ but $\{a, c\} \notin \mathcal{T}_1 \cup \mathcal{T}_2$.

(b) Let $\{\mathcal{T}_\alpha\}_{\alpha \in A}$ be a family of topologies on X. Show that there is a unique smallest topology on X containing all the collections $\mathcal{T}_\alpha$, and a unique largest topology contained in all $\mathcal{T}_\alpha$.

Firstly, let $\{\mathcal{T}'_\beta\}_{\beta \in B}$ be the collection of all topologies on X which contain $\bigcup_{\alpha \in A} \mathcal{T}_\alpha$.

We know that $\{\mathcal{T}'_\beta\}_{\beta \in B}$ is not empty because it must at least have $\mathcal{P}(X)$ as an element. Hence, we can apply part (a) of the problem to know that $\bigcap_{\beta \in B} \mathcal{T}'_\beta$ is a topology on X.

Importantly, by virtue of being an intersection, that topology is smaller than all other topologies containing $\bigcup_{\alpha \in A} \mathcal{T}_\alpha$. At the same time, we know it contains $\bigcup_{\alpha \in A} \mathcal{T}_\alpha$.

So it is the unique smallest topology on X containing all the collections $\mathcal{T}_\alpha$.

The second part of this question is trivial from part (a). If a topology $\mathcal{T}''$ is contained in all $\mathcal{T}_\alpha$, then we know that $\mathcal{T}'' \subseteq \bigcap_{\alpha \in A} \mathcal{T}_\alpha$. Clearly, the largest topology satisfying this is $\bigcap_{\alpha \in A} \mathcal{T}_\alpha$.

Exercise 13.5: Show that if $\mathcal{A}$ is a basis for a topology on X, then the topology generated by $\mathcal{A}$ equals the intersection of all topologies on X that contain $\mathcal{A}$.

Let $\mathcal{T}$ be the topology generated by $\mathcal{A}$, and suppose $\mathcal{T}'$ is any topology containing $\mathcal{A}$. Then consider any $U \in \mathcal{T}$. By Lemma 13.1, we know that $U = \bigcup_{\beta \in B} A_\beta$ where $\{A_\beta\}$ is some collection of sets in $\mathcal{A}$. Hence, U is a union of sets in $\mathcal{T}'$, meaning $U \in \mathcal{T}'$.

Since $\mathcal{T} \subseteq \mathcal{T}'$ for all $\mathcal{T}'$ containing $\mathcal{A}$, we thus know that $\mathcal{T}$ is the unique smallest topology containing $\mathcal{A}$. At the same time, by exercise 13.4.a, we know that the intersection $\mathcal{T}''$ of all topologies containing $\mathcal{A}$ is a topology. By virtue of being an intersection, we know it is smaller than all topologies containing $\mathcal{A}$, and that it contains $\mathcal{A}$. So, $\mathcal{T} \subseteq \mathcal{T}'' \subseteq \mathcal{T} \implies \mathcal{T} = \mathcal{T}''$.

Prove the same if $\mathcal{A}$ is a subbasis.

Let $\mathcal{T}$ be the topology generated by $\mathcal{A}$ and suppose $\mathcal{T}'$ is any topology containing $\mathcal{A}$. Then consider any $U \in \mathcal{T}$. We know that $U = \bigcup_{\beta \in B} U_\beta$ where $\{U_\beta\}_{\beta \in B}$ is a collection of finite intersections of sets in $\mathcal{A}$.

Because each U_β must be in $\mathcal{T}'$, we thus know that $U \in \mathcal{T}'$. So $\mathcal{T} \subseteq \mathcal{T}'$.

The rest of the proof goes exactly the same as before.

This fact can be used as a shortcut for finding the unique smallest topology containing all topologies in a collection.

Exercise 13.1: Let X be a topological space and let A be a subset of X. Suppose that for each $x \in A$, there is an open set U such that $x \in U \subseteq A$. Then A is open in X.

For all $x \in A$, pick an open set U_x such that $x \in U_x \subseteq A$. Then $A = \bigcup_{x \in A} U_x$ is a union of open sets.

(A.O.C. usage!!)

9/27/2024

If X is simply-ordered, the standard topology for X (called the order topology) is defined as follows:

Given any $a, b \in X$ with $a < b$, we define the sets (a, b), $[a, b)$, $(a, b]$ and $[a, b]$ as you would expect. These are the open, closed, and half-open intervals on X.

Now let $\mathcal{B}$ be the collection of all sets of the form:

- Open intervals (a, b) in X.
- Intervals of the form $[a_0, b)$ where a_0 is the smallest element of X (if one exists).
- Intervals of the form $(a, b_0]$ where b_0 is the largest element of X (if one exists).

The collection $\mathcal{B}$ is a basis for a topology on X which is called the order topology.

It's fairly trivial to show that this is a basis. It's just that for the second condition of a basis, there are a bunch of cases that need to be mentioned.

Another way we can define the order topology is through rays. Given any $a \in X$, we define the sets $(a, +\infty)$, $(-\infty, a)$, $[a, +\infty)$ and $(-\infty, a]$ as you would expect.

Let $\mathcal{S}$ be the set of open rays: $(a, +\infty)$ and $(-\infty, a)$. This is a subbasis for the order topology on X.

> To see this, firstly note that all open rays are open sets in the order topology of X. So, every set in the topology generated by $\mathcal{S}$ will be an open set in our original order topology. Hence if $\mathcal{T}$ is the order topology on X and $\mathcal{T}'$ is the topology generated by S, we know that $\mathcal{T}' \subseteq \mathcal{T}$.
>
> At the same time, every interval in the previously defined basis of $\mathcal{T}$ is the intersection of two (or one if the interval contains the greatest or least element of X) rays. Hence, $\mathcal{T} \subseteq \mathcal{T}'$.

9/29/2024

Today I'm going to be studying from Michael Artin's textbook *Algebra, second edition*. My reasoning is that I need to learn more group theory in order to be ready for 188.

A law of compositon on a set S is map from $S \times S$ to S. Given the ordered pair $(a, b) \in S \times S$, we denote the element the pair is mapped to as either ab, $a \times b$, $a \circ b$, $a + b$, or etc.

> Typically, $+$ is used if the composition is commutative. Meanwhile, the multiplicative symbols don't imply commutativity.

Proposition 2.1.4: Let an associative law of composition be given on S. Then we can uniquely define for all $n \in \mathbb{N}$ a product of n elements $a_1, \ldots, a_n$ of S, temporarily denoted by $[a_1 \cdots a_n]$, with the following properties:

(i) The product $[a_1]$ of one element is a_1.

(ii) The product $[a_1 a_2]$ of two elements is given by the law of composition.

(iii) For any integer i in the range $1 \leq i < n$, $[a_1 \cdots a_n] = [a_1 \cdots a_i][a_{i+1} \cdots a_n]$.

Proof:
We proceed by induction on n.

Let us define $[a_1 \cdots a_n] = [a_1 \cdots a_{n-1}][a_n]$ and suppose that the analogous definition of $[a_1 \cdots a_r]$ satisfies our properties for all $1 < r < n$. Then for any $1 \leq i < n - 1$, we have that:

$$\begin{aligned}
[a_1 \cdots a_n] &= [a_1 \cdots a_{n-1}][a_n] && \text{(by definition)} \\
&= ([a_1 \cdots a_i][a_{i+1} \cdots a_{n-1}])[a_n] && \text{(by inductive hypothesis)} \\
&= [a_1 \cdots a_i]([a_{i+1} \cdots a_{n-1}][a_n]) && \text{(by associativity)} \\
&= [a_1 \cdots a_i][a_{i+1} \cdots a_{n-1} a_n]
\end{aligned}$$

Based on the previous proposition, it's safe to just denote the product of $a_1, \ldots, a_n$ as $a_1 \cdots a_n$.

An identity for a law of composition is an element e of S satisfying that:

$$ea = a \text{ and } ae = a \text{ for all } a \in S.$$

We denote the identity of a law of composition as 0 or 1 (depending on whether we are using multiplicative or additive notation). We can only have one identity element.

Proof:
Suppose e and e' are both identity elements. Then $e = ee' = e'$.

An element a of S is invertible if there is another element $b \in S$ such that $ab = 1$ and $ba = 1$. We call b the inverse of a and denote b as $-a$ or a^{-1} depending on whether additive or multiplicative notation is being used.

Exercise 1.2:

- If an element a has both a left inverse l and a right inverse r, then $l = r$, a is invertible, and r is its inverse.

  Suppose $la = 1$ and $ar = 1$. Then we have that:

  $$r = 1r = (la)r = l(ar) = l1 = l$$

- If a is invertible, its inverse is unique.

  Suppose b and b' are both inverses of a. Then:

  $$b = 1b = (b'a)b = b'(ab) = b'1 = b'$$

- If a and b are invertible, then ab is invertible with $(ab)^{-1} = b^{-1}a^{-1}$.

  Proof:

  $$abb^{-1}a^{-1} = a1a^{-1} = aa^{-1} = 1 \text{ and } b^{-1}a^{-1}ab = b^{-1}1b = b^{-1}b = 1$$

A group is a set G together with a law of composition such that:

1. The law of composition is associative.
2. G has an identity element.
3. Every element of G has an inverse.

An abelian group is a group whose law of composition is commutative.

The order of a group G is the number of elements it contains. We denote the order $|G|$. If $|G|$ is finite, we say G is a finite group. Otherwise, we say G is an infinite group.

The $n \times n$ general linear group is the group of all invertible $n \times n$ matrices. It's denoted GL_n. If we want to specify whether we are working with real or complex matrices, we write $GL_n(\mathbb{R})$ or $GL_n(\mathbb{C})$.

The symmetric group of a set is the set of permutations of the set (with the law of composition being function composition). We denote S_n the group of permutations of the indices $1, 2, \ldots n$.

S_3 has order 6 (this is an easy fact from combinatorics). Now let $x = (1\ \ 2\ \ 3)$ and $y = (1\ \ 2)(3)$ (this is using cyclic notation). Then we can express all the elements of S_3 as products of x and y.

$$\begin{array}{lll} 1 = (1)(2)(3) & x = (1\ \ 2\ \ 3) & x^2 = (1\ \ 3\ \ 2) \\ y = (1\ \ 2)(3) & xy = (1\ \ 3)(2) & x^2y = (1)(2\ \ 3) \end{array}$$

Also note that $x^3 = 1$, $y^2 = 1$, and $yx = x^2y$.

Exercise 2.1: Make a multiplication table for the symmetric group S_3.

Just to clarify, to read this table, take the row element to be on the left side of the product and the column element to be on the right side of the product.

$\times$	1	x	x^2	y	xy	x^2y
1	1	x	x^2	y	xy	x^2y
x	x	x^2	1	xy	x^2y	y
x^2	x^2	1	x	x^2y	y	xy
y	y	x^2y	xy	1	x^2	x
xy	xy	y	x^2y	x	1	x^2
x^2y	x^2y	xy	y	x^2	x	1

12/14/2024

My goal for while I'm flying home and can't work on grading is to go prove the following extra exercise of homework 2 in my math 188 class.

Let $V = \{F(x) \in \mathbb{C}[[x]] : F(0) = 0\}$ and $W = \{F(x) \in \mathbb{C}[[x]] : G(0) = 1\}$.

(a) Given $F(x) \in V$, show that $\boldsymbol{E}(F(x)) = \sum_{n \geq 0} \frac{F(x)^n}{n!}$ is the *unique* formal power series $G(x) \in W$ such that $\mathrm{D}G(x) = \mathrm{D}F(x) \cdot G(x)$. This defines a function $\boldsymbol{E} : V \longrightarrow W$.

Note that we use the convention $F(x)^0 = 1$ even if $F(x) = 0$.

Firstly, note that we have $G(x) \coloneqq \boldsymbol{E}(F(x)) = \exp(F(x))$ where $\exp(x) \coloneqq \sum_{n\geq 0} \frac{x^n}{n!}$. Also, you can check $\exp(x)$ is it's own derivative. Thus by chain rule:

$$G(x) = \mathrm{D}F(x) \cdot \mathrm{D}(\exp)(F(x)) = \mathrm{D}F(x) \cdot \exp(F(x)) = \mathrm{D}F(x) \cdot G(x)$$

Next, suppose $H(x)$ is another formal power series in W satisfying that $\mathrm{D}H(x) = \mathrm{D}F(x) \cdot H(x)$. Note that since $H(0) \neq 0 \neq G(0)$, we can write that $\frac{\mathrm{D}H(x)}{H(x)} = \mathrm{D}F(x) = \frac{\mathrm{D}G(x)}{G(x)}$. Therefore, we get that:

$$(*) \quad \mathrm{D}H(x) \cdot G(x) = \mathrm{D}G(x) = H(x)$$

Let $H(x) = \sum_{n\geq 0} h_n x^n$ and $G(x) = \sum_{n\geq 0} g_n x^n$. Since we assumed that $H(x), G(x) \in W$, we know that $h_0 = g_0 = 1$. Then, proceeding by induction (assuming that $h_i = g_i$ for all $0 \leq i \leq n$), when we take the nth. coefficient of $(*)$ we get:

$$\begin{aligned}(n+1)h_{n+1} + \sum_{i=0}^{n-1}(i+1)h_{i+1}g_{n-i} \\ &= \sum_{i=0}^{n}(i+1)h_{i+1}g_{n-i} = \sum_{i=0}^{n}(i+1)g_{i+1}h_{n-i} \\ &= (n+1)g_{n+1} + \sum_{i=0}^{n-1}(i+1)g_{i+1}h_{n-i}\end{aligned}$$

But by induction we have $\sum_{i=0}^{n-1}(i+1)g_{i+1}h_{n-i} = \sum_{i=1}^{n-1}(i+1)h_{i+1}g_{n-i}$.

So subtracting out the sum from $i = 0$ to $n-1$ and then dividing by $n+1$ which is crucially nonzero, we then have that $h_{n+1} = g_{n+1}$.

(b) Given $G(x) \in W$, show that there is a *unique* formal power series $F(x) \in V$ such that $\mathrm{D}F(x) = \frac{\mathrm{D}G(x)}{G(x)}$. This let's us define the function $\boldsymbol{L} : W \longrightarrow V$ by $\boldsymbol{L}(G(x)) = F(x)$.

Since $G(0) = 1$, we know that $G(x)$ is invertible. So there is a unique formal power series $A(x) = \sum_{n\geq 0} a_n x^n$ such that $A(x) = \frac{\mathrm{D}G(x)}{G(x)}$.

Then if $F(x) = \sum_{n\geq 0} f_n x^n$ satisfies that $\mathrm{D}F(x) = \frac{\mathrm{D}G(x)}{G(x)}$, then we can solve that $f_n = \frac{a_{n-1}}{n}$ for all $n \geq 1$. This shows that f_n is uniquely determined for all $n \geq 1$. Also, since we are forcing $F(0) = 0$, we know that $f_0 = 0$. So $F(x)$ is a unique power series.

From this it's also hopefully clear to see how one can solve for $F(x)$ in order to show that $F(x)$ exists.

(c) Show that $\boldsymbol{E}$ and $\boldsymbol{L}$ are inverses of each other.

Firstly, we'll show $\boldsymbol{L}(\boldsymbol{E}(F(x))) = F(x)$ for all $F(x) \in V$.

Let $F(x) \in V$, $G(x) = \boldsymbol{E}(F(x))$, and $H(x) = \boldsymbol{L}(G(x))$. Then we have that:

$$\mathrm{D}H(x) = \frac{\mathrm{D}G(x)}{G(x)} = \frac{\mathrm{D}F(x)\cdot G(x)}{G(x)} = \mathrm{D}F(x)$$

This proves that $[x^n]H(x) = [x^n]F(x)$ for all $n \geq 1$. And since both $H(x), F(x) \in V$, we know that $H(0) = F(0)$. so $H(x) = F(x)$.

Secondly, we'll show $\boldsymbol{E}(\boldsymbol{L}(F(x))) = F(x)$ for all $F(x) \in W$.

Let $F(x) \in V$, $G(x) = \boldsymbol{L}(F(x))$, and $H(x) = \boldsymbol{E}(G(x))$. Then we have that:

$$\mathrm{D}H(x) = \mathrm{D}G(x) \cdot H(x) = \frac{\mathrm{D}F(x)}{F(x)} \cdot H(x)$$

Thus we know that $\mathrm{D}H(x) \cdot F(x) = \mathrm{D}F(x) \cdot H(x)$. Since both $H(x)$ and $F(x)$ are in W, we can employ identical logic as that of part (a) to show that $H(x) = F(x)$.

(d) Show that $\boldsymbol{E}(F_1(x) + F_2(x)) = \boldsymbol{E}(F_1(x))\boldsymbol{E}(F_2(x))$ for all $F_1(x), F_2(x) \in V$.

Note that:

$$\begin{aligned}
\boldsymbol{E}(F_1(x) + F_2(x)) &= \lim_{m\to\infty} \sum_{n=0}^{m} \frac{(F_1(x)+F_2(x))^n}{n!} \\
&= \lim_{m\to\infty} \sum_{n=0}^{m} \frac{1}{n!} \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} F_1(x)^i F_2(x)^{n-i} \\
&= \lim_{m\to\infty} \sum_{n=0}^{m} \sum_{i=0}^{n} \frac{F_1(x)^i}{i!} \cdot \frac{F_2(x)^{n-i}}{(n-i)!} \\
&= \lim_{m\to\infty} \left((\sum_{n=0}^{m} \frac{F_1(x)^n}{n!})(\sum_{n=0}^{m} \frac{F_2(x)^n}{n!}) + R_m(x) \right)
\end{aligned}$$

In the above manipulations $R_m(x)$ is the formal power series negating all the terms which are in $(\sum_{n=0}^{m} \frac{F_1(x)^n}{n!})(\sum_{n=0}^{m} \frac{F_2(x)^n}{n!})$ but aren't in $\sum_{n=0}^{m} \sum_{i=0}^{n} \frac{F_1(x)^i}{i!} \cdot \frac{F_2(x)^{n-i}}{(n-i)!}$.

In other words, $R_m(x)$ contains all the terms of the form $\frac{1}{i!j!} F_1(x)^i F_2(x)^j$ where $i + j > m$. Importantly, because $F_1(0) = 0 = F_2(0)$, we know that $\mathrm{mdeg}\ R(x) > m$. So, $R_m(x) \to 0$ as $m \to \infty$.

In turn:

$$\lim_{m\to\infty} \left((\sum_{n=0}^{m} \frac{F_1(x)^n}{n!})(\sum_{n=0}^{m} \frac{F_2(x)^n}{n!}) + R_m(x) \right)$$
$$= \lim_{m\to\infty} \sum_{n=0}^{m} \frac{F_1(x)^n}{n!} \cdot \lim_{m\to\infty} \sum_{n=0}^{m} \frac{F_2(x)^n}{n!} + \lim_{m\to\infty} R_m(x) = \boldsymbol{E}(F_1(x))\boldsymbol{E}(F_2(x))$$

(e) Show that $\boldsymbol{L}(F_1(x)F_2(x)) = \boldsymbol{L}(F_1(x)) + \boldsymbol{L}(F_2(x))$ for all $F_1(x), F_2(x) \in W$.

By part (d), we know that:

- $\boldsymbol{E}(\boldsymbol{L}(F_1(x)) + \boldsymbol{L}(F_2(x))) = \boldsymbol{E}(\boldsymbol{L}(F_1(x)))\boldsymbol{E}(\boldsymbol{L}(F_2(x)))$

Meanwhile, by part (c) we know that:

- $\boldsymbol{E}(\boldsymbol{L}(F_1(x)F_2(x))) = F_1(x)F_2(x)$
- $\boldsymbol{E}(\boldsymbol{L}(F_1(x)))\boldsymbol{E}(\boldsymbol{L}(F_2(x))) = F_1(x)F_2(x)$

Hence, we've shown that $\boldsymbol{E}$ maps the left- and right-hand sides of the above claimed equation to the same formal power series. But from part (c) we know that $\boldsymbol{E}$ is an injective map. So we must have that the two sides of the equation are in fact equal.

(f) If m is a positive integer and $G(x) \in W$, show that $\boldsymbol{E}(\frac{\boldsymbol{L}(G(x))}{m})$ is an mth. root of $G(x)$.

Let $F(x) = \boldsymbol{E}(\frac{\boldsymbol{L}(G(x))}{m})$. This implies $m \cdot \boldsymbol{L}(F(x)) = \boldsymbol{L}(G(x))$. Then by part (e), we know that $\boldsymbol{L}(F(x)^m) = \boldsymbol{L}(G(x))$. And finally, plugging both sides into $\boldsymbol{E}$ we get $F(x)^m = G(x)$.

Also, $F(0) = 1$. So $F(x)$ is the unique mth. root of $G(x)$ with 1 as its constant coefficient.

Because of part (f), we can now extend our definition of the zth. power of a formal power series $G(x) \in W$ to any complex number z. Specifically, for any $G(x) \in W$ we define $G(x)^z \coloneqq \boldsymbol{E}(z \cdot \boldsymbol{L}(G(x)))$. Then, we've shown in part (f) that this definition agrees with our more restricted definition from math 188 on at least the positive rational numbers.

Two important identities (where $G(x) \in W$ and $z, w \in \mathbb{C}$):

- $G(x)^z G(x)^w = G(x)^{z+w}$?

  Proof:
  $$\begin{aligned} G(x)^z G(x)^w &= \boldsymbol{E}(z \cdot \boldsymbol{L}(G(x)))\boldsymbol{E}(w \cdot \boldsymbol{L}(G(x))) \\ &= \boldsymbol{E}(z \cdot \boldsymbol{L}(G(x)) + w \cdot \boldsymbol{L}(G(x))) = \boldsymbol{E}((z+w) \cdot \boldsymbol{L}(G(x))) \\ &= G(x)^{z+w} \end{aligned}$$

- $(G(x)^z)^w = G(x)^{zw}$

  Proof:
  $$\begin{aligned} (G(x)^z)^w &= (\boldsymbol{E}(z \cdot \boldsymbol{L}(G(x))))^w \\ &= \boldsymbol{E}(w \cdot \boldsymbol{L}(\boldsymbol{E}(z \cdot \boldsymbol{L}(G(x))))) = \boldsymbol{E}(w \cdot (z \cdot \boldsymbol{L}(G(x)))) \\ &= \boldsymbol{E}(wz \cdot \boldsymbol{L}(G(x))) = G(x)^{zw} \end{aligned}$$

Using those identities, here are some corollaries:

- $G(x)^{-a}$ gives the multiplicative inverse of $G(x)^a$

  (this proves that this definition of exponentiation agrees with our more restricted definition from math 188 on all rational numbers).

  Proof:
  Firstly note that:
  $$G(x)^{-a}G(x)^a = G(x)^{-a+a} = G(x)^0 = \boldsymbol{E}(0 \cdot \boldsymbol{L}(G(x))) = \boldsymbol{E}(0)$$

  Seconly, note that $\boldsymbol{E}(0) = \sum_{n\geq 0} \frac{0^n}{n!} = \frac{0^0}{0!} = 1$ (as a reminder we are using the convention that $0^0 = 1$). Therefore, $G(x)^{-a}G(x)^a = 1$, meaning $G(x)^{-a}$ is the multiplicative inverse of $G(x)^a$.

- V is a subgroup of $\mathbb{C}[[x]]$ under addition, W is a group under multiplication, and $\boldsymbol{E}$ is a group isomorphism between $(V, +)$ and $(W, \cdot)$.

  Proof:
  One can easily see without our prior reasoning that $(V, +)$ is subgroup of $\mathbb{C}[[x]]$ under addition (with 0 as it's identity).

  Meanwhile, by our previous corollary we can see that all $G(x) \in W$ have a multiplicative inverse in W. Specifically since $G(x) = G(x)^1$, we know by the previous corollary that $G(x)$ has the inverse $G(x)^{-1}$ inside W. Combining that with the fact that $1 \in W$ and $G(x)H(x) \in W$ when $G(x), H(x) \in W$, we know now that W is a group under multiplcation.

  Finally, we know $\boldsymbol{E}$ is a group isomorphism because of part (c) of this exercise as well as the fact that $\boldsymbol{E}(0) = 1$.

- Power Rule: Given $G(x) \in W$, if $H(x) = G(x)^z$, then
  $$\mathrm{D}H(x) = z\mathrm{D}G(x)G(x)^{z-1}.$$

  Proof:
  $$\begin{aligned}\mathrm{D}H(x) &= \mathrm{D}(\boldsymbol{E}(z \cdot \boldsymbol{L}(G(x))))(x)\\ &= \mathrm{D}(z \cdot \boldsymbol{L}(G(x)))(x) \cdot \boldsymbol{E}(z \cdot \boldsymbol{L}(G(x)))\\ &= z \cdot \mathrm{D}(\boldsymbol{L}(G(x)))(x) \cdot G(x)^z = z\frac{\mathrm{D}G(x)}{G(x)}G(x)^z\\ &= z\mathrm{D}G(x)G(x)^{-1}G(x)^z\\ &= z\mathrm{D}G(x)G(x)^{z-1}\end{aligned}$$

- Binomial Theorem: Given $z \in \mathbb{C}$, we have that:
  $(1+x)^z = \sum\limits_{n\geq 0} \binom{z}{n} x^n$ where $\binom{z}{n} = \frac{z(z-1)\cdots(z-n+1)}{n!}$ when $n > 0$ and 1 when $n = 0$.

  Proof:
  Note that $[x^n](1+x)^z = \frac{1}{n!}\mathrm{D}^n((1+x)^z)(0)$. Also, by induction using the power rule we can say for $n > 0$ that:

$$\begin{aligned}\mathrm{D}^n((1+x)^z)(x) &= z\mathrm{D}^{n-1}((1+x)^z)(x)\\ &= z(z-1)\mathrm{D}^{n-2}((1+x)^z)\\ &= \cdots = z(z-1)\cdots(z-n+1)(1+x)\end{aligned}$$

Therefore $\mathrm{D}^n((1+x)^z)(0) = z(z-1)\cdots(z-n+1)(1+0)$ and we thus have that for $n > 0$.

$$[x^n](1+x)^z = \tfrac{1}{n!}z(z-1)\cdots(z-n+1) = \tbinom{z}{n}.$$

Meanwhile, if $n = 0$, then $[x^0](1+x)^z = 1 = \binom{z}{n}$ (because $(1+x)^z \in W$).

Before going on to parts (g) and (h), here are two more identities (where $F(x) \in V$, $G(x) \in W$, and $z \in \mathbb{C}$):

- $(\boldsymbol{E}(F(x)))^z = \boldsymbol{E}(z \cdot \boldsymbol{L}(\boldsymbol{E}(F(x)))) = \boldsymbol{E}(zF(x))$
- $\boldsymbol{L}(G(x)^z) = \boldsymbol{L}(\boldsymbol{E}(z \cdot \boldsymbol{L}(G(x)))) = z\boldsymbol{L}(G(x))$

(g) Show that if $\sum_{i\geq 0} F_i(x)$ converges to $F(x)$, then $\prod_{i\geq 0} \boldsymbol{E}(F_i(x))$ converges to $\boldsymbol{E}(F(x))$.

We start by proving the following lemma: If $A(x) \in \mathbb{C}[[x]]$ and $(B_i(x))_{i\in\mathbb{N}}$ is a sequence in $\mathbb{C}[[x]]$ converging to $B(x)$ as $i \to \infty$ and satisfying that $B_i(0) = 0$ for all i, then $A(B_i(x)) \to A(B(x))$ as $i \to \infty$.

Proof:
For notation, we'll denote:

$A(x) = \sum_{n\geq 0} a_n x^n$, $\;B_i(x) = \sum_{n\geq 0} b_n^{(i)} x^n$, and $B(x) = \sum_{n\geq 0} b_n x^n$

To start, note that for all integers $m \geq 0$, we have that $B_i(x)^m \to B(x)^m$. Also, since $B_i(0) = 0$ for all i, we know that $\operatorname{mdeg} B_i(x)^m \geq m$ for all integers i and m, and also that $\operatorname{mdeg} B(x)^m \geq m$ for all integers m. Thus, fixing $n \geq 0$ we can say that:

$$[x^n]A(B_i(x)) = [x^n]\sum_{m=0}^{n} a_m B_i(x)^m \text{ and } [x^n]A(B(x)) = [x^n]\sum_{m=0}^{n} a_m B(x)^m$$

Next, let I_m be large enough that $[x^n]B_i(x)^m = [x^n]B(x)^m$ for all $i \geq I_m$. Then set $I = \max(I_0, I_1, \ldots, I_m)$ and note that for all $i \geq I$, we have:

$$\begin{aligned}[x^n]\sum_{m=0}^{n} a_m B_i(x)^m &= \sum_{m=0}^{n} a_m [x^n](B_i(x)^m)\\ &= \sum_{m=0}^{n} a_m [x^n](B(x)^m) = [x^n]\sum_{m=0}^{n} a_m B(x)^m\end{aligned}$$

So for all $i \geq I$, we have that $[x^n]A(B_i(x)) = [x^n]A(B(x))$. This proves $A(B_i(x)) \to A(B(x))$.

I should have proved this in my math 188 notes when I was showing that $((A+B)\circ C)(x) = (A\circ C)(x) + (B\circ C)(x)$ and $(AB\circ C)(x) = (A\circ C)(x)(B\circ C)(x)$. But in my defense the professor didn't mention any of these three facts in his notes.

As a reminder: $\boldsymbol{E}(F(x)) = \exp(F(x))$ where $\exp = \sum_{n\geq 0} \frac{1}{n!}x^n$. Also, by our previous lemma, we know that:

$$\exp(F(x)) = \exp(\lim_{n\to\infty} \sum_{i=0}^{n} F_i(x)) = \lim_{n\to\infty} \exp(\sum_{i=0}^{n} F_i(x))$$

But then $\exp(\sum_{i=0}^{n} F_i(x)) = \boldsymbol{E}(\sum_{i=0}^{n} F_i(x)) = \prod_{i=0}^{n} \boldsymbol{E}(F_i(x))$.

So, we have shown that $\boldsymbol{E}(F(x)) = \lim_{n\to\infty} \prod_{i=0}^{n} \boldsymbol{E}(F_i(x)) = \prod_{i\geq 0} \boldsymbol{E}(F_i(x))$

Side note: The lemma we proved in this part also tells us that if $(B_i(x))_{i\in\mathbb{N}}$ is a sequence in V converging to $B(x)$, then $\boldsymbol{E}(B_i(x)) \to \boldsymbol{E}(B(x))$ as $i \to \infty$. In other words, $\boldsymbol{E}$ is a continuous map.

(If $\rho(A(x), B(x)) = \frac{1}{\text{mdeg }(A-B)(x)}$, then $(\mathbb{C}[[x]], \rho)$ is a metric space in which formal power series convergence is equivalent to convergence in this metric space.)

(h) Show that if $\prod_{i\geq 0} G_i(x)$ converges to $G(x)$, then $\sum_{i\geq 0} \boldsymbol{L}(G_i(x))$ converges to $\boldsymbol{L}(G(x))$.

Unfortunately, unlike with $\boldsymbol{E}$ we do not (currently) have a formal power series $A(x)$ for which we can generally say $\boldsymbol{L}(B(x)) = A(B(x))$. Thus, we can't move the limit from inside $\boldsymbol{L}$ to outside $\boldsymbol{L}$ as easily as we did in part (g) for $\boldsymbol{E}$.

However, consider that $\lim_{n\to\infty} \boldsymbol{E}(\sum_{i=0}^{n} \boldsymbol{L}(G_i(x)))$ exists. Specifically:

$$\begin{aligned}\lim_{n\to\infty} \boldsymbol{E}(\sum_{i=0}^{n} \boldsymbol{L}(G_i(x))) &= \lim_{n\to\infty} \boldsymbol{E}(\sum_{i=0}^{n} \boldsymbol{L}(G_i(x))) \\ &= \lim_{n\to\infty} \prod_{i=0}^{n} \boldsymbol{E}(\boldsymbol{L}(G_i(x))) = \lim_{n\to\infty} \prod_{i=0}^{n} G_i(x) = G(x)\end{aligned}$$

Thus, if we can show $\sum_{i\geq 0} \boldsymbol{L}(G_i(x))$ converges, then we can use the lemma from part (g) to see that:

$$\boldsymbol{E}(\lim_{n\to\infty} \sum_{i=0}^{n} \boldsymbol{L}(G_i(x))) = \lim_{n\to\infty} \boldsymbol{E}(\sum_{i=0}^{n} \boldsymbol{L}(G_i(x))) = G(x),$$

and then by applying $\boldsymbol{L}$ to the left and right sides of this equation, we will get our desired result.

We now show that $\sum_{i\geq 0} \boldsymbol{L}(G_i(x))$ converges. Firstly, note that after fixing $n \geq 1$, we have that:

$$[x^n](\boldsymbol{L}(G_i(x)))(x) = \frac{1}{n}[x^{n-1}]\frac{\mathrm{D}G_i(x)}{G_i(x)}.$$

Secondly, note that $\prod_{i\geq 0} G_i(x)$ converging to $G(x)$ and $G_i(0) = 1$ for all i implies that $\mathrm{mdeg}\ (G_i(x) - 1) \to \infty$ as $i \to \infty$ and thus $\mathrm{mdeg}\ \mathrm{D}G_i(x) = \mathrm{mdeg}\ (G_i(x) - 1) - 1 \to \infty$ as $i \to \infty$.

Hence, there exists $I_n \geq 0$ such that $i \geq I_n$ implies that $[x^0]\mathrm{D}G_i(x)$, $[x^1]\mathrm{D}G_i(x), \ldots, [x^{n-1}]\mathrm{D}G_i(x)$ are all 0. In turn, we have for all $i \geq I_n$ that $[x^{n-1}](\mathrm{D}G_i(x) \cdot \frac{1}{G_i(x)}) = 0$ and $[x^n](\boldsymbol{L}(G_i(x)))(x) = \frac{1}{n} \cdot 0 = 0$.

Since we also have by the definition of $\boldsymbol{L}$ that $(\boldsymbol{L}(G_i(x)))(0) = 0$ for all i, we can thus conclude that: $\lim\limits_{i\to\infty} \mathrm{mdeg}\ (\boldsymbol{L}(G_i(x))) = 0$.

This proves that $\sum_{i\geq 0} \boldsymbol{L}(G_i(x))$ converges.

To finish off, here is an identity: If $z, \alpha \in \mathbb{C}$ and $G(x) = (1 - \alpha x)^z$, then $\boldsymbol{L}(G(x)) = \sum\limits_{n\geq 1} \frac{-z}{n}\alpha^n x^n$.

Proof:
We already showed that $\boldsymbol{L}((1-\alpha x)^z) = z \cdot \boldsymbol{L}(1 - \alpha x)$. Also:

$$\mathrm{D}(\boldsymbol{L}(1-\alpha x))(x) = \frac{\mathrm{D}(1-\alpha x)}{1-\alpha x} = -\alpha \sum_{n\geq 0} \alpha^n x^n = \sum_{n\geq 0} -\alpha^{n+1} x^n$$

Thus, we have that $z \cdot \boldsymbol{L}(1-\alpha x) = z \cdot \sum\limits_{n\geq 1} \frac{-1}{n}\alpha^{(n+1)-1}x^n = \sum\limits_{n\geq 1} \frac{-z}{n}\alpha^n x^n$.

Some thoughts on this:

- $\boldsymbol{L}(\frac{1}{1-x}) = \sum_{n\geq 1} \frac{+1}{n}(1)^n x^n = \sum_{n\geq 1} \frac{1}{n}x^n$ which is importantly the same as what we found in class.
- Given that $G(x)$ is a polynomial with constant coefficient 1, we can combine this identity with part (e) to calculate $\boldsymbol{L}(G(x))$.

  Specifically, let $G(x) = (1 - \frac{1}{\gamma_1}x)\cdots(1 - \frac{1}{\gamma_k}x)$ where $\gamma_1, \ldots, \gamma_k$ are the roots of $G(x)$. Since $G(0) = 1$ by assumption, we know that $\gamma_j \neq 0$ for all j. Then we have that:

$$\begin{aligned}\boldsymbol{L}(G(x)) &= \boldsymbol{L}((1 - \tfrac{1}{\gamma_1}x)\cdots(1 - \tfrac{1}{\gamma_k}x)) \\ &= \boldsymbol{L}(1 - \tfrac{1}{\gamma_1}x) + \ldots + \boldsymbol{L}(1 - \tfrac{1}{\gamma_k}x) \\ &= \sum_{n\geq 1} \tfrac{-1}{n}{\gamma_1}^{-n}x^n + \ldots + \sum_{n\geq 1} \tfrac{-1}{n}{\gamma_k}^{-n}x^n = \sum_{n\geq 1} \tfrac{-1}{n}\left(\sum_{j=1}^{k} {\gamma_j}^{-n}\right)x^n\end{aligned}$$

And now I'm out of ideas of what else to do with this homework problem.

12/19/2024

For the next while I want to work through some of the exercises in Folland's *Real Analysis* about the Cantor set and Cantor function. Assume for this section that $\mathbb{R}$ is equipped with the standard metric ρ and that we are using the complete Lebesgue measure space $(\mathbb{R}, \mathcal{L}, m)$.

If I is a bounded interval and $\alpha \in (0,1)$, then call the open interval with the same midpoint as I and length equal to α times the length of I the "open middle αth" of I. If $(\alpha_j)_{j\in\mathbb{N}}$ is a sequence of numbers in $(0,1)$, then we can define a decreasing sequence $(K_j)_{j\in\mathbb{N}}$ of closed sets by setting $K_0 = [0,1]$ and obtaining K_j by removing the open middle α_jth from the intervals that make up K_{j-1}. Then $K = \bigcap_{j\in\mathbb{N}} K_j$ is called a generalized Cantor set.

The ordinary Cantor set C is obtained by setting all α_j equal to $1/3$.

Exercise 2.27: Let $K = \bigcap K_j$ be a generalized Cantor set created using the sequence $(\alpha_j)_{j\in\mathbb{N}}$ in $(0,1)$. Prove that K is compact, perfect (i.e. closed and has no isolated points), nowhere dense (i.e. not dense in any nonempty open set), and totally disconnected (i.e. the only connected subsets of K are single points).

- K is closed because it is an intersection of closed sets. Also, K is a bounded set in $\mathbb{R}$ because it is a subset of $[0,1]$. Thus, K is compact.

- Let $x \in K$. Then for any $\varepsilon > 0$, pick $J \in \mathbb{N}$ with $2^{-J} < \varepsilon$. Note that all intervals of K_j have a length at most 2^{-j}. After all, when going from K_{j-1} to K_j, we split all the intervals of K_{j-1} in half and then remove an additional amount of length determined by α_j. So, let I be the interval of K_J containing x. Then both endpoints of I are in K and also in $B(\varepsilon, x)$. And, at least one of those endpoints is not x. So x is a limit point of K. Since K is also closed, we have that K is perfect.

- Let $x, y \in K$ and without loss of generality assume $x < y$. Then we know there must exist some integer $J \in \mathbb{N}$ such that x and y are in different intervals of K_J. Afterall, as previously mentioned, points in the same interval of K_j are within 2^{-j} distance of each other. So if no such J exists, then $\rho(x,y) < 2^{-j}$ for all j, meaning $x = y$.

  We can specifically choose J to be the least integer such that x and y are in two different intervals of K_J. Then both x and y are in the same interval I of K_{J-1}, but the midpoint of that interval z is not in $K_J \subseteq K$ and $x < z < y$. By a theorem in 140A, this proves that K is totally disconnected since for all $x, y \in K$, $[x,y] \not\subseteq K$ unless $x = y$.

- Since K is perfect, we know that K is only dense on subsets of K. However, since all open sets in $\mathbb{R}$ are countable unions of open intervals and K contains no nonempty open intervals since K is totally disconnected, we know that K has no nonempty open subsets.

Trying to explicitely write out the bijection between $[0,1]$ and a generalized Cantor set $K = \bigcap K_j$ would be really time consuming and awkward. So I'm going to be more handwavey

We can define an injection from $\{0,1\}^\omega$ to K as follows:

Given $\boldsymbol{x} \in \{0,1\}^\omega$, we can define a convergent subsequence in K. Specifically, set $a_0 = 0$. Then recursively for $j > 0$, we know a_{j-1} falls into some interval I of K_{j-1}. Futhermore, we know that I gets split into two disjoint intervals I_0 and I_1 when going from K_{j-1} to K_j (take I_0 to be the lower interval). Then, let a_j be the left bound on I_n where n is the value at the jth index of $\boldsymbol{x}$.

Since the endpoints of the intervals in each K_j are all in the final intersection, we know that $(a_j)_{j\in\mathbb{N}}$ is a sequence contained in K. Also, $\rho(a_j, a_{j+1}) < 2^{-j}$ for all $j \geq 0$. From that you can easily work out that $(a_j)_{j\in\mathbb{N}}$ is Cauchy. Thus, since K is closed, we know that $(a_j)_{j\in\mathbb{N}}$ converges to some number $y \in K$.

The mapping $\boldsymbol{x} \mapsto y$ is injective because $\boldsymbol{x}$ uniquely determines which interval of K_j that y is in for all j (specifically the same interval as a_j for each j). If $\boldsymbol{x}'$ is another sequence of 0s and 1s mapped to y', and $\boldsymbol{x}$ and $\boldsymbol{x}'$ differ at position J, then y and y' will be in two different intervals of K_J. Since those intervals are disjoint, we know that $y \neq y'$.

It is possible to show that our above injection is also surjective. However, it's quicker to just say $\mathfrak{c} = \mathrm{card}(\{0,1\}^\omega) \leq \mathrm{card}(K) \leq \mathrm{card}(\mathbb{R}) = \mathfrak{c}$. Thus generalized Cantor sets have the cardinality of the continuum.

Finally, note that given a generalized Cantor set $K = \bigcap K_j$, because $m(K_1) < 1$ and $(K_j)_{j\in\mathbb{N}}$ is a decreasing sequence of sets, we know that:

$$m(K) = \lim_{j\to\infty} m(K_j) = \lim_{j\to\infty} 2^j \prod_{i=1}^{j} \tfrac{(1-\alpha_i)}{2} = \prod_{j=1}^{\infty}(1-a_j)$$

Exercise 1.32:

(a) Suppose $(a_j)_{j\in\mathbb{N}}$ is a sequence in $(0,1)$. $\prod_{j=1}^\infty (1-a_j) > 0$ if and only if $\sum_{j=1}^\infty a_j < \infty$.

To start, note that for all $x \in [0,1)$, we have that $0 \leq x \leq -\log(1-x)$. After all, $x + \log(1-x)$ equals 0 at $x = 0$. Also, its derivative $1 - \frac{1}{1-x}$ is negative on $[0,1)$. This tells us that $x + \log(1-x)$ is strictly decreasing as x increases, meaning that the difference of x and $-\log(1-x)$ is less than 0 for all $x > 0$.

This lets us conclude that if $\sum_{j=1}^{\infty} -\log(1-a_j)$ converges, then by comparison test we must also have that $\sum_{j=1}^{\infty} a_j$ converges.

Meanwhile, for all $x \in [0, 1/2)$ we have that $0 \leq -\log(1-x) \leq 2x$. To see this, note that $2x + \log(1-x)$ also equals 0 at $x = 0$. But it's derivative: $2 - \frac{1}{1-x}$, is positive for $x < 1/2$. This tells us that $2x + \log(1-x)$ is strictly increasing for $x \in (0, 1/2)$. So, the difference of $2x$ and $-\log(1-x)$ is greater than 0 for all $x \in (0, 1/2)$.

Importantly, if $\sum_{j=1}^{\infty} a_j$ converges, then we know that all a_j after a certain index J will be in the interval $(0, 1/2)$. Then, since the sum of the $-\log(1-a_j)$ for $j \leq J$ will be finite and since we can use comparison test on the remaining terms, we know that $\sum_{j=1}^{\infty} -\log(1-a_j)$ also converges.

In other words, we've shown that:

$$\sum_{j=1}^{\infty} a_j < \infty \text{ if and only if } \sum_{j=1}^{\infty} -\log(1-a_j) < \infty.$$

Next, note that $\sum_{j=1}^{\infty} -\log(1-a_j)$ converges if and only if $\sum_{j=1}^{\infty} \log(1-a_j)$ converges.

Finally, consider that $\prod_{j=1}^{\infty}(1-a_j) > 0$ if and only if $\sum_{j=1}^{\infty} \log(1-a_j) > -\infty$.

If $\prod_{j=1}^{\infty}(1-a_j) = \alpha > 0$, then we know that $\log(\alpha)$ is a finite negative value. And because $\log$ is a continuous function, we know:

$$\begin{aligned} \log(\alpha) &= \log(\lim_{N\to\infty} \textstyle\prod_{j=1}^{N}(1-a_j)) \\ &= \lim_{N\to\infty} \log(\textstyle\prod_{j=1}^{N}(1-a_j)) = \lim_{N\to\infty} \displaystyle\sum_{j=1}^{N} \log(1-a_j) = \sum_{j=1}^{\infty} \log(1-a_j) \end{aligned}$$

Meanwhile, if $\prod_{j=1}^{\infty}(1-a_j) = \lim_{N\to\infty} \prod_{j=1}^{N}(1-a_j) = 0$, then we know that:

$$\sum_{j=1}^{\infty} \log(1-a_j) = \lim_{N\to\infty} \sum_{j=1}^{N} \log(1-a_j) = \lim_{N\to\infty} \log(\prod_{j=1}^{N}(1-a_j)) = -\infty$$

Note that we always have that $\prod_{j=1}^{\infty}(1-a_j) \in [0, 1)$.

(b) Given $\beta \in (0, 1)$, there exists a sequence $(a_j)_{j\in\mathbb{N}}$ in $(0, 1)$ such that $\prod_{j=1}^{\infty}(1-a_j) = \beta$.

Let $c_j = \frac{1}{2^j}(1-\beta) + \beta$ for all $j \in \mathbb{N}$. That way $(c_j)_{j\in\mathbb{N}}$ is a strictly decreasing sequence in $(0, 1)$ converging to β. Then, we want to make $\prod_{i=1}^{j}(1-a_j) = c_j$ for all j. To do this, set $a_j = 1 - \frac{c_j}{c_{j-1}}$ for all j. Because $0 < c_j < c_{j-1}$, we know that $\frac{c_j}{c_{j-1}} \in (0, 1)$. And thus, $a_j \in (0, 1)$ for all j as well and $\prod_{j=1}^{\infty}(1-a_j) = \beta$.

Letting $C = \bigcap C_j$ be the standard Cantor set (i.e. where all $\alpha_j = 1/3$), we now define the Cantor function:

Note that if $x \in C$, then there exists a unique sequence $(a_j)_{j\in\mathbb{N}}$ with $x = \sum_{j=1}^{\infty} a_j \frac{1}{3^j}$ and all a_j equal to either 0 or 2. (This is because each choice of a_j as either 0 or 2 corresponds to which subinterval of C_j that x is in.) Let $f(x) = \sum_{j=1}^{\infty} b_j 2^{-j}$ where $b_j = \frac{a_j}{2}$. Note that $f(x)$ is the binary expansion of a number in $[0,1]$.

Observe that for all $y \in [0,1]$ there exists $x \in C$ with $f(x) = y$. Also, for $x_1, x_2 \in C$ with $x_1 < x_2$, we have that $f(x_1) \le f(x_2)$ with equality if and only if x_1 and x_2 are the end points of a removed interval (thus making x_1 and x_2 correspond to the binary expansions $0.b_1b_2\ldots0\overline{1}$ and $0.b_1b_2\ldots1\overline{0}$). This allows us to continuously extend f to all $[0,1]$ by making f constant on all the intervals between points of C.

Exercise 2.9: Let $f : [0,1] \longrightarrow [0,1]$ be the Cantor function, and let $g(x) = f(x) + x$.

(a) g is a bijection from $[0,1]$ to $[0,2]$, and $h = g^{-1}$ is continuous from $[0,2]$ to $[0,1]$.

To start, note that g is easily checked to be a strictly increasing function. This proves both that g is injective and that g has a range of $[0,2]$ since $g(0) = 0$ and $g(1) = 2$. Also, note that g is continuous since both $f(x)$ and x are continuous. Thus, by applying I.V.T, we can say that g is surjective from $[0,1]$ to $[0,2]$. This proves that g is a bijection.

The proof that $h = g^{-1}$ is continuous works for any strictly increasing continuous function.

Given $y \in [0,2]$ there exists $x \in [0,1]$ such that $g(x) = y$.

Now let $\varepsilon > 0$ and set $\alpha = \max(0, x-\varepsilon)$ and $\beta = \min(1, x+\varepsilon)$. Then for all $y' \in (g(\alpha), g(\beta))$, we have that $h(y') \in B(\varepsilon, x)$. So we can set $\delta = \min(|g(\alpha) - y|, |g(\beta) - y|)$. This fulfills the definition of continuity.

(b) $m(g(C)) = 1$ where $C = \bigcap C_j$ is the Cantor set.

Because g^{-1} is continuous, we know that $g(C)$ and $g(C_j)$ are measurable for all j since C and each C_j are Borel sets. Also $(g(C_j))_{j\in\mathbb{N}}$ is a decreasing sequence of sets with $m(g(C_1)) \le 2$ and $\bigcap_{j\in\mathbb{N}} g(C_j) = g(C)$. Thus $m(g(C)) = \lim_{j\to\infty} m(g(C_j))$.

Next note that C_j has 2^j many intervals, each with width 3^{-j}. Also, if $[\alpha, \alpha + 3^{-j}]$ is one of those intervals, then:

$$\begin{aligned} g(\alpha + 3^{-j}) - g(\alpha) &= f(\alpha + 3^{-j}) + \alpha + 3^{-j} - f(\alpha) - 3^{-j} \\ &= \textstyle\sum_{i=1}^{j}(b_i 2^{-i}) + 2^{-j} + \alpha + 3^{-j} - \sum_{i=1}^{j}(b_i 2^{-i}) - \alpha \\ &= 2^{-j} + 3^{-j} \end{aligned}$$

Thus $m(g(C_j)) = 2^j(2^{-j} + 3^{-j}) = 1 + (\frac{2}{3})^j$. Taking $j \to \infty$ we get the desired result.

To do the next parts of that exercise, we first need to do a different exercise.

Exercise 1.29:

(a) Suppose $E \subseteq V$ where V is a Vitali set (see the tangent on page 22) and $E \in \mathcal{L}$. Prove that $m(E) = 0$.

For all $r \in \mathbb{Q} \cap [-1, 1]$, define $E_r = \{v + r : v \in E\}$. By translation invariance, we know that E_r is measurable with $m(E_r) = m(E)$ for all r. Also each E_r is disjoint and $\bigcup_{r \in \mathbb{Q} \cap [-1,1]} E_r \subseteq [-1, 2]$. It follows that $\bigcup_{r \in \mathbb{Q} \cap [-1,1]} E_r$ is measurable and:

$$3 \geq m(\bigcup_{r \in \mathbb{Q} \cap [-1,1]} E_r) = \sum_{r \in \mathbb{Q} \cap [-1,1]} m(E_r) = \sum_{r \in \mathbb{Q} \cap [-1,1]} m(E)$$

The only way this is possible is if $m(E) = 0$.

(b) If $m(E) > 0$, then there exists a nonmeasurable set $N \subseteq E$.

Sidenote: the converse of this statement is trivially true because $(\mathbb{R}, \mathcal{L}, m)$ is complete.

To start, it suffices to show this for $E \subseteq [0, 1]$. After all, we can use the translation invariance of the Lebesgue measure to move N from $[0, 1]$ to where ever E is (we can't have that $N + r$ is measurable because that would imply $(N + r) - r = N$ is measurable).

Now, let V be a Vitali set and $V_r = \{v + r : v \in V\}$ for all $r \in [-1, 1] \cap \mathbb{Q}$. If $E \cap V_r$ is not measurable for some r, then we are done. So, suppose $E \cap V_r$ is measurable for all r. Then note $\bigcup_{r \in [-1,1] \cap \mathbb{Q}} (E \cap V_r) = E \cap \bigcup_{r \in [-1,1] \cap \mathbb{Q}} (V_r)$. Since $[0, 1]$ is a subset of $\bigcup_{r \in [-1,1] \cap \mathbb{Q}} (V_r)$ and $E \subseteq [0, 1]$, we thus know that $E \cap \bigcup_{r \in [-1,1] \cap \mathbb{Q}} (V_r) = E$. Additionally, since each $E \cap V_r$ is disjoint, we know that:

$$m(E) = \sum_{r \in [-1,1] \cap \mathbb{Q}} m(E \cap V_r)$$

Now hopefully it's clear how part (a) of this exercise extends to each nonmeasurable set V_r. Thus, since we assumed each $E \cap V_r$ is a measurable set, we know that $m(E \cap V_r) = 0$. It follows that $m(E) = 0$, a contradiction of our problem.

Now we return to exercise 2.9.

(c) $g(C)$ contains a Lebesgue nonmeasurable set A by exercise 1.29. Let $B = g^{-1}(A)$. Then B is Lebesgue measurable but not Borel.

Since C is a measurable null set in the complete measure space $(\mathbb{R}, \mathcal{L}, m)$, we have that all subsets of C including B must be measurable. So $B \in \mathcal{L}$.

Side note: since $\mathrm{card}(C) = \mathrm{card}(\mathbb{R})$, we know that:

$$\mathrm{card}(\mathcal{P}(\mathbb{R})) = \mathrm{card}(\mathcal{P}(C)) \leq \mathrm{card}(\mathcal{L}) \leq \mathrm{card}(\mathcal{P}(\mathbb{R})).$$

However, because g^{-1} is continuous, we know that g^{-1} is a Borel measurable function. Hence, if B was borel, then we would have to have that $g(B) = A$ is also Borel, thus contradicting that A is not measurable. So, we know B is measurable but not Borel.

(d) There exists a Lebesgue measurable function F and continuous function G on $\mathbb{R}$ such that $F \circ G$ is not Lebesgue measurable.

Define G by continuously extending $g^{-1}(x)$ to all $\mathbb{R}$ (One way to do this would be to set $G(x) = x$ when $x < 0$ and $G(x) = 1$ when $x > 2$). Then set $F = \chi_B$ where B is the set found in part (c). Now $(F \circ G)^{-1}(\{1\}) = A$ is not Lebesgue measurable. So $F \circ G$ is not a Lebesgue measurable function.

The significance of this result is that we've proven that G is continuous but not Lebesgue measurable.

One more interesting observation Folland makes is that the collection of Borel sets $\mathcal{B}_{\mathbb{R}}$ only has the cardinality of the continuum, meaning that most measurable sets are not Borel.

To prove this, firstly note that by exercise 1.3 in my LaTeX math 240A notes (page 11), we know that $\mathrm{card}(\mathcal{B}_{\mathbb{R}}) \geq \mathfrak{c}$.

Also, consider the following lemmas:

1. **Proposition 0.14:**
   (a) If $\mathrm{card}(X) \leq \mathfrak{c}$ and $\mathrm{card}(Y) \leq \mathfrak{c}$, then $\mathrm{card}(X \times Y) \leq \mathfrak{c}$.

   Proof:
   It suffices to take $X = Y = \mathcal{P}(\mathbb{N})$ since then both X and Y have the largest cardinality we are allowing. Next, define $\psi, \phi : \mathbb{N} \to \mathbb{N}$ by $\psi(n) = 2n$ and $\phi(n) = 2n - 1$. Then $f : \mathcal{P}(\mathbb{N}) \times \mathcal{P}(\mathbb{N}) \to \mathcal{P}(\mathbb{N})$ given by $f(A, B) = \psi(A) \cup \phi(B)$ is a bijection.

   (b) If $\mathrm{card}(A) \leq \mathfrak{c}$ and $\mathrm{card}(\mathcal{E}_\alpha) \leq \mathfrak{c}$ for all $\alpha \in A$, then $\mathrm{card}(\bigcup_{\alpha \in A} \mathcal{E}_\alpha) \leq \mathfrak{c}$.

   Proof:
   For each $\alpha \in A$ there is a surjection $f_\alpha : \mathbb{R} \longrightarrow \mathcal{E}_\alpha$. So define the function $f : \mathbb{R} \times A \longrightarrow \bigcup_{\alpha \in A} \mathcal{E}_\alpha$ by $f(x, \alpha) = f_\alpha(x)$. This is a surjection. So, we know that $\mathrm{card}(\bigcup_{\alpha \in A} \mathcal{E}_\alpha) \leq \mathrm{card}(\mathbb{R} \times A)$ and the latter set by part (a) has no greater than the cardinality of the continuum.

2. If $\mathrm{card}(\mathcal{E}) \leq \mathfrak{c}$, then $\mathrm{card}(\mathcal{E}^\omega) \leq \mathfrak{c}$.

   To see this, we can assume $\mathcal{E} = \{0,1\}^\omega$ since then $\mathrm{card}(\mathcal{E}) = \mathfrak{c}$. Then note that we can use a diagonalization argument to create a bijection between between $\mathcal{E}$ and $\mathcal{E}^\omega$. Writing it out would be a pain so do it yourself.

Now recall from Folland's proposition 1.23 (the bonus proposition written on page 38 of my LaTeX notes for math 240A) the following construction of $\mathcal{B}_\mathbb{R}$.

Let S_Ω be a minimal uncountable set (by constructing S_Ω from $\mathbb{R}$ using the construction I copied from Munkres on page 14 of this pdf, we can guarentee that $S_\Omega \subseteq \mathbb{R}$ and thus $\mathrm{card}(S_\Omega) \leq \mathfrak{c}$).

Next, using 0 to refer to the least element of S_Ω, let $\mathcal{E}_0$ be the set of all rays of the form $[a, \infty)$ where $a \in \mathbb{R}$. Then for all other $\alpha \in S_\Omega$:

- If α has a direct predecessor β, then let $\mathcal{E}_\alpha$ be the collection of all countable unions of and complements of sets from $\mathcal{E}_\beta$.
- If α does not have a direct predecessor, then set $\mathcal{E}_\alpha = \bigcup_{\beta \in S_\alpha} \mathcal{E}_\alpha$.

Finally, $\mathcal{B}_\mathbb{R} = \bigcup_{\alpha \in S_\Omega} \mathcal{E}_\alpha$.

We obviously have that $\mathrm{card}(\mathcal{E}_0) = \mathfrak{c}$. Then using transfinite induction along with our two previously mentioned lemmas, we can conclude that $\mathrm{card}(\mathcal{E}_\alpha) \leq \mathfrak{c}$ for all $\alpha \in S_\Omega$. So by part (b) of proposition 0.14, we conclude that:

$$\mathrm{card}(\mathcal{B}_\mathbb{R}) = \mathrm{card}(\textstyle\bigcup_{\alpha \in S_\Omega} \mathcal{E}_\alpha) \leq \mathfrak{c}.$$

Since $\mathfrak{c} \leq \mathrm{card}(\mathcal{B}_\mathbb{R}) \leq \mathfrak{c}$, we know that $\mathrm{card}(\mathcal{B}_\mathbb{R}) = \mathfrak{c}$.

7/5/2025

I'm going to be taking more analysis notes from Folland. I'm starting with the section: <u>The Dual of $C_0(X)$</u>. Here, X refers to an LCH space.

To start out, we shall identify all positive bounded linear functionals on $C_0(X)$. Note that if I is such a functional on $C_0(X)$, then we know it is also a positive bounded linear functional on the subspace $C_c(X)$. Meanwhile going in reverse, we have that if $I(f) = \int f d\mu$ is a positive linear function on $C_c(X)$ that is bounded, then we can uniquely extend it to a positive bounded linear functional on $C_0(X)$ by defining $I(f) = \lim_{n\to\infty} I(f_n)$ for any $f \in C_0(X)$ where $\{f_n\}_{n\in\mathbb{N}}$ is any sequence in $C_c(X)$ converging to f uniformly. So, given any Radon measure μ, we need to determine when $I(f) = \int f d\mu$ is bounded.

Since X is open and μ is Radon, by the Riesz Representation theorem:

$$\mu(X) = \sup\{I(f) : f \in C_c(X),\ \mathrm{supp}(f) \subseteq X,\ 0 \leq f \leq 1\}.$$

The second condition is redundant and $I(f) = \int f \mathrm{d}\mu$. So we can rewrite this as $\mu(X) = \sup\{\int f \mathrm{d}\mu : f \in C_c(X),\ 0 \leq f \leq 1\}$. We now claim I is bounded iff $\mu(X) < \infty$, and that when I is bounded, $\|I\|_{\mathrm{op}} = \mu(X)$.

($\Longrightarrow$)
Suppose $\mu(X) = \infty$. Then for any $N > 0$, there is a function $f \in C_c(X)$ with $0 \leq f \leq 1$ such that $\int f \mathrm{d}\mu \geq N$. And since $\|f\|_u \leq 1$, we know that if $|I(f)| \leq C\|f\|_u$, then $N \leq |\int f \mathrm{d}\mu| = |I(f)| \leq C$. This proves no finite C works for all $f \in C_c(X)$, and thus I is unbounded.

($\Longleftarrow$)
Suppose $\mu(X) < \infty$ and then consider any $f \in C_c(X)$ with $\|f\|_u = 1$. Note that $|I(f)| = |\int f \mathrm{d}\mu| \leq \int |f| \mathrm{d}\mu$. Then since $0 \leq |f| \leq 1$, we know that $\int |f| \mathrm{d}\mu \leq \mu(X)$. So $\|I\|_{\mathrm{op}}$ exists and is at most $\mu(X)$.

To prove that $\mu(X) = \|I\|_{\mathrm{op}}$, let $\varepsilon > 0$ and pick $f \in C_c(X)$ with $0 \leq f \leq 1$ such that $\int f \mathrm{d}\mu > \mu(X) - \varepsilon$. Thus we have that $\|I\|_{\mathrm{op}}\|f\|_u > \mu(X) - \varepsilon$. Then since $\|f\|_u \leq 1$, we have that $\|I\|_{\mathrm{op}} > \mu(X) - \varepsilon$. Taking $\varepsilon \to 0$ finishes the proof.

So, the positive bounded linear functionals on $C_0(X)$ are precisely given by integration against finite Radon measures (and this correspondence is one-to-one by the Riesz Representation theorem). Next, we identify the other linear functionals on $C_0(X)$.

<u>Lemma 7.15:</u> If $I \in C_0(X, \mathbb{R})^*$, there exists positive functionals $I^{\pm} \in C_0(X, \mathbb{R})^*$ such that $I = I^+ - I^-$.

Proof:
If $f \in C_0(X, [0, \infty))$, define:

$$I^+(f) = \sup\{I(g) : g \in C_0(X, \mathbb{R}),\ 0 \leq g \leq f\}.$$

If $c \geq 0$, then clearly $I^+(cf) = cI^+(f)$. Meanwhile, let $f_1, f_2 \in C_0(X, [0, \infty))$. To show that $I^+(f_1 + f_2) = I^+(f_1) + I^+(f_2)$, first suppose $0 \leq g_1 \leq f_1$ and $0 \leq g_2 \leq f_2$. Then $0 \leq g_1 + g_2 \leq f_1 + f_2$. So, $I^+(f_1 + f_2) \geq I(g_1 + g_2) = I(g_1) + I(g_2)$. By taking $I(g_1) \to I^+(f_1)$ and $I(g_2) \to I^+(f_2)$, we then get that $I^+(f_1 + f_2) \geq I^+(f_1) + I^+(f_2)$.

On the other hand, if $0 \leq g \leq f_1 + f_2$, let $g_1 = \min(g, f_1)$ and $g_2 = g - g_1$. Thus $0 \leq g_1 \leq f_1$, $0 \leq g_2 \leq f_2$, and g_1, g_2 are continuous. This guarentees $g_1, g_2 \in C_0(X, [0, \infty))$. So, $I(g) = I(g_1) + I(g_2) \leq I^+(f_1) + I^+(f_2)$. And, taking $I(g) \to I^+(f_1 + f_2)$ gets us $I^+(f_1 + f_2) \leq I^+(f_1) + I^+(f_2)$.

Now, we extend I^+ to a positive linear functional in $C_0(X, \mathbb{R})^*$. Given $f \in C_0(X, \mathbb{R})$, let f^+ and f^- be the positive and negative parts of f. Then, $f^+, f^- \in C_0(X, [0, \infty))$. So, define $I^+(f) = I^+(f^+) - I^+(f^-)$. This is linear because if $c \in \mathbb{R}$, then ignoring the trivial edge case where $c = 0$:

$$\begin{aligned} I^+(cf) &= \operatorname{sgn}(c)\,(I^+(|c|f^+) - I^+(|c|f^-)) \\ &= \operatorname{sgn}(c)|c|\,(I^+(f^+) - I^+(f^-)) = cI^+(f). \end{aligned}$$

Also, suppose $f = g + h$ where $g, h \in C_0(X, \mathbb{R})$. Then $f^+ + g^- + h^- = f^- + g^+ + h^+$ where all the functions in that expression are in $C_0(X, [0, \infty))$. So, we know from our earlier work that:

$$\begin{aligned} I^+(f^+) + I^+(g^-) + I^+(h^-) &= I^+(f^+ + g^- + h^-) \\ &= I^+(f^- + g^+ + h^+) = I^+(f^-) + I^+(g^+) + I^+(h^+) \end{aligned}$$

Or in other words:

$$I^+(f) = I^+(f^+) - I^+(f^-) = I^+(g^+) - I^+(g^-) + I^+(h^+) - I^+(h^-) = I^+(g) + I^+(h)$$

To show that I^+ is bounded, first note that if $f \in C_0(X, [0, \infty))$, then since $|I(g)| \leq \|I\|\|g\|_u \leq \|I\|\|f\|_u$ for all $0 \leq g \leq f$ and $I(0) = 0$, we have $0 \leq I^+(f) \leq \|I\|\|f\|_u$. (Note, this also proves I^+ is positive). Meanwhile, if $f \in C_0(X, \mathbb{R})$, then $I^+(f) = I^+(f^+) - I^+(f^-)$ where both terms in that difference are positive. Hence, we can say that:

$$|I^+(f)| \leq \max(I^+(f^+), I^-(f^-)) \leq \|I\| \max(\|f^+\|_u, \|f^-\|_u) = \|I\|\|f\|_u$$

Thus, we've finished constructing I^+. So now define $I^-(f) = I^+(f) - I(f)$. Then we know $I^- \in C_0(X, \mathbb{R})^*$ because $C_0(X, \mathbb{R})^*$ is a real vector space. Also, I^- is positive because if $f \geq 0$, then you can see from our definition of $I^+(f)$ on $C_0(X, [0, \infty))$ that $I^+(f) \geq I(f)$. Hence, $I^-(f) = I^+(f) - I(f)$ is also nonnegative. ■

Now any $I \in C_0(X)^*$ is uniquely determined by its restriction J to $C_0(X, \mathbb{R})$.

Why:
$I(f) = I(\operatorname{Re}(f) + i\operatorname{Im}(f)) = I(\operatorname{Re}(f)) + iI(\operatorname{Im}(f)) = J(\operatorname{Re}(f)) + iJ(\operatorname{Im}(f))$.

Next, there are two real linear functionals $J_1, J_2 \in C_0(X, \mathbb{R})^*$ such that $J = J_1 + iJ_2$. Specifically, set $J_1(f) = \operatorname{Re}(J(f))$ and $J_2 = \operatorname{Im}(J(f))$. Then clearly J_1 and J_2 are real linear functionals and they are bounded with $\|J_i\| \leq \|I\|$.

Using our lemma, we can decompose J_1 and J_2 into differences of positive bounded linear real functionals. I.e., we write $J = J_1^+ - J_1^- + i(J_2^+ - j_2^-)$.

Finally, define $I_1^+, I_1^-, I_2^+, I_2^-$ such that $I_1^+(f) = J_1^+(\operatorname{Re}(f)) + iJ_1^+(\operatorname{Im}(f))$ and the others have analogous definitions. Then all of our $I_i^\pm$ are well-defined complex linear functionals on $C_0(X)$ that are bounded since:

$$|I_i^\pm(f)| \leq \|J_i^\pm\|(\|\operatorname{Re}(f)\|_u + \|\operatorname{Im}(f)\|_u) \leq 2\|J_i^\pm\|\|f\|_u.$$

Also, all the $I_i^\pm$ are positive since if f is nonnegative, then $I_i^\pm(f) = J_i^\pm(f)$. This means that there are finite Radon measures $\mu_1, \mu_2, \mu_3, \mu_4$ such that $I_1^+(f) = \int f d\mu_1$, $I_1^-(f) = \int f d\mu_2$, $I_2^+(f) = \int f d\mu_3$. and $I_2^-(f) = \int f d\mu_4$.

Additionally:

$$\begin{aligned}I(f) &= J(\mathrm{Re}(f)) + iJ(\mathrm{Im}(f)) \\ &= J_1(\mathrm{Re}(f)) + iJ_2(\mathrm{Re}(f)) + iJ_1(\mathrm{Im}(f)) + i^2J_2(\mathrm{Im}(f)) \\ &= J_1^+(\mathrm{Re}(f)) - J_1^-(\mathrm{Re}(f)) + iJ_2^+(\mathrm{Re}(f)) - iJ_2^-(\mathrm{Re}(f)) \\ &\quad + iJ_1^+(\mathrm{Im}(f)) - iJ_1^-(\mathrm{Im}(f)) + i^2J_2^+(\mathrm{Im}(f)) - i^2J_2^-(\mathrm{Im}(f)) \\ &= I_1^+(f) - I_2^-(f) + iI_2^+(f) - iI_2^-(f)\end{aligned}$$

So for any $I \in C_0(X)^*$, there are finite Radon measures $\mu_1, \mu_2, \mu_3, \mu_4$ such that $I(f) = \int f \mathrm{d}\mu$ where $\mu = \mu_1 - \mu_2 + i\mu_3 - i\mu_4$.

7/6/2025

I'm continuing on in Folland where I left off.

A signed Radon measure is a signed Borel measure such that it's positive and negative variations are Radon.

A complex Radon measure is a complex Borel measure such that it's real and imaginary variations are signed Radon measures.

Side note: Complex Borel measures are always finite on compact sets. Thus if X is an LCH space in which every open set is σ-compact, we know by theorem 7.8 that all complex Borel measures are Radon. In particular, if X is a second countable LCH space, then all complex Borel measures are Radon.

We denote the space of complex Radon measures on $(X, \mathcal{B}_X)$ as $M(X)$ and for $\mu \in M(X)$ we define $\|\mu\| = |\mu|(X)$ where $|\mu|$ is the total variation of μ.

Proposition 7.16: If μ is a complex Borel measure, then μ is Radon iff $|\mu|$ is Radon. Moreover, $M(X)$ is a vector space and $\mu \mapsto \|\mu\|$ is a norm on that space.

Proof:

By proposition 7.5 (which says that Radon measures are inner regular on all their σ-finite sets), we know that a finite positive measure $|\mu|$ is Radon iff for any Borel set E and $\varepsilon > 0$ there is an open set U and a compact set K with $K \subseteq E \subseteq U$ and $\mu(U - K) < \varepsilon$. From this we show the first assertion as follows. If $\mu = \mu_1 - \mu_2 + i\mu_3 - i\mu_4$ where all the μ_j are finite positive measures, and $|\mu|(U - K) < \varepsilon$, then $\mu_j(U - K) < \varepsilon$ for all j.

Why: (Also I'm going into more detail cause I am having trouble remembering how to work with the total variation of a complex measure.) Let ν be some positive measure with $\mu \ll \nu$. Then $\mu_j \ll \nu$ for each j, so for each j there are functions f_j with $\mathrm{d}\mu_j = f_j \mathrm{d}\nu$. Also, $\mathrm{d}\mu = (f_1 - f_2 + i(f_3 - f_4))\mathrm{d}\nu$ and $\mathrm{d}|\mu| = |f_1 - f_2 + i(f_3 - f_4)|\mathrm{d}\nu$.

Now since all the f_j are real-valued, we have:
$$|f_1 - f_2 + i(f_3 - f_4)| \geq \max(|f_1 - f_2|, |f_3 - f_4|).$$

Next, since $\mu_1 \perp \mu_2$ and $\mu_3 \perp \mu_4$ and all the measures are positive, we know that $\min(f_1, f_2) = 0$ and $\min(f_3, f_4) = 0$ ν-a.e. Hence,
$$\max(|f_1 - f_2|, |f_3 - f_4|) \geq \max(f_1, f_2, f_3, f_4) \text{ a.e.}$$

And so, we get $|\mu|(E) \geq \max(\mu_1(E), \mu_2(E), \mu_3(E), \mu_4(E))$ for all $E \in \mathcal{B}_X$.

Meanwhile if we can pick U_j, K_j for all j such that $\mu_j(U_j - K_j) < \varepsilon/4$, then set $U = \bigcap U_j$ and $K = \bigcup K_j$. Now, $|\mu|(U - K) < 4 \cdot \varepsilon/4 = \varepsilon$.

Why: By proposition 3.14,
$$\begin{aligned} |\mu| = |\mu_1 - \mu_2 + i\mu_3 - i\mu_4| &\leq |\mu_1| + |-\mu_2| + |i\mu_3| + |-i\mu_4| \\ &= \mu_1 + \mu_2 + \mu_3 + \mu_4. \end{aligned}$$

Then since $\mu_j(U) \leq \mu_j(U_j)$ and $\mu_j(K) \geq \mu(K_j)$ for all j, the claim holds.

Similar reasoning to that right above can show that $M(X)$ is closed under addition, and that $\|\mu_1 + \mu_2\| \leq \|\mu_1\| + \|\mu_2\|$. Also if $\mathrm{d}\mu = f\mathrm{d}\nu$ for some positive measure ν, then $c\mathrm{d}\mu = cf\mathrm{d}\nu$. So $|c\mathrm{d}\mu| = |c|\mathrm{d}|\mu|$, and from that it is clear that $|\mu|$ being Radon implies $|c\mathrm{d}\mu|$ is Radon. So, $M(X)$ is closed under scalar multiplication. Note this also shows that $\|c\mu\| = |c|\|\mu\|$ for all $c \in \mathbb{C}$ and $\mu \in M(X)$.

Finally, suppose $\mu \in M(X)$ with $\mu \neq 0$. Then there is some set $E \in \mathcal{B}_X$ such that $\mu(E) \neq 0$. Next $0 < |\mu(E)| \leq |\mu|(E)$ (see proposition 3.13). And since $|\mu|$ is Radon, we know that:
$$0 < |\mu|(E) = \inf\{|\mu|(U) : E \subseteq U \text{ where } U \text{ is open}\} \leq |\mu|(X) = \|\mu\|.$$

This proves, $M(X)$ is a normed vector space when equipped with $\|\cdot\|$. ■

7/7/2025

Before getting to the next theorem, I'd like to return to when I showed that for any $I \in C_0(X)^*$, there are finite Radon measures $\mu_1, \mu_2, \mu_3, \mu_4$ such that $I(f) = \int f\mathrm{d}\mu$ where $\mu = \mu_1 - \mu_2 + i\mu_3 - i\mu_4$.

A thing that Folland neglected to show is that while it's clear that $\mu_1 - \mu_2$ and $\mu_3 - \mu_4$ are the real and imaginary variations of μ respectively, it's not necessarily clear that μ_1 and μ_2 are the positive and negative variations respectively of $(\mu_1 - \mu_2)$ and likewise for μ_3 and μ_4 with respect to $(\mu_3 - \mu_4)$. So, I want to show that today since this will be relevant to the next propositions that Folland covers.

Lemma (Riesz Representation theorem on $C_0(X, \mathbb{R})^*$): There is a one-to-one correspondance between positive linear functionals in $C_0(X, \mathbb{R})^*$ and finite Radon measures on $(X, \mathcal{B}_X)$.

Proof:

Recall that there is a bijection between $C_0(X, \mathbb{R})^*$ and $C_0(X)^*$. Namely given any $I \in C_0(X, \mathbb{R})^*$, define $J \in C_0(X)^*$ by setting $J(f) = I(\mathrm{Re}(f)) + iI(\mathrm{Im}(f))$ for all f. And to go the other way, just restrict the domain of J.

Now in that bijection, it's clear that I is positive iff J is positive. Also, it's clear that $I(f) = \int f d\mu$ for all $f \in C_0(X, \mathbb{R})$ iff $J(f) = \int f d\mu$ for all $f \in C_0(X)$. So, we can apply the Riesz Representation theorem we already proved to say that there is a bijective correspondence between finite Radon measures on X and $\{I \in C_0(X, \mathbb{R})^* : I \text{ is positive}\}$.

Now suppose $I \in C_0(X, \mathbb{R})^*$ and let $I = I^+ - I^-$ where $I^\pm \in C_0(X, \mathbb{R})^*$ are as we constructed in Lemma 7.15. As we just demonstrated, there are finite Radon measures μ_1 and μ_2 such that $I^+(f) = \int f d\mu_1$ and $I^-(f) = \int f d\mu_2$. In turn, setting $\mu = \mu_1 - \mu_2$ we have that $I(f) = \int f d\mu$.

Exercise 7.16: The positive and negative variations of μ are the Radon measures μ_1 and μ_2 respectively.

Let μ^+ and μ^- be the positive and negative variations of μ, and let $E \in \mathcal{B}_X$ be a set such that $\mu^+(E) = 0$ and $\mu^-(E^{\mathsf{C}}) = 0$.

Fixing $f \in C_0(X, [0, \infty))$, note that:

$$\begin{aligned} I^+(f) &= \sup\{I(g) : g \in C_0(X, \mathbb{R}), 0 \leq g \leq f\} \\ &= \sup\{\textstyle\int g d\mu^+ - \int g d\mu^- : g \in C_0(X, \mathbb{R}), 0 \leq g \leq f\} \\ &\leq \sup\{\textstyle\int g d\mu^+ : g \in C_0(X, \mathbb{R}), 0 \leq g \leq f\} = \int f d\mu^+ \end{aligned}$$

On the other hand, since μ_1, μ_2 are finite Radon measures and I showed yesterday that $M(X)$ is a vector space, I know that $\mu = \mu_1 - \mu_2$ is also a finite Radon measure. Also from yesterday, I know that that is equivalent to saying that $|\mu| = \mu^+ + \mu^-$ is Radon. Plus, μ being finite implies $|\mu|$ is finite. Hence, f vanishes outside of a set with finite measure (that set being all of X). So, for any $\varepsilon > 0$ we can apply Lusin's theorem to get a function $\phi \in C_c(X)$ with $\phi = f\chi_{E^{\mathsf{C}}}$ except on a set $F \in \mathcal{B}_X$ with $|\mu|(F) < \varepsilon$.

If we then set $\psi = \min(\mathrm{Re}(\phi)^+, f)$, we still have that $\psi = f\chi_{E^{\mathsf{C}}}$ except on F. But then also $\psi \in C_C(X, \mathbb{R}) \subseteq C_0(X, \mathbb{R})$ with $0 \leq \psi \leq f$. So:

$$\begin{aligned} I^+(f) \geq \textstyle\int \psi d\mu &= \textstyle\int \psi d\mu^+ - \int \psi d\mu^- \\ &= \textstyle\int_F \psi d\mu^+ + \int_{F^{\mathsf{C}}} f\chi_{E^{\mathsf{C}}} d\mu^+ - \int_F \psi d\mu^- - \int_{F^{\mathsf{C}}} f\chi_{E^{\mathsf{C}}} d\mu^- \\ &\geq 0 + \textstyle\int_{F^{\mathsf{C}}} f\chi_{E^{\mathsf{C}}} d\mu^+ - \int_F \psi d\mu^- - \int_{F^{\mathsf{C}}} f\chi_{E^{\mathsf{C}}} d\mu^- \\ &= \textstyle\int_{F^{\mathsf{C}}} f d\mu^+ - \int_F \psi d\mu^- - 0 \\ &= \textstyle\int f d\mu^+ - \int_F f d\mu^+ - \int_F \psi d\mu^- - 0 \\ &\geq \textstyle\int f d\mu^+ - 2\|f\|_u \mu(F) > \int f d\mu^+ - 2\varepsilon\|f\|_u \end{aligned}$$

Since f was fixed, by taking $\varepsilon \to 0$ we have thus proven that $I^+(f) = \int f \mathrm{d}\mu^+$ for all $f \in C_0(X, [0, \infty))$. Then by considering positive and negative parts and making use of the linearity of both sides, we can easily see $I^+(f) = \int f \mathrm{d}\mu^+$ for all $f \in C_0(x, \mathbb{R})$. This proves that μ^+ is the unique Radon measure associated with I^+. Hence, $\mu^+ = \mu_1$.

Also, since $I^-(f) = I^+(f) - I(f) = \int f \mathrm{d}\mu^+ - (\int f \mathrm{d}\mu^+ - \int f \mathrm{d}\mu^-)$, we have $I^-(f) = \int f \mathrm{d}\mu^-$ for all $f \in C_0(X, \mathbb{R})$. So $\mu^- = \mu_2$. ■

Now as seen in the first lemma I showed today, if we extend $I^\pm \in C_0(X, \mathbb{R})^*$ to be a linear functional in $C_0(X)^*$, it doesn't change the measure μ at all. So, I'm done.

7/8/2025

Firstly, I'm going to finish describing $C_0(X)^*$.

Proposition 7.17 (The Riesz Representation Theorem): Let X be an LCH space, and for $\mu \in M(X)$ and $f \in C_0(X)$, let $I_\mu(f) = \int f \mathrm{d}\mu$. Then the map $\mu \mapsto I_\mu$ is an isometric isomorphism from $M(X)$ to $C_0(X)^*$.

Proof:
We already have shown that every $I \in C_0(X)^*$ is of the form I_μ for some $\mu \in M(X)$. On the other hand, if $\mu \in M(X)$, then we already know that I_μ is a linear function. Also, by proposition 3.13:

$$|\int f \mathrm{d}\mu| \leq \int |f| \mathrm{d}|\mu| \leq \|f\|_u \|\mu\|.$$

So, I_μ is bounded with $\|I_\mu\| \leq \|\mu\|$.

All we have left to do is show $\|\mu\| \leq \|I_\mu\|$. So let $h = \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}$. Then since $|h| = 1$ by proposition 3.13 and $|\mu|$ is a finite Radon measure, we know by Lusin's theorem that for any $\varepsilon > 0$ there exists $f \in C_c(X)$ such that $\|f\|_u \leq \|h\|_u$ and $f = \overline{h}$ except on a set E with $|\mu|(E) < \varepsilon/2$. (Note that since $|\overline{h}| = 1$ almost everywhere, we have that $\|f\|_u = 1$.)

Now:

$$\begin{aligned}\|\mu\| = \int 1 \mathrm{d}|\mu| = \int |h|^2 \mathrm{d}|\mu| \\ = \int \overline{h} \mathrm{d}\mu \leq |\int f \mathrm{d}\mu| + |\int (f - \overline{h}) \mathrm{d}\mu| \\ \leq |I_\mu(f)| + \int |f - \overline{h}| \mathrm{d}|\mu| \leq \|I_\mu\| + 2|\mu|(E) \\ < \|I_\mu\| + \varepsilon\end{aligned}$$

Thus $\|\mu\| \leq \|I_\mu\|$ and we are done. ■

Corollary 7.18: If X is a compact Hausdorff space, then $C(X)^*$ is isometrically isomorphic to $M(X)$.

Next, I plan on taking a break from Folland chapter 7 in order to do some of the section 8.3 exercises in Folland that I never started or finished during the past Spring quarter.

Exercise 8.14 (Wirtinger's Inequality) If $f \in C^1([a,b])$ and $f(a) = f(b) = 0$, then:

$$\int_a^b |f(x)|^2 \mathrm{d}x \leq \left(\frac{b-a}{\pi}\right)^2 \int_a^b |f'(x)|^2 \mathrm{d}x$$

Hint: By a change of variables it suffices to assume $a = 0$ and $b = \frac{1}{2}$. Extend f To $[-\frac{1}{2}, \frac{1}{2}]$ by setting $f(-x) = -f(x)$, and then extend f to be periodic on $\mathbb{R}$. Check that f, thus extended, is in $C^1(\mathbb{T})$ and apply the Parseval identity.

Given our f, we can define $g(x) \coloneqq f(a + 2x(b-a))$. Then $g \in C^1([0, \frac{1}{2}])$ with $g(0) = g(\frac{1}{2}) = 0$ and $f(x) = g(\frac{x-a}{2(b-a)})$. Now suppose we prove the inequality for g. I.e., we show $\int_0^{1/2} |g(x)|^2 \mathrm{d}x \leq (\frac{1}{2\pi})^2 \int_0^{1/2} |g'(x)|^2 \mathrm{d}x$. Then:

- $\int_0^{1/2} |g(x)|^2 \mathrm{d}x = \int_0^{1/2} |f(a + 2x(b-a))|^2 \mathrm{d}x = 2(b-a) \int_a^b |f(y)|^2 \mathrm{d}y$,
- $(\frac{1}{2\pi})^2 \int_0^{1/2} |g'(x)|^2 \mathrm{d}x = (\frac{1}{2\pi})^2 \int_0^{1/2} |2(b-a) f'(a + 2x(b-a))|^2 \mathrm{d}x$
  $= (\frac{1}{2\pi})^2 (2(b-a))^3 \int_a^b |f'(y)|^2 \mathrm{d}y = 2(b-a)(\frac{b-a}{\pi})^2 \int_a^b |f'(y)|^2 \mathrm{d}y.$

By canceling out the $2(b-a)$ term (which is positive since $b > a$), we see the result still holds for f if it held for g.

But now we need to actually prove the result for g. To do this, extend out g to all of $\mathbb{R}$ by first setting $g(-x) \coloneqq -g(x)$ for $x \in [0, \frac{1}{2}]$, and then extending g to be periodic on $\mathbb{R}$. Note that this is well defined specifically because $g(0) = g(\frac{1}{2}) = 0$.

To see that g is in $C^1(\mathbb{T})$, note that since $g(x) = -g(-x)$ for $x \in (-\frac{1}{2}, 0)$ we have that $g'(x) = g'(-x)$ on $(-\frac{1}{2}, 0)$. Thus g is continuously differentiable on $(-\frac{1}{2}, 0)$ since we already know g is continuously differentiable on $(0, \frac{1}{2})$.

As for at $x = 0$, note that:

$$\lim_{h \to 0^-} \tfrac{g(h) - g(0)}{h} = \lim_{h \to 0^-} \tfrac{g(h)}{h} = \lim_{h \to 0^-} \tfrac{-g(-h)}{h} = \lim_{h \to 0^+} \tfrac{-g(h)}{-h} = \lim_{h \to 0^+} \tfrac{g(h)}{h} = \lim_{h \to 0^+} \tfrac{g(h) - g(0)}{h}$$

Thus $g'(0)$ still exists on the extended domain. Also, since $\lim_{x \to 0^+} g'(x) = g'(0)$, we know that $\lim_{x \to 0^-} g'(x) = \lim_{x \to 0^-} g'(-x) = \lim_{x \to 0^+} g'(x) = g'(0)$. So, g' is continuous at $t = 0$. Similar reasoning also works at $x = \frac{1}{2}$, although the looping structure of $\mathbb{T}$ makes the expressions slightly messier.

Now since $\mathbb{T}$ is compact, we know that $C(\mathbb{T}) \subseteq L^p$ for all p (and in particular, for $p = 2$). Thus both g and g' are in L^2. Applying Parseval's identity to g we get that:

$$\int_{-1/2}^{1/2} |g(x)|^2 dx = \|g\|_{L^2(\mathbb{T})}^2 = \sum_{k \in \mathbb{Z}} |\widehat{g}(k)|^2 = \sum_{k \in \mathbb{Z}} |\int_{-1/2}^{1/2} g(x) e^{-2\pi i k x} \mathrm{d}x|^2$$

If we do integration by parts, then since $g(\frac{1}{2}) = g(-\frac{1}{2}) = 0$, we get for all $k \neq 0$ that:

$$\widehat{g}(k) = \int_{-1/2}^{1/2} g(x) e^{-2\pi i k x} \mathrm{d}x = \tfrac{1}{2\pi i k} \int_{-1/2}^{1/2} g'(x) e^{-2\pi i k x} \mathrm{d}x = \tfrac{1}{2\pi i k} \widehat{g'}(k)$$

Meanwhile, because of the way we extended g, we know g is an odd function. Thus, $\widehat{g}(0) = \int_{-1/2}^{1/2} g(x)\mathrm{d}x = 0$ and we've thus shown that:

$$\int_{-1/2}^{1/2} |g(x)|^2 dx \leq \sum_{\substack{k\in\mathbb{Z}\\k\neq 0}} \left|\frac{1}{2\pi i k}\widehat{g'}(k)\right|^2 = \frac{1}{4\pi^2}\sum_{\substack{k\in\mathbb{Z}\\k\neq 0}} \frac{1}{k^2}\left|\widehat{g'}(k)\right|^2$$

Now $\sum_{\substack{k\in\mathbb{Z}\\k\neq 0}} \frac{1}{k^2}\left|\widehat{g'}(k)\right|^2 \leq \sum_{k\in\mathbb{Z}} 1\cdot\left|\widehat{g'}(k)\right|^2$ and the latter is equal to $\|g'\|^2_{L^2(\mathbb{T})} = \int_{-1/2}^{1/2}|g'(x)|^2\mathrm{d}x$ by Parseval's identity.

Hence, we've proven that $\int_{-1/2}^{1/2} |g(x)|^2 dx \leq \left(\frac{1}{2\pi}\right)^2 \int_{-1/2}^{1/2} |g'(x)|^2\mathrm{d}x$.

Finally, since $|g(x)|^2$ and $|g'(x)|^2$ are both even on account of g being odd, we know that $\int_{-1/2}^{1/2} |g(x)|^2 dx = 2\int_0^{1/2} |g(x)|^2 dx$ and $\int_{-1/2}^{1/2} |g'(x)|^2\mathrm{d}x = 2\int_0^{1/2} |g'(x)|^2\mathrm{d}$. After canceling out the factor of 2, we've thus proven our desired inequality. ■

Exercise 8.16: Let $f_k = \chi_{[-1,1]} * \chi_{[-k,k]}$. (Also assume $k \in \mathbb{N}$ with $k > 0$).

(a) Compute $f_k(x)$ explicitely and show that $\|f\|_u = 2$.

You can fairly easily see that for any $x \in \mathbb{R}$, $f_k(x) = \int_{-k}^{k} \chi_{[-1,1]}(x-y)\mathrm{d}y$. Evaluating that gives the formula:

$$f_k(x) = \begin{cases} 2 & \text{if } |x| \leq k-1 \\ k-x+1 & \text{if } k-1 \leq x \leq k+1 \\ x+1+k & \text{if } -k-1 \leq x \leq -k+1 \\ 0 & \text{if } |x| \geq k+1 \end{cases}$$

From that it is hopefully clear that $\|f\|_u = 2$. After all, $f_k(0) = 2$. Also, $f_k(x) = \int_{-k}^{k} \chi_{[-1,1]}(x-y)\mathrm{d}y \leq \int \chi_{[-1,1]}(x-y)\mathrm{d}y = 2$.

(b) Show $f_k^\vee(x) = (\pi x)^{-2}\sin(2\pi x)\sin(2\pi k x)$, and $\|f_K^\vee\|_1 \to \infty$ as $k \to \infty$.

Recall from the homework that $\chi^\wedge_{[-a,a]} = \chi^\vee_{[-a,a]} = 2a\frac{\sin(2a\pi x)}{2\pi a x} = \frac{\sin(2\pi a x)}{\pi x}$.

Also, for any $f, g \in L^1$, by identical reasoning as we used to show $\widehat{f * g} = \widehat{f}\widehat{g}$, we know that $(f*g)^\vee = f^\vee g^\vee$. Therefore:

$$\begin{aligned} f_k^\vee(x) = \chi^\vee_{[-1,1]}(x)\chi^\vee_{[-k,k]}(x) &= \left(\frac{\sin(2\pi x)}{\pi x}\right)\left(\frac{\sin(2\pi k x)}{\pi x}\right) \\ &= (\pi x)^{-2}\sin(2\pi x)\sin(2\pi k x). \end{aligned}$$

Next, let $y = 2\pi k x$. Then:

$$\begin{aligned} \int |f_k^\vee(x)|\mathrm{d}x &= \int_{-\infty}^{\infty} |(\pi x)^{-2}\sin(2\pi x)\sin(2\pi k x)|\mathrm{d}x \\ &= \frac{1}{2\pi k}\int_{-\infty}^{\infty} \left|\frac{4k^2}{y^2}\sin\left(\frac{y}{k}\right)\sin(y)\right|\mathrm{d}y = \frac{2}{\pi}\int_{-\infty}^{\infty}\left|\frac{\sin(y/k)}{y/k}\cdot\frac{\sin(y)}{y}\right|\mathrm{d}y \\ &= \frac{4}{\pi}\int_0^{\infty}\left|\frac{\sin(y/k)}{y/k}\cdot\frac{\sin(y)}{y}\right|\mathrm{d}y. \end{aligned}$$

(the last equality holds because the integrand is even)

Now, because $\frac{\sin(x)}{x} \to 1$ as $x \to 0$, we know that $|\frac{\sin(y/k)}{y/k} \cdot \frac{\sin(y)}{y}|$ converges pointwise to $|\frac{\sin(y)}{y}|$ as $k \to \infty$. Also, observe that $|\frac{\sin(x)}{x}| \leq 1$ for all $x > 0$.

Proof:
Let $g(x) = \frac{\sin(x)}{x}$. Then clearly $|g(x)| \leq \frac{1}{x} \leq 1$ when $x \geq 1$.

Meanwhile, if $x < 1$, note that $g'(x) = \frac{x\cos(x) - \sin(x)}{x^2}$. Since $x^2 > 0$, it suffices to show that the numerator: $h(x) = x\cos(x) - \sin(x)$, is negative when $x < 1$ in order to prove that $g'(x)$ is not positive when $x < 1$, Luckily, note that $h(0) = 0$ and $h'(x) = -x\sin(x)$. Since $\sin(x) \geq 0$ for $x \leq \pi \approx 3.14$, we thus know that $h'(x) \leq 0$ for all $x \in [0,1]$. In turn, we know that $h(x) \leq h(0) = 0$ for all $x \in [0,1]$. So, we've proven that $g'(x)$ is not positive on $(0,1]$.

This proves that $g(x)$ is monotonically decreasing on $(0,1)$. And since $\lim_{x \to 0} g(x) = 1$, this proves that $g(x) \leq 1$ for all $x \in (0,1]$. Also, since $\sin(x) > 0$ when $0 < x < \pi \approx 3.14$, we know that $g(x) > 0$ for all $x \in (0,1]$. So, $|g(x)| \leq 1$ for all $x > 0$.

If we fix a constant $b > 0$, we have that: $|\frac{\sin(y/k)}{y/k} \cdot \frac{\sin(y)}{y} \cdot \chi_{[0,b]}(y)| \leq 1 \cdot 1 \cdot \chi_{[0,b]}(y)$ for all $k \in \mathbb{N}$. Hence by the dominated convergence theorem:

$$\liminf_{k\to\infty} \frac{4}{\pi} \int_0^\infty |\frac{\sin(y/k)}{y/k} \cdot \frac{\sin(y)}{y}| \mathrm{d}y \geq \lim_{k\to\infty} \frac{4}{\pi} \int_0^b |\frac{\sin(y/k)}{y/k} \cdot \frac{\sin(y)}{y}| \mathrm{d}y = \frac{4}{\pi} \int_0^b |\frac{\sin(y)}{y}| \mathrm{d}y$$

But now note that $\int_0^\infty |\frac{\sin(y)}{y}| \mathrm{d}y = \infty$.

$$\begin{aligned}
\int_0^\infty |\frac{\sin(y)}{y}| \mathrm{d}y \geq \sum_{n=0}^{\infty} \int_{n\pi + \frac{\pi}{6}}^{n\pi + \frac{5\pi}{6}} |\frac{\sin(y)}{y}| \mathrm{d}y &= \sum_{n=0}^{\infty} \int_{n\pi + \frac{\pi}{6}}^{n\pi + \frac{5\pi}{6}} \frac{1}{2y} \mathrm{d}y \\
&\geq \frac{1}{2} \sum_{n=0}^{\infty} \int_{n\pi + \frac{\pi}{6}}^{n\pi + \frac{5\pi}{6}} \frac{1}{n\pi + \frac{5\pi}{6}} \mathrm{d}y \\
&= \frac{\pi}{3} \sum_{n=0}^{\infty} \frac{1}{n\pi + \frac{5\pi}{6}} \geq \frac{\pi}{3} \sum_{n=1}^{\infty} \frac{1}{n\pi} = \frac{1}{3} \sum_{n=1}^{\infty} \frac{1}{n} = \infty
\end{aligned}$$

Thus, we can make $\int_0^b |\frac{\sin(y)}{y}| \mathrm{d}y$ arbitrarily big by making b big enough. Hence, we've proven that $\lim_{k\to\infty} \|f_k^\vee\|_1 = \lim_{k\to\infty} \frac{4}{\pi} \int_0^b |\frac{\sin(y)}{y}| \mathrm{d}y = \infty$.

Side note: while $\int_0^\infty \frac{\sin(y)}{y} \mathrm{d}y$ is not defined as a Lebesgue integral, it is defined as an improper Riemann integral and we can calculate that integral as follows.

Let $s > 0$. Then note that $\frac{\sin(y)}{y}$ and $e^{-sy}\chi_{[0,\infty)}$ are both in L^2. After all, $|\frac{\sin(y)}{y}|^2 \leq \chi_{[-1,1]}(y) + \frac{1}{y^2}\chi_{[-1,1]^{\mathsf{c}}}(y)$ and the right side is in L^2. Meanwhile, $\|e^{-sy}\|_2 = \frac{1}{2s}$. Thus by the Plancharel theorem, we know:

$$\int_0^\infty \frac{\sin(y)}{y} e^{-sy} \mathrm{d}y = \int_{-\infty}^\infty \mathcal{F}(\tfrac{\sin(y)}{y}) \overline{\mathcal{F}(e^{-sy}\chi_{[0,\infty)}(y))} \mathrm{d}y$$

Now since $\chi_{[-a,a]}^{\vee} = \frac{\sin(2\pi a x)}{\pi x}$ for any $a \geq 0$, we can see that: $\mathcal{F}(\frac{\sin(y)}{y}) = \pi\chi_{[-\frac{1}{2\pi},\frac{1}{2\pi}]}(\xi)$. Meanwhile:

$$\mathcal{F}(e^{-sy}\chi_{[0,\infty)}) = \int_0^\infty e^{-(s+2\pi i\xi)y}\mathrm{d}y = \frac{-1}{s+2\pi i\xi}(0-1) = \frac{1}{s+2\pi i\xi}$$

Hence, we've shown that:

$$\int_0^\infty \frac{\sin(y)}{y}e^{-sy}\mathrm{d}y = \pi\int_{-\frac{1}{2\pi}}^{\frac{1}{2\pi}} \overline{\left(\frac{1}{s+2\pi i\xi}\right)}\mathrm{d}\xi = \pi\int_{-\frac{1}{2\pi}}^{\frac{1}{2\pi}} \frac{s+2\pi i\xi}{s^2+4\pi^2\xi^2}\mathrm{d}\xi = \pi\int_{-\frac{1}{2\pi}}^{\frac{1}{2\pi}} \frac{s}{s^2+4\pi^2\xi^2}\mathrm{d}\xi$$

(Note, the last equality follows because we know that the imaginary part of the integral has to cancel since $\int_0^\infty \frac{\sin(y)}{y}e^{-sy}\mathrm{d}y$ is purely real-valued.)

Now:

$$\pi\int_{-\frac{1}{2\pi}}^{\frac{1}{2\pi}} \frac{s}{s^2+4\pi^2\xi^2}\mathrm{d}\xi = \frac{s\pi}{4\pi^2}\int_{-\frac{1}{2\pi}}^{\frac{1}{2\pi}} \frac{1}{(\frac{s}{2\pi})^2+\xi^2}\mathrm{d}\xi$$
$$= \frac{s}{4\pi}\left(\frac{2\pi}{s}\right)\left[\arctan(\tfrac{2\pi}{s}\xi)\right]_{\xi=-\frac{1}{2\pi}}^{\xi=\frac{1}{2\pi}} = \frac{1}{2}\left(\arctan(\tfrac{1}{s}) - \arctan(-\tfrac{1}{s})\right) = \arctan(\tfrac{1}{s})$$

Thus, we've proven that $\int_0^\infty \frac{\sin(y)}{y}e^{-sy}\mathrm{d}y = \arctan(\frac{1}{s})$ for all $s > 0$.

Taking the limit as $s \to 0$, we get that $\int_0^\infty \frac{\sin(y)}{y}e^{-sy}\mathrm{d}y \to \frac{\pi}{2}$. That said, some care is needed since $\int_0^\infty \frac{\sin(y)}{y}\mathrm{d}y$ and $\lim_{s\to 0}\int_0^\infty \frac{\sin(y)}{y}e^{-sy}\mathrm{d}y$ are defined differently. In fact, we still have not showed that the former which is equal to $\lim_{b\to\infty}\int_0^b \frac{\sin(y)}{y}\mathrm{d}y$ exists. So, let's do that now.

Note that for any $b \in (0,\infty)$, there are unique $n \in \mathbb{Z}_{\geq 0}$ and $\alpha \in [0,\pi)$ such that $b = n\pi + \alpha$. Then for all $s \geq 0$, we have that:

$$\int_0^b \frac{\sin(y)}{y}e^{-sy}\mathrm{d}y = \sum_{j=0}^{n-1}(-1)^j\int_{j\pi}^{(j+1)\pi}|\frac{\sin(y)}{y}e^{-sy}|\mathrm{d}y + \int_{n\pi}^{n\pi+\alpha}|\frac{\sin(y)}{y}e^{-sy}|\mathrm{d}y$$

Now, the leftover term will approach 0 as $b \to \infty$ since it is at most $\frac{\alpha}{n\pi}$ when $b \geq 1$ and $n \to \infty$ as $b \to \infty$. Hence, letting $c_n = \int_{j\pi}^{(j+1)\pi}|\frac{\sin(y)}{y}e^{-sy}|\mathrm{d}y$, we know that: $\lim_{b\to\infty}\int_0^b \frac{\sin(y)}{y}e^{-sy}\mathrm{d}y = \sum_{n=0}^\infty(-1)^n c_n$. It's easily verified using the alternating series test that the series converges. This proves that our improper Riemann integral exists for all $s \geq 0$ (including $s = 0$).

Importantly, this series also converges uniformly over all $s \in [0,\infty)$. To see why, observe that since $(c_n)_{n\in\mathbb{N}}$ is a strictly decreasing sequence, for any $N \geq 0$ we have that: $c_N \geq |\sum_{n=N}^\infty(-1)^n a_n|$. This can be proven via induction fairly easily. Next, making s larger makes all the c_n strictly smaller. So, by picking N large enough so that $c_N < \varepsilon$ when $s = 0$, we can guarentee that $c_N < \varepsilon$ for all s. It then follows that the error from the limit point: $|\sum_{n=N}^\infty(-1)^n c_n|$, is also less than ε for all s.

With that, we know there is some $b > 0$ such that:

$$\left|\int_0^\infty \frac{\sin(y)}{y}e^{-sy}\mathrm{d}y - \int_0^b \frac{\sin(y)}{y}e^{-sy}\mathrm{d}y\right| < {}^{\varepsilon}/_{4} \text{ for all } s \geq 0.$$

Also, by dominated convergence theorem (its 4am and I don't want to type out verifications for all the conditions), we know that $\int_0^b \frac{\sin(y)}{y} e^{-sy} \mathrm{d}y \to \int_0^b \frac{\sin(y)}{y} \mathrm{d}y$ as $s \to 0$. So, there is some $s > 0$ such that:

$$\left| \int_0^b \frac{\sin(y)}{y} \mathrm{d}y - \int_0^b \frac{\sin(y)}{y} e^{-sy} \mathrm{d}y \right| < \varepsilon/4.$$

Also, by making s potentially smaller, we can also guarentee that:

$$\left| \frac{\pi}{2} - \int_0^\infty \frac{\sin(y)}{y} e^{-sy} \mathrm{d}y \right| < \varepsilon/4.$$

And chaining those together, we get that:

$$\begin{aligned} \left| \int_0^\infty \frac{\sin(y)}{y} \mathrm{d}y - \frac{\pi}{2} \right| &\leq \left| \int_0^\infty \frac{\sin(y)}{y} \mathrm{d}y - \int_0^b \frac{\sin(y)}{y} \mathrm{d}y \right| + \left| \int_0^b \frac{\sin(y)}{y} \mathrm{d}y - \int_0^b \frac{\sin(y)}{y} e^{-sy} \mathrm{d}y \right| \\ &\quad + \left| \int_0^b \frac{\sin(y)}{y} e^{-sy} \mathrm{d}y - \int_0^\infty \frac{\sin(y)}{y} e^{-sy} \mathrm{d}y \right| + \left| \int_0^\infty \frac{\sin(y)}{y} e^{-sy} - \frac{\pi}{2} \right| \\ &< \varepsilon/4 + \varepsilon/4 + \varepsilon/4 + \varepsilon/4 = \varepsilon \end{aligned}$$

Taking $\varepsilon \to 0$, we've finally shown that $\lim\limits_{b \to \infty} \int_0^b \frac{\sin(y)}{y} \mathrm{d}y = \frac{\pi}{2}$.

(c) Prove that $\mathcal{F}(L^1)$ is a proper subset of C_0.

To start off with, recall that if $f \in L^1$ and $\widehat{f} = 0$, then $f = 0$ a.e. As a corollary to this, we have that if $f, g \in L^1$ and $\widehat{f} = \widehat{g}$, then $f = g$ a.e. This is because $(f - g)^\wedge = 0$ implies that $f - g = 0$ a.e. So, we know that $\mathcal{F}$ is an injective map from L^1 to C_0. If $\mathcal{F}$ was also surjective, then we would know that $\mathcal{F}$ is a bijection, and that therefore a function $\mathcal{F}^{-1} : C_0 \to L^1$ exists. Also, by the open map theorem, we would know that $\mathcal{F}^{-1}$ is bounded.

However, in part (b) we found that $\|f_k\|_u = 2$ for all $k \in \mathbb{N}$ but $\|f_k^\vee\|_1 \to \infty$ as $k \to \infty$. Importantly, we can see from our work earlier that $f_k^\wedge = f_k^\vee$ and $\|f_k^\vee\|_1 < \infty$ for all k. After all, $f_k^\vee(y)$ is bounded by 1 when $|y| \leq 1$ and by $\frac{k}{y^2}$ when $|y| \geq 1$. So by the Fourier inversion theorem, we know that $(f_k^\vee)^\wedge = f_k$ (with equality holding everywhere since both sides are continuous). And so, $\mathcal{F}^{-1}(f_k) = f_k^\vee$.

This proves that $\mathcal{F}^{-1}$ is not bounded since $\|\mathcal{F}^{-1}(f_k)\|_1$ can be made arbitrarily large even while $\|f_k\|_u = 2$ for all k.

7/10/2025

Today I'm gonna do more problems from chapter 8 of Folland.

Recall that for $f \in L^p(\mathbb{R})$, if there exists $h \in L^p(\mathbb{R})$ such that $\lim_{y \to 0} \|y^{-1}(\tau_{-y} f - f) - h\|_p = 0$, we call h the (strong) L^p derivative of f. If $f \in L^p(\mathbb{R}^n)$, L^p partial derivatives of f are defined similarly. (Also, the notation $\tau_y f(x)$ refers to $f(x - y)$.)

Exercise 8.8: Suppose that p and q are conjugate exponents, $f \in L^p$, $g \in L^q$, and the L^p derivative $\partial_j f$ exists. Then $\partial_j(f * g)$ exists (in the ordinary sense) and equals $(\partial_j f) * g$.

Note that:

$\lim\limits_{t \to 0} \frac{(f*g)(x+te_j) - (f*g)(x)}{t} = ((\partial_j f) * g)(x)$ iff $\lim\limits_{t \to 0} \left| \frac{(f*g)(x+te_j) - (f*g)(x)}{t} - ((\partial_j f) * g)(x) \right| = 0.$

Now for all $t \neq 0$:

$$0 \leq \left|\frac{(f*g)(x+te_j)-(f*g)(x)}{t} - ((\partial_j f) * g)(x)\right|$$
$$= \left|t^{-1}\int (f(x+te_j-y) - f(x-y))\, g(y)\mathrm{d}y - \int \partial_j f(x-y)g(y)\mathrm{d}y\right|$$
$$= \left|\int t^{-1}(\tau_{-te_j} f - f)(x-y)g(y)\mathrm{d}y - \int \partial_j f(x-y)g(y)\mathrm{d}y\right|$$
$$\leq \int |t^{-1}(\tau_{-te_j} f - f)(x-y) - \partial_j f(x-y)||g(y)|\mathrm{d}y$$
$$\leq \|t^{-1}(\tau_{-te_j} f - f) - \partial_j f\|_p \|g\|_q$$

Since $\|g\|_q$ is fixed and $\|t^{-1}(\tau_{-te_j} f - f) - \partial_j f\|_p \to 0$ as $t \to 0$, we've thus shown that $\left|\frac{(f*g)(x+te_j)-(f*g)(x)}{t} - ((\partial_j f) * g)(x)\right| \to 0$ as $t \to 0$.

Exercise 8.9: Let $1 \leq p < \infty$. If $f \in L^p(\mathbb{R})$, the L^p derivative of f (call it h; see Exercise 8) exists iff f is absolutely continuous on every bounded interval (perhaps after modification on a null set) and its pointwise derivative f' is in L^p, in which case $h = f'$ a.e.

($\Longrightarrow$)

Suppose f has an L^p derivative h. Then setting $\varphi(x) = (1-|x|)\chi_{[-1,1]}$, note that $\int \varphi(x)\mathrm{d}x = 1$ and $0 \leq \varphi \leq 1 \leq \frac{4}{(1+|x|)^2}$. Thus, φ satisfies the hypothesis of theorem 8.15 (see page 31 of my paper notes) and so we know that:

- $(f * \varphi_{1/n})(x) = \int f(x-y) \cdot n\varphi(ny)\mathrm{d}y \to f(x)$ as $n \to \infty$ for all $x \in L_f$,
- $(h * \varphi_{1/n})(x) = \int h(x-y) \cdot n\varphi(ny)\mathrm{d}y \to h(x)$ as $n \to \infty$ for all $x \in L_h$

(where L_f and L_h are the Lebesgue sets of f and h respectively).

Side note: If $g \in L^1_{\text{loc}}$, then we know that $(L_g)^{\mathrm{C}}$ has measure zero. Now while it's obvious that $L^1, L^\infty \subseteq L^1_{\text{loc}}$, I'm currently realizing I've never justified to myself why $L^p \subseteq L^1_{\text{loc}}$ for all $1 < p < \infty$.

If E is a measurable set with finite measure, then the measure restricted to E does not have any sets of arbitrarily large measure. Thus for any $p, q \in (0, \infty)$ with $p < q$, we have that $L^q(E) \subseteq L^p(E)$. Specifically, this means that for any $1 < p < \infty$, $L^p(E) \subseteq L^1(E)$. It follows that $L^p \subseteq L^1_{\text{loc}}$ for all $1 < p < \infty$.

Also, if q is the conjugate exponent of p, we know that $\phi_{1/n} \in L^q$. Therefore, by the previous exercise we know that $(f * \phi_{1/n})' = h * \varphi_{1/n}$. Additionally:

$$|h * \varphi_{1/n}(x)| \leq \int |h(x-y)\varphi_{1/n}(y)|\mathrm{d}y \leq \|h\|_p \left(\int_{-1/n}^{1/n} |n(1-|nx|)|^q \mathrm{d}x\right)^{1/q}$$
$$\leq n\|h\|_p \left(\int_{-1/n}^{1/n} 1\mathrm{d}x\right)^{1/q} = n\left(\tfrac{2}{n}\right)^{1/q} \|h\|_p \leq 2^{1/q}\|h\|_p$$

This tells us that for all $n \in \mathbb{N}$, $f * \varphi_{1/n}$ has a bounded derivative. It then follows by the mean value theorem that $f * \varphi_{1/n}$ is absolutely continuous. So for any $a, x \in \mathbb{R}$ with $a < x$, we have that:

$$(f * \varphi_{1/n})(x) - (f * \varphi_{1/n})(a) = \int_a^x (f * \varphi_{1/n})'(y)\mathrm{d}y = \int_a^x (h * \varphi_{1/n})(y)\mathrm{d}y$$

And, since $h * \varphi_{1/n} \to h$ pointwise a.e. and $2^{1/q}\|h\|_p \chi_{[a,x]} \in L^1$, we know by dominated convergence theorem that:

$$\lim_{n\to\infty} ((f * \varphi_{1/n})(x) - (f * \varphi_{1/n})(a)) = \lim_{n\to\infty} \int_a^x (h * \varphi_{1/n})(y)\mathrm{d}y = \int_a^x h(y)\mathrm{d}y$$

Now, we're finally ready to show the right hand side of our implication. Suppose $a, b \in L_f$ are fixed with $a < b$. Then for any $x \in L_f \cap [a, b]$, we have that:

$$f(x) - f(a) = \lim_{n\to\infty} ((f * \varphi_{1/n})(x)) - \lim_{n\to\infty} ((f * \varphi_{1/n})(a)) = \int_a^x h(y)\mathrm{d}y$$

By redefining f on the null space $(L_f)^{\mathsf{C}} \cap [a, b]$, we can thus guarentee that $f(x) - f(a) = \int_a^x h(y)\mathrm{d}y$ for all $x \in [a, b]$. In turn, by the fundamental theorem of calculus we know f is absolutely continuous on $[a, b]$ and that $h = f'$ a.e. on $[a, b]$.

If $I \subseteq \mathbb{R}$ is any arbitrary bounded interval, then we can still apply the former reasoning by finding $a, b \in L_f$ such that $I \subseteq [a, b]$. Then f being absolutely continuous on $[a, b]$ implies that f is absolutely continuous I. Also, since $\mathbb{R}$ can be completed covered by these intervals, we know that $f' = h$ a.e. The only snag we still have to sort out is to show that our redefinitions of $f(x)$ for $x \in (L_f)^{\mathsf{C}}$ are well defined (i.e. not dependent on our choice of $a, b \in L^f$.)

Suppose $a_1, a_2 \in L_f$ and without loss of generality assume $a_1 < a_2 < x$. Then:

$$\left(\int_{a_1}^x h(y)\mathrm{d}y + f(a_1)\right) - \left(\int_{a_2}^x h(y)\mathrm{d}y + f(a_2)\right) = \int_{a_1}^{a_2} h(y)\mathrm{d}y - (f(a_2) - f(a_1)).$$

Since $a_1, a_2 \in L^f$, we know that $\int_{a_1}^{a_2} h(y)\mathrm{d}y = f(a_2) - f(a_1)$. So our above expression equals 0 and we've shown that:

$$f(x) = \int_{a_1}^x h(y)\mathrm{d}y + f(a_1) = \int_{a_2}^x h(y)\mathrm{d}y + f(a_2) \text{ is well defined.}$$

($\Longleftarrow$)

Note that if $y > 0$, then our assumptions about f tell us that:

$$\begin{aligned}\tfrac{f(x+y)-f(x)}{y} - f'(x) = \tfrac{1}{y}\int_x^{x+y} f'(t)\mathrm{d}t - f'(x) &= \tfrac{1}{y}\int_x^{x+y} f'(t) - f'(x)\mathrm{d}t \\ &= \tfrac{1}{y}\int_0^y f'(x+t) - f'(x)\mathrm{d}t\end{aligned}$$

Similarly, if $y < 0$, then we know:

$$\begin{aligned}\tfrac{f(x+y)-f(x)}{y} - f'(x) = \tfrac{-1}{y}\int_{x+y}^{x} f'(t)\mathrm{d}t - f'(x) &= \tfrac{-1}{y}\int_{x+y}^{x} f'(t) - f'(x)\mathrm{d}t \\ &= \tfrac{-1}{y}\int_y^0 f'(x+t) - f'(x)\mathrm{d}t\end{aligned}$$

In either case, we can see that:

$$\left|\tfrac{f(x+y)-f(x)}{y} - f'(x)\right| \leq \int_{-|y|}^{|y|} \tfrac{1}{|y|}|\tau_{-t}f'(x) - f'(x)|\mathrm{d}t$$

Thus by Minkowski's inequality for integrals:

$$\left\|\tfrac{f(x+y)-f(x)}{y} - f'(x)\right\|_p \leq \left\|\int_{-|y|}^{|y|} \tfrac{1}{|y|}|\tau_{-t}f'(x) - f'(x)|\mathrm{d}t\right\|_p \leq \tfrac{1}{|y|}\int_{-|y|}^{|y|} \|\tau_{-t}f'(x) - f'(x)\|_p\mathrm{d}t$$

And since translation is continuous with respect to the L^p norm for $1 \leq p < \infty$, we know that $\|\tau_{-t}f'(x) - f'(x)\|_p \to 0$ as $t \to 0$. Hence given $\varepsilon > 0$, we have for $|y|$ small enough that:

$$\tfrac{1}{|y|}\int_{-|y|}^{|y|} \|\tau_{-t}f'(x) - f'(x)\|_p\mathrm{d}t < \tfrac{1}{|y|}\int_{-|y|}^{|y|} \varepsilon\mathrm{d}t = \tfrac{2|y|\varepsilon}{|y|} = 2\varepsilon$$

By taking $\varepsilon \to 0$, this proves that $\left\|\frac{f(x+y)-f(x)}{y} - f'(x)\right\|_p \to 0$ as $y \to 0$. Hence f' is an L^p derivative of f. ■

So what's the significance of this result?

- A function on $\mathbb{R}$ having an L^p derivative is a strictly stronger assumption than the function just being differentiable almost everywhere.
- Any two L^p derivatives of a function are equal a.e. to the ordinary derivative of the function. Thus there's at most one L^p derivative of any function in $L^p(\mathbb{R})$.
- Any function $L^p(\mathbb{R})$ that is differentiable a.e. and whose derivative is bounded and also in L^p has an L^p derivative.

7/11/2025

Exercise 8.18: Suppose $f \in L^2(\mathbb{R})$.

(a) The L^2 derivative f' exists iff $\xi\widehat{f} \in L^2$, in which case $\widehat{f'}(\xi) = 2\pi i\xi\widehat{f}(\xi)$.

($\Longrightarrow$)
Once again set $\varphi(x) = (1-|x|)\chi_{[-1,1]}$. Then by theorem 8.14(a) (see page 29 of my paper notes): $f * \varphi_{1/n} \to f$ in L^2 as $n \to \infty$. In turn, since the Fourier transform is continuous on L^2, we know that $(f * \varphi_{1/n})^\wedge \to \widehat{f}$ as $n \to \infty$.

Next, note that $f * \varphi_{1/n} \in C^1$.

Why: Recall from exercise 8.8 that $(f * \varphi_{1/n})' = f' * \varphi_{1/n}$. Also, since $f' \in L^1_{\text{loc}}$ and $\varphi_{1/n} \in C^0$ has compact support, we know from exercise 8.7 (which was a homework problem in Math 240C), that $f' * \varphi_{1/n} \in C^0$.

Also, since f, f' and $\varphi_{1/n}$ are all in L^2, we know by proposition 8.8 (see page 25 of my paper notes) that $f * \varphi_{1/n} \in C_0$, and we know by Young's inequality (see page 26 of my paper notes) that $f * \varphi_{1/n}, f' * \varphi_{1/n} \in L^1$. All together, this lets us conclude via integration by parts that:

$$(f * \varphi_{1/n})^\wedge = \tfrac{1}{2\pi i\xi}((f * \varphi_{1/n})')^\wedge = \tfrac{1}{2\pi i\xi}(f' * \varphi_{1/n})^\wedge.$$

Finally, since $f' * \varphi_{1/n} \to f'$ in L^2 as $n \to \infty$ and the Fourier transform is continuous on L^2, we know that:

$$\widehat{f}(\xi) = \lim_{n\to\infty}(f * \varphi_{1/n})^\wedge(\xi) = \tfrac{1}{2\pi i\xi}\lim_{n\to\infty}(f' * \varphi_{1/n})^\wedge(\xi) = \tfrac{1}{2\pi i\xi}\widehat{f'}(\xi) \text{ a.e.}$$

Since $\frac{1}{2\pi i}\widehat{f'}$ is in L^2, this thus proves that $\xi\widehat{f} \in L^2$. Also, by rearranging out expression we get that $\widehat{f'} = 2\pi i\xi\widehat{f}(\xi)$.

($\Longleftarrow$)
Define $h(\xi) = 2\pi i\xi\widehat{f}(\xi)$. Then by assumption we know that $h \in L^2$. So, there exists a function $H \in L^2$ such that $\widehat{H} = h$. And since the Fourier transform is a continuous isometric linear operator on L^2, we know that for all $y \neq 0$:

$$\begin{aligned}\|\tfrac{1}{y}(\tau_{-y}f - f) - H\|_2 &= \|\mathcal{F}(\tfrac{1}{y}(\tau_{-y}f - f) - H)\|_2 \\ &= \|\tfrac{1}{y}(\mathcal{F}(\tau_{-y}f) - \mathcal{F}(f)) - h\|_2\end{aligned}$$

Now we claim that $\mathcal{F}(\tau_{-y}f) = e^{2\pi i\xi y}\mathcal{F}(f)$ for all $f \in L^2$.

Proof:
Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of Schwartz functions converging to $f \in L^2$. Then since $\int g = \int \tau_{-y}g$ for all functions g, we can easily see that $\tau_{-y}f_n \to \tau_{-y}f$ in L^2. Therefore, $\mathcal{F}(\tau_{-y}f) = \lim\limits_{n\to\infty} \mathcal{F}(\tau_{-y}f_n)$.

Next, since $f_n \in L^1$ for all n, we know that $\mathcal{F}(\tau_{-y}f_n)(\xi) = e^{2\pi i\xi y}\widehat{f_n}(\xi)$. Then finally, since the Fourier transform is continuous on L^2, we have that $\widehat{f_n} \to \widehat{f}$ in L^2 as $n \to \infty$. By passing to a subsequence, we can assume $\widehat{f_n} \to \widehat{f}$ pointwise a.e. And so, $\mathcal{F}(\tau_{-y}f) = \lim\limits_{n\to\infty} e^{2\pi i\xi y}\widehat{f_n}(\xi) = e^{2\pi i\xi y}\widehat{f}(\xi)$ a.e.

Thus, we know that:

$$\begin{aligned}\left|\tfrac{1}{y}(\mathcal{F}(\tau_{-y}f) - \mathcal{F}(f)) - h\right|^2 &= \left|\left(\tfrac{1}{y}e^{2\pi i\xi y} - \tfrac{1}{y} - 2\pi i\xi\right)\widehat{f}(\xi)\right|^2 \\ &= \left|\left((\tfrac{\cos(2\pi\xi y)}{y} - \tfrac{1}{y}) + i(\tfrac{\sin(2\pi\xi y)}{y} - 2\pi\xi)\right)\widehat{f}(\xi)\right|^2 \\ &= \left|\left((\tfrac{\cos(2\pi\xi y)-1}{\xi y}) + i(\tfrac{\sin(2\pi\xi y)}{y\xi} - 2\pi)\right)\xi\widehat{f}(\xi)\right|^2\end{aligned}$$

Now, note that $\lim_{y\to 0}\frac{\cos(2\pi\xi y)-1}{\xi y} = 2\pi\lim_{t\to 0}\frac{\cos(t)-1}{t} = 2\pi\cdot 0 = 0$ for all $\xi \neq 0$. Similarly, we have that $\lim_{y\to 0}\frac{\sin(2\pi\xi y)}{y\xi} = 2\pi\lim_{t\to 0}\frac{\sin(t)}{t} = 2\pi\cdot 1$ for all $\xi \neq 0$. This proves that $|\frac{1}{y}(\mathcal{F}(\tau_{-y}f) - \mathcal{F}(f)) - h|^2 \to 0$ pointwise a.e. as $y \to 0$.

Meanwhile, note that $\left|\frac{\sin(2\pi x)}{x}\right|$ and $\left|\frac{\cos(2\pi x)-1}{x}\right|$ are both less than or equal to 2π on their domains. Therefore, we can get that for all $\xi \neq 0$ and $y \neq 0$, we have that:

$$|(\tfrac{\cos(2\pi\xi y)-1}{\xi y}) + i(\tfrac{\sin(2\pi\xi y)}{y\xi} - 2\pi)| \leq |(\tfrac{\cos(2\pi\xi y)-1}{\xi y})| + |(\tfrac{\sin(2\pi\xi y)}{y\xi} - 2\pi)| \leq 6\pi$$

(I'm not sure how to prove $|\frac{\cos(2\pi x)-1}{x}| \leq 2\pi$ without pulling out numerical methods. But you'll see that it is true if you graph it.)

Thus using $36\pi^2|\xi\widehat{f}(\xi)|^2$ as our upper bound function (which is in L^1 since $\xi\widehat{f} \in L^2$), we can conclude via the dominated convergence theorem that:

$$\lim_{y\to 0}\|\tfrac{1}{y}(\mathcal{F}(\tau_{-y}f) - \mathcal{F}(f)) - h\|_2^2 = \lim_{y\to 0}\int\left|\tfrac{1}{y}(\mathcal{F}(\tau_{-y}f) - \mathcal{F}(f)) - h\right|^2 = 0$$

So, f has $H = h^\vee$ as it's L^2 derivative.

7/12/2025

Ok. I think that in order to prove part (b) of exercise 8.18, I need to make a pit stop in the exercises of section 3.5 of Folland. This is because Folland's hinted solution route is to use integration by parts. However, right now I've only shown that you can do integration by parts if the two functions in your integrand are continuously differentiable. Yet that's not guarenteeable in exercise 8.18(b). So, my current objective is to weaken my requirements for doing integration by parts.

Exercise 3.35: If F and G are absolutely continuous on $[a,b]$, then so is FG and:

$$\int_a^b (FG' + GF')(x)\mathrm{d}x = F(b)G(b) - F(a)G(a)$$

Proof:
By extreme value theorem, there exists $M \geq 0$ such that $\max(|F(x)|, |G(x)|) \leq M$ for all $x \in [a,b]$. Now for any $\varepsilon > 0$, let $\delta > 0$ be such that for all finite collections of disjoint intervals $(a_1, b_1), \ldots, (a_n, b_n) \subseteq [a,b]$ with $\sum_{i=1}^n |b_i - a_i| < \delta$, we have:

$$\sum_{i=1}^n |F(b_i) - F(a_i)| < \tfrac{\varepsilon}{2M} \text{ and } \sum_{i=1}^n |G(b_i) - G(a_i)| < \tfrac{\varepsilon}{2M}$$

Then we have:

$$\begin{aligned}
\sum_{i=1}^n |F(b_i)G(b_i) - F(a_i)G(a_i)| &\leq \sum_{i=1}^n |F(b_i)G(b_i) - F(b_i)G(a_i)| + \sum_{i=1}^n |F(b_i)G(a_i) - F(a_i)G(a_i)| \\
&= \sum_{i=1}^n |F(b_i)||G(b_i) - G(a_i)| + \sum_{i=1}^n |G(a_i)||F(b_i) - F(a_i)| \\
&\leq M\sum_{i=1}^n |G(b_i) - G(a_i)| + M\sum_{i=1}^n |F(b_i) - F(a_i)| \\
&< M\tfrac{\varepsilon}{2M} + M\tfrac{\varepsilon}{2M} = \varepsilon
\end{aligned}$$

Hence, FG is also absolutely continuous on $[a,b]$. It follows by the fundamental theorem of calculus for Lebesgue integrals that:

$$F(b)G(b) - G(b)G(a) = \int_a^b (FG)'(x)\mathrm{d}x = \int_a^b (FG' + GF')(x)\mathrm{d}x. \blacksquare$$

The following is also tangentially relevant to exercise 8.18(b) in addition to being interesting in its own right. A function $f : \mathbb{R} \to \mathbb{C}$ is singular if f is continuous everywhere, f' exists a.e. with $f' = 0$ a.e., and f is not a constant function.

Recalling page 53 of this journal, it's easy to see that the Cantor function is singular (if you continuously extend it to be constant outside $[0,1]$). However, it's also constant on a small enough neighborhood around almost every point. Hence, it doesn't defy our intution too overly much. In the next two exercises however, we'll construct a strictly increasing singular function.

Exercise 3.39: If $(F_k)_{k\in\mathbb{N}}$ is a sequence of nonnegative increasing functions on $[a,b]$ such that $F(x) = \sum_{k=1}^\infty F_k(x) < \infty$ for all $x \in [a,b]$, then $F'(x) = \sum_{k=1}^\infty F_k'(x)$ for a.e. $x \in [a,b]$.

For all k, define:

$$G_k(x) = \begin{cases} F_k(a+) & \text{if } x \leq a \\ F_k(x+) & \text{if } x \in [a,b) \\ F_k(b) & \text{if } x \geq b \end{cases} \quad \text{and} \quad G(x) = \begin{cases} F(a+) & \text{if } x \leq a \\ F(x+) & \text{if } x \in [a,b) \\ F(b) & \text{if } x \geq b \end{cases}$$

Then by theorem 3.23, we know that $G_k' = F_k'$; $G' = F'$ a.e. on $[a,b]$ Also, each G_k is a nonnegative monotone increasing function with $G(x) = \sum_{k=1}^{\infty} G_k(x) < \infty$ for all $x \in \mathbb{R}$.

Why: For any $x \in [a,b)$, we can apply the dominated convergence theorem to $l^1(\mathbb{N})$ using the upper bound $F(b) = \sum_{k=1}^{\infty} F_k(b)$ in order to get that:

$$G(x) = F(x+) = \lim_{t \to 0^+} \sum_{k=1}^{\infty} F_k(x+t) = \sum_{k=1}^{\infty} \lim_{t \to 0^+} F_k(x+t) = \sum_{k=1}^{\infty} F_k(x+) = \sum_{k=1}^{\infty} G_k(x)$$

Taking things one step further, define $H_k(x) = G_k(x) - G_k(a)$ and $H(x) = G(x) - G(a)$. Since adding by a constant doesn't change the derivative of a function at all, we still know that $H_k' = F_k'$; $H' = F'$ a.e. on $[a,b]$ Also, since G_k and G are monotone increasing, we know that $G_k(x) \geq G_k(a)$ and $G(x) \geq G(a)$ for all x. Hence, all of our H_k and H are still nonnegative monotone increasing functions on $\mathbb{R}$. And clearly $H(x) = \sum_{k=1}^{\infty} H_k(x)$ for all $x \in \mathbb{R}$.

The significance of this is that if we now prove that $H'(x) = \sum_{k=1}^{\infty} H_k'(x)$ for a.e. $x \in [a,b]$, then we will have also shown that $F'(x) = \sum_{k=1}^{\infty} F_k'(x)$ for a.e. $x \in [a,b]$. But importantly, all our H_k and H are in NBV. After all, they are in BV because they are bounded and monotone increasing. Also, they are all right continuous and $H_k(-\infty) = 0 = H(-\infty)$. It then follows that there are unique finite Borel measures μ_{H_k} and μ_H such that $\mu_{H_k}((-\infty, x]) = H_k(x)$ and $\mu_H((-\infty, x]) = H(x)$.

Now let $\mathrm{d}\mu_{H_k} = \mathrm{d}\lambda_k + f_k \mathrm{d}m$ and $\mathrm{d}\mu_H = \mathrm{d}\lambda + f \mathrm{d}m$ be the Radon-Nikodym representations of μ_{H_k} and μ_H with respect to the Lebesgue measure. Then since μ_H and μ_{H_k} are both finite measures in the separable locally compact Hausdorff space $\mathbb{R}$, we know by theorem 7.8 that μ_{H_k} and μ_H are regular. Also, if we let $E_r(x) = (x, x+r]$ for all $r > 0$ and $x \in \mathbb{R}$, then we know that E_r shrinks nicely to x for all x. Therefore, by the generalized Lebesgue differentiation theorem, we have:

$$H'(x) = \lim_{h \to 0^+} \frac{H(x+h) - H(x)}{h} = \lim_{h \to 0^+} \frac{\mu_H((x, x+h])}{h} = \lim_{h \to 0^+} \frac{\mu_H(E_h(x))}{m(E_h(x))} = f(x) \text{ for } m\text{-a.e. } x.$$

(And similarly we have that $H_k'(x) = f_k(x)$ for m-a.e. x.)

Next note that for any $(a,b) \subseteq \mathbb{R}$, we have that:

$$\mu_H((a,b)) = \lim_{\beta \to b-} \mu_H((a,\beta]) = \lim_{\beta \to b-} (H(\beta) - H(a)) = \lim_{\beta \to b-} \sum_{k=1}^{\infty} (H_k(\beta) - H_k(a)).$$

Once again, $H_k(\beta) - H_k(a) \geq 0$ for all k and our series is bounded from above by $\sum_{k=1}^{\infty} (H_k(b) - H_k(a)) = \mu_{H_k}((a,b]) < \infty$. So, by applying the dominated convergence theorem we get that:

$$\sum_{k=1}^{\infty} (H_k(\beta) - H_k(a)) = \sum_{k=1}^{\infty} \lim_{\beta \to b-} (H_k(\beta) - H_k(a)) = \sum_{k=1}^{\infty} \lim_{\beta \to b-} \mu_{H_k}((a,\beta]) = \sum_{k=1}^{\infty} \mu_{H_k}((a,b)).$$

This in turn proves that $\mu_H = \sum_{k=1}^{\infty} \mu_{H_k}$ on all open sets by the countable additivity of measures, and all measurable sets in general by outer regularity. Hence, we know that $\mathrm{d}\lambda + H' \mathrm{d}m = \sum_{k=1}^{\infty} (\mathrm{d}\lambda_k + H_k' \mathrm{d}m) = \sum_{k=1}^{\infty} \mathrm{d}\lambda_k + \sum_{k=1}^{\infty} H_k' \mathrm{d}m$.

Lastly, note that $\sum_{k=1}^{\infty} \lambda_k \perp m$ and $\sum_{k=1}^{\infty} H_k' \mathrm{d}m = \left(\sum_{k=1}^{\infty} H_k'\right) \mathrm{d}m \ll m$. By the Radon-Nikodym theorem, we thus know that $\lambda = \sum_{k=1}^{\infty} \lambda_k$ and $H' \mathrm{d}m = \left(\sum_{k=1}^{\infty} H_k'\right) \mathrm{d}m$ with $H' = \sum_{k=1}^{\infty} H_k'$ m-a.e. ■

Exercise 3.40: Let F denote the Cantor function on $[0,1]$ and set $F(x) = 0$ for $x > 1$ and $F(x) = 1$ for $x > 1$. Let $\{[a_n, b_n]\}_{n\in\mathbb{N}}$ be an enumeration of the closed subintervals of $[0,1]$ with distinct rational endpoints, and let $F_n(x) = F(\frac{x-a_n}{b_n-a_n})$. Then $G = \sum_{n=1}^{\infty} 2^{-n} F_n$ is continuous and strictly increasing on $[0,1]$, and $G' = 0$ a.e.

Since $\frac{x-a_n}{b_n-a_n}$ is continuous and increasing, we know that each F_n is still continuous and monotone increasing. Also, we clearly have that if $x \geq b_n$, then $F_n(x) = 1$. Meanwhile, if $x \leq a_n$, then $F_n(x) = 0$. Thus, it's easy to see that:

- $G(x) = 0$ for $x \leq 0$ and $G(x) = 1$ for $x \geq 1$
- G is monotone increasing
- $\sum_{n=1}^{\infty} 2^{-n} F_n$ converges uniformly to G, thus making G continuous.
- By an easy application of exercise 3.39, $G' = 0$ a.e. since F' being zero almost everywhere implies $(2^{-n}F_n)' = 0$ a.e. for each n.

Reminder, for any x not in the Cantor set, we know either $x \notin [0,1]$ or there is an open interval containing x that was removed to form the Cantor set. In either scenario, we have that f is constant on a neighborhood of x. So, $f'(x) = 0$.

Finally, to show that G is strictly increasing on $[0,1]$, note that for any $x, y \in [0,1]$ with $x < y$, we know there is a closed subinterval $[a_n, b_n]$ with $x < a_n < b_n < y$. In turn, we know that $F_n(x) = 0$ while $F_n(y) = 1$. Then since $F_n(y) \geq F_n(x)$ for all other n, we know that $G(x)$ is strictly less than $G(y)$.

Note: We can also fairly easily see now that $\sum_{n\in\mathbb{Z}}(n + G(x-n))\chi_{[n,n+1]}$ is strictly increasing and continuous everywhere with a derivative equal to zero almost everywhere.

This poses a challenge because in exercise 8.18(b), we're going to need to be able say that a function having a derivative of zero almost everywhere implies that the function is constant. So here is one more lemma.

<u>Lemma:</u> If $f : \mathbb{R} \to \mathbb{C}$ is absolutely continuous on $[a,b]$ and $f' = 0$ a.e. on $[a,b]$, then f is constant on $[a,b]$.

Why: By the fundamental theorem of calculus for Lebesgue integrals, we know that if $x \in [a,b]$, then $f(x) - f(a) = \int_a^x f'(t)\mathrm{d}t = \int_a^x 0\mathrm{d}t = 0$. So, $f(x) = f(a)$.

7/13/2025

Exercise 8.18 (continued):

(b) If the L^2 derivative f' exists, then: $[\int |f(x)|^2 \mathrm{d}x]^2 \leq 4 \int |xf(x)|^2 \mathrm{d}x \int |f'(x)|^2 \mathrm{d}x$.

To start out, we need to make sure this inequality is well defined. Note that since $f, f' \in L^2$, we know that $\int |f(x)|^2 \mathrm{d}x < \infty$ and $\int |f'(x)|^2 \mathrm{d}x < \infty$. So, to guarentee that this inequality is well defined, we just need to show that if $\int |f'(x)|^2 \mathrm{d}x = 0$, then we will never have that $\int |xf(x)|^2 \mathrm{d}x = \infty$ (thus making the right-hand side $4(\infty \cdot 0)$). Luckily, by exercise 8.9, we know that f having an L^2 derivative means that f is absolutely continuous on every bounded interval. So by the lemma I ended yesterday with, we know that if $\int |f'(x)|^2 \mathrm{d}x = 0$, then f must be constant on every bounded interval since the ordinary derivative of f is zero almost everywhere. This proves that $f = c$ where c is some constant. But since $f \in L^2$, we must have that $\int_{-\infty}^{\infty} |c|^2 \mathrm{d}x < \infty$. The only way this is possible is if $c = 0$. So, $f = 0$ a.e. and we've thus shown that $\int |xf(x)|^2 \mathrm{d}x = 0$ as well.

Next, for any $a < b$ note that $|f|^2$ is absolutely continuous on $[a, b]$. This is because as mentioned before, f is absolutely continuous on $[a, b]$. Then in turn, it is easy to see that $\overline{f}$ is absolutely continuous on $[a, b]$. So, by exercise 3.35, we know that $f\overline{f} = |f|^2$ is absolutely continuous on $[a, b]$. Also $g(x) = x$ is absolutely continuous on $[a, b]$. Thus by exercise 3.35, we know that:

$$\int_a^b (1|f(x)|^2 + x\tfrac{\mathrm{d}}{\mathrm{d}x}|f(x)|^2) = b|f(b)|^2 - a|f(a)|^2$$

Or in other words: $\int_a^b |f(x)|^2\mathrm{d}x = b|f(b)|^2 - a|f(a)|^2 - \int_a^b x\frac{\mathrm{d}}{\mathrm{d}x}|f(x)|^2\mathrm{d}x$.

Also, note:

$$\tfrac{\mathrm{d}}{\mathrm{d}x}|f(x)|^2 = \tfrac{\mathrm{d}}{\mathrm{d}x}(f(x)\overline{f(x)}) = f'(x)\overline{f(x)} + f(x)\overline{f'(x)} = 2\mathrm{Re}(f'(x)\overline{f(x)}).$$

Hence, for any $a < b$, we have:

$$\int_a^b |f(x)|^2\mathrm{d}x = b|f(b)|^2 - a|f(a)|^2 - 2\mathrm{Re}(\int_a^b xf'(x)\overline{f(x)}\mathrm{d}x)$$

Now since the inequality we want to prove is trivial if $\int |xf(x)|^2\mathrm{d}x = \infty$, we can safely assume $\int |xf(x)|^2\mathrm{d}x < \infty$. This is important because it guarentees that for any $n \in \mathbb{N}$ and $\varepsilon > 0$, we can pick $a_n < -n$ and $b_n > n$ such that $|a_nf(a_n)|^2 < {}^1\!/\!{}_n$ and $|b_nf(b_n)|^2 < {}^1\!/\!{}_n$. In turn, this lets us say that $a_n|f(a_n)|^2 \in (-{}^1\!/\!{}_n, 0]$ and $b_n|f(b_n)|^2 \in [0, {}^1\!/\!{}_n)$ since $a_n < -1$ and $b_n > 1$.

Now by an application of dominated convergence theorem, we know that:

$$\begin{aligned}\int |f(x)|^2\mathrm{d}x &= \lim_{n\to\infty} \int_{a_n}^{b_n} |f(x)|^2\mathrm{d}x \\ &= \lim_{n\to\infty} \Big(b_n|f(b_n)|^2 - a_n|f(a_n)|^2 - 2\mathrm{Re}(\int_{a_n}^{b_n} xf'(x)\overline{f(x)}\mathrm{d}x)\Big) \\ &= -2 \cdot \lim_{n\to\infty} \mathrm{Re}(\int_{a_n}^{b_n} xf'(x)\overline{f(x)}\mathrm{d}x)\end{aligned}$$

Then by the Cauchy-Schwartz inequality (using the fact that $f'\chi_{[a_n,b_n]}, xf \in L^2$), we can say that:

$$\begin{aligned}-2 \cdot \lim_{n\to\infty} \mathrm{Re}(\int_{a_n}^{b_n} xf'(x)\overline{f(x)}\mathrm{d}x) &\leq 2\lim_{n\to\infty} \left|\int_{a_n}^{b_n} xf'(x)\overline{f(x)}\mathrm{d}x\right| \\ &\leq 2\lim_{n\to\infty} \Big(\int_{a_n}^{b_n} |f'(x)|^2\mathrm{d}x\Big)^{1/2}(\int |xf(x)|^2\mathrm{d}x)^{1/2}\end{aligned}$$

By a final application of dominated convergence theorem using an upper bound of $|f'(x)|^2$, we get that:

$$\lim_{n\to\infty} \Big(\int_{a_n}^{b_n} |f'(x)|^2\mathrm{d}x\Big)^{1/2} = (\int |f'(x)|^2\mathrm{d}x)^{1/2}$$

So, $\int |f(x)|^2\mathrm{d}x \leq (\int |f'(x)|^2\mathrm{d}x)^{1/2}(\int |xf(x)|^2\mathrm{d}x)^{1/2}$. Squaring both sides gives the desired inequality.

(c) (**Heisenberg's Inequality**) For any $b, \beta \in \mathbb{R}$,

$$\int (x-b)^2|f(x)|^2\mathrm{d}x \int (\xi - \beta)^2|\widehat{f}(\xi)|^2\mathrm{d}\xi \geq \tfrac{\|f\|_2^4}{16\pi^2}$$

To start off, $f = 0$ a.e. if and only if $\widehat{f} = 0$ a.e. It follows that we will never have a $0 \cdot \infty$ situation on the left-hand side, and thus our inequality is well-defined. Also, if either of the two left hand integrals are infinite, then the inequality is trivial. So, we may assume both integrals are finite.

Next note that if we consider $g(x) = f(x+b)$, then I already showed on page 71 that $\widehat{g}(\xi) = e^{2\pi i\xi b}\widehat{f}(\xi)$. In turn, we know that:

- $\int x^2|g(x)|^2\mathrm{d}x = \int x^2|f(x+b)|^2\mathrm{d}x = \int(x-b)^2|f(x)|^2\mathrm{d}x$,
- $\int(\xi-\beta)^2|\widehat{g}(\xi)|^2\mathrm{d}\xi = \int(\xi-\beta)^2|e^{2\pi i\xi b}\widehat{f}(\xi)|^2\mathrm{d}\xi = \int(\xi-\beta)^2|e^{2\pi i\xi b}\widehat{f}(\xi)|^2\mathrm{d}\xi$,
- $\|g\|_2 = \|f\|_2$.

So, by proving our inequality for g when $b=0$, we've also proven it for f when b is anything.

Going a step further, set $h(x) = e^{-2\pi i\beta x}g(x)$. Then $h^\vee(\xi) = g^\vee(\xi-\beta) = \widehat{g}(\beta-\xi)$. So, we know that $\widehat{h}(\xi) = \widehat{g}(\beta+\xi)$ since $\widehat{g}(\xi) = g^\vee(-\xi)$. In turn:

- $\int x^2|h(x)|^2\mathrm{d}x = \int x^2|e^{-2\pi i\beta x}g(x)|^2\mathrm{d}x = \int x^2|g(x)|^2\mathrm{d}x$,
- $\int \xi^2|\widehat{h}(\xi)|^2\mathrm{d}\xi = \int \xi^2|\widehat{g}(\xi+\beta)|^2\mathrm{d}\xi = \int(\xi-\beta)^2|\widehat{g}(\xi)|^2\mathrm{d}\xi$,
- $\|h\|_2 = \|g\|_2$.

So, by proving our inequality for h when $b=0$ and $\beta=0$, we've also proven it for f when b and β are anything. Luckily, proving that for h is easy due to what we've already proven in parts (a) and (b) of this exercise.

Since $\int\xi^2|\widehat{h}(\xi)|^2\mathrm{d}\xi < \infty$, we know from part (a) that h has an L^2 derivative h' which satisfies that:

$$\tfrac{1}{2\pi i\xi}\widehat{h'}(\xi) = \widehat{h}(\xi).$$

In turn, we can rewrite $\int\xi^2|\widehat{h}(\xi)|^2\mathrm{d}\xi = \frac{1}{4\pi^2}\int|\widehat{h'}(\xi)|^2\mathrm{d}\xi$ and the latter is just $\frac{1}{4\pi^2}\int|h'(\xi)|^2\mathrm{d}\xi$ by the Plancharel theorem. Finally, by applying part (b) we get that:

$$\tfrac{1}{4\pi^2}\int x^2|h(x)|^2\mathrm{d}x\int|h'(\xi)|^2\mathrm{d}\xi \geq \tfrac{\|h\|_2^4}{4}\cdot\tfrac{1}{4\pi^2} = \tfrac{\|h\|_2^4}{16\pi^2}. \ \blacksquare$$

This inequality is the cause of the quantum uncertainty principle. To see why, first note that in quantum mechanics, a property of a particle at a given point in time is modeled as a probability density function whose density at a point x is $|f(x)|^2$ where f is some function in L^2 (importantly this means $\|f\|_2 = 1$ always in this context).

In turn, $\int(x-b)^2|f(x)|^2\mathrm{d}x$ is the formula for the variance of that probability distribution around b. So, that integral evaluates to something small precisely when the probability distribution of the property of the particle has a small standard deviation and b is close to the mean of the distribution.

Next, note that in quantum mechanics, pairs of properties are related to each other by a Fourier transformation. Hence, $|\widehat{f}(\xi)|^2\mathrm{d}x$ is the probability density function of another property of the particle.

Similarly to before, $\int(x-\beta)^2|\widehat{f}(x)|^2\mathrm{d}x$ is the formula for the variance of that probability distribution around β, and that will be small precisely when the probability distribution of the property has a small standard deviation and β is close to the mean of the distribution.

Now finally, $\int(x-b)^2|f(x)|^2\mathrm{d}x\int(x-\beta)^2|\widehat{f}(x)|^2\mathrm{d}x \geq \frac{1}{16\pi^2}$ for all $b,\beta\in\mathbb{R}$ means that it's impossible for both probability distributions to simultaneously have a standard deviation less than $\frac{1}{2\sqrt{\pi}}$, and decreasing one of the standard deviations beyond that value necessarily requires increasing the other. This is the quantum uncertainty principle.

Exercise 8.19: If $f\in L^2(\mathbb{R}^n)$ and the set $S=\{x : f(x)\neq 0\}$ has finite measure, then for any measurable $E\subseteq\mathbb{R}^n$, $\int_E|\widehat{f}|^2\leq\|f\|_2^2 m(S)m(E)$.

By Minkowski's inequality for integrals, we have:

$$\begin{aligned}\int_E|\widehat{f}|^2 &= \int\chi_E(\xi)|\int f(x)e^{-2\pi i\xi\cdot x}\mathrm{d}x|^2\mathrm{d}\xi \\ &= \int|\int f(x)e^{-2\pi i\xi\cdot x}\sqrt{\chi_E(\xi)}\mathrm{d}x|^2\mathrm{d}\xi \\ &\leq \left[\int(\int|f(x)e^{-2\pi i\xi\cdot x}\sqrt{\chi_E(\xi)}|^2\mathrm{d}\xi)^{1/2}\mathrm{d}x\right]^2 = \left[\int|f(x)|(\int_E\mathrm{d}\xi)^{1/2}\mathrm{d}x\right]^2 = m(E)\left(\int|f(x)|\mathrm{d}x\right)^2\end{aligned}$$

Next, by Hölder's inequality we have:

$$\int|f(x)|\mathrm{d}x = \int|\chi_S(x)f(x)|\mathrm{d}x \leq \|\chi_S\|_2\|f\|_2 = \sqrt{m(S)}\|f\|_2.$$

Thus $\int_E|\widehat{f}|^2\leq m(E)(\sqrt{m(S)}\|f\|_2)^2 = m(E)m(S)\|f\|_2^2$. ■

This inequality is another cause/statement of the quantum uncertainty principle. This is because to optimize the precision of our measurement of the property associated to the wave $|\widehat{f}|^2$, we'd want to maximize $\int_E|\widehat{f}|^2$ while simultaneously minimizing $m(E)$. But, this inequality says that doing that requires increasing $m(S)$. I.e., it requires us to know less about the property associated to the wave $|f|^2$.

7/15/2025

Welp, I'm currently sick. Anyways, now that I've scanned my paper notes from winter quarter, I'm thinking I want to finally learn vector calculus properly since I never really learned it in math 20E. Also, since I'm crashing the physics 4 sequence, I'm going to eventually need to finally learn Stokes' theorem and Divergence theorem when they cover E&M.

For now, my plan is to sort of follow along with Munkres' Analysis on Manifolds, starting at chapter 5. That said, I want to work with Lebesgue integrals. So, I might go on some tangents and or come up with different proofs for things. Also, I might skip something if I'm bored.

Conventions:

- $\{e_1, \ldots, e_n\}$ will refer to the standard basis on $\mathbb{R}^n$.
- If $f \in C^r(U)$ where $U \subseteq \mathbb{R}^n$ is open and $r \geq 1$, then $\mathrm{D}f$ will refer to the derivative matrix of f with respect to the standard bases. I.e, $\mathrm{D}f$ is the matrix of partial derivatives of f.

Lemma: Let W be a k-dimensional linear subspace of $\mathbb{R}^n$. Then there is an orthogonal (i.e. unitary) linear transformation $h : \mathbb{R}^n \to \mathbb{R}^n$ that carries W onto the subspace $\mathbb{R}^k \times 0^{n-k}$ of $\mathbb{R}^n$.

Proof:
Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis for W such that $\{v_1, \ldots, v_k\}$ is an orthonormal basis for W. Then if g is the linear map whose matrix with respect to the standard basis has columns, $v_1, \ldots, v_n$, we know g is orthogonal and $g(e_i) = v_i$ for all i. Now just set $h = g^{-1}$.

Theorem: Let $k, n \in \mathbb{N}$ with $0 < k \leq n$. There is a unique function V that assigns to each k-tuple: $(x_1, \ldots, x_k)$, of elements in $\mathbb{R}^n$ a nonnegative number such that:

1. If $h : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal transformation, then:
$$V(h(x_1), \ldots, h(x_k)) = V(x_1, \ldots, x_k)$$
2. If $y_1, \ldots, y_k$ belong to the subspace $\mathbb{R}^k \times 0^{n-k}$ with $y_i = \left[\begin{smallmatrix} z_i \\ 0 \end{smallmatrix}\right]$ where $z_i \in \mathbb{R}^k$, then $V(y_1, \ldots, y_k) = |\det(z_1, \ldots, z_k)|$.

Specifically, we have, $V(x_1, \ldots, x_k) = \left(\det(X^\mathsf{T} X)\right)^{1/2}$ where X is the $n \times k$ matrix $X = [x_1, \ldots, x_k]$.

(Note: we will typically abreviate $V(x_1, \ldots, x_k)$ as $V(X)$...)

Proof:
Given $X = [x_1, \ldots, x_k]$, define $F(X) = \det(X^\mathsf{T} X)$. Then note:

- If $h : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal linear map, then $h(x) = Ax$ where A is an orthogonal matrix. Now $[h(x_1), \ldots, h(x_k)] = [Ax_1, \ldots, Ax_k] = AX$. Thus:
$F(h(X)) = f(AX) = \det\,(AX)^\mathsf{T} AX = \det X^\mathsf{T} A^\mathsf{T} AX = \det X^\mathsf{T} X = F(X)$
- If Z is a $k \times k$ matrix and Y is the $n \times k$ matrix $\left[\begin{smallmatrix} Z \\ 0 \end{smallmatrix}\right]$, then:
$$F(Y) = \det\left(\begin{bmatrix} Z^\mathsf{T} & 0 \end{bmatrix} \begin{bmatrix} Z \\ 0 \end{bmatrix}\right) = \det\left(Z^\mathsf{T} Z\right) = (\det(Z))^2$$

With that, all we need to do is show that F is nonnegative so that we can take the square root of F. Luckily, if $\{x_1, \ldots, x_k\}$ are any k-tuple of vectors in $\mathbb{R}^n$, then we know there is a k-dimensional linear subspace of $\mathbb{R}^n$ containing $x_1, \ldots, x_k$. By our prior lemma, there exists an orthogonal linear map h taking W to $\mathbb{R}^k \to 0^{n-k}$. Next by our first bullet, we know that $F(h(X)) = F(X)$. And finally, by our second bullet we know $F(h(X)) = (\det(h(X)))^2 \geq 0$.

Now just define $V(X) = \sqrt{F(X)}$. This proves the existence part of the theorem.

To prove uniqueness, suppose U also satisfies our two axioms. Then for any $\{x_1, \ldots, x_k\} \subseteq \mathbb{R}^n$, let $h(x) = Ax$ be an orthogonal linear map taking $\{x_1, \ldots, x_k\}$ into $\mathbb{R}^k \times 0^{n-k}$. Thus $U(X) = U(h(X)) = |\det(AX)| = V(h(X)) = V(X)$

Side note: if $\{x_1, \ldots, x_k\}$ are not linearly independent, then $V(X) = 0$.

Also, hopefully it's clear that the significance of V is that we now have a way of defining the k-dimensional volume of a k-dimensional parallelepiped in $\mathbb{R}^n$.

With that, we're ready to start integrating on manifolds. We'll start with the simple case of a manifold parametrized by a single function.

Let $k \leq n$. Let A be open in $\mathbb{R}^k$ and let $\alpha : A \to \mathbb{R}^n$ be an injective C^r map (where $r \geq 1$). The set $Y = \alpha(A)$ together with the map α constitute a parametrized-manifold of dimension k. We denote this parametrized manifold Y_α. For a topology, we equip Y_α with the subspace topology of Y with respect to $\mathbb{R}^n$. That way α is still a continuous map.

Next, we define a Borel measure on Y_α. Given any set $E \in \mathcal{B}_{Y_\alpha}$, we define:

$$V(E) := \int_{\alpha^{-1}(E)} V(\mathrm{D}\alpha)$$

(unfortunately the measure is typically called V even though we already named another function that.)

Note, $V(\mathrm{D}\alpha)$ is Borel measurable because the matrix determinant is a continuous function with respect to all the matrix entries and all the entries of $\mathrm{D}\alpha$ are continuous since $\alpha \in C^1$. It follows that our integral is well-defined.

Since $V(\emptyset) = 0$ and V is clearly countably additive, we know V is a measure. Also, the naturalness of this measure is hopefully clear. After all, you can imagine that we are approximating the k-dimensional volume using a bunch of tiny parallelepipeds.

Then, given any measurable function $f : Y \to \mathbb{C}$ on our manifold, we can integrate via the formula: $\int_{Y_\alpha} f \mathrm{d}V = \int_A (f \circ \alpha) V(\mathrm{D}\alpha) \mathrm{d}m$.

Why: This formula clearly holds for simple functions. Then you can extend that to nonnegative functions via the monotone convergence theorem and then to real and complex functions in the standard way.

Theorem: Let $g : A \to B$ be a diffeomorphism of open sets in $\mathbb{R}^k$ and let $\beta : B \to \mathbb{R}^n$ be an injective C^r map (with $r \geq 1$). If we define $\alpha = \beta \circ g$, then $\alpha : A \to \mathbb{R}^n$ is also an injective C_r map and $\alpha(A) = \beta(B) = Y$. Then, a function $f : Y \to \mathbb{C}$ is integrable over Y_β if and only if it is integrable over Y_α, in which case:

$$\int_{Y_\alpha} f \mathrm{d}V_\alpha = \int_{Y_\beta} f \mathrm{d}V_\beta$$

Proof:
The measurability of f is independent of our parametrization since the topology of Y_α and Y_β was not defined using α or β. Next note that by change of variables:

$$\begin{aligned}\int_{Y_\beta} f\mathrm{d}V_\beta = \int_B (f\circ\beta)V(\mathrm{D}\beta)\mathrm{d}m &= \int_A (f\circ\beta\circ g)(V(\mathrm{D}\beta)\circ g)|\det(\mathrm{D}g)|\mathrm{d}m \\ &= \int_A (f\circ\alpha)(V(\mathrm{D}\beta)\circ g)|\det(\mathrm{D}g)|\mathrm{d}m\end{aligned}$$

Thus, we just need to show that $(V(\mathrm{D}\beta)\circ g)|\det(\mathrm{D}g)| = V(\mathrm{D}\alpha)$. To do that, note by chain rule that $\mathrm{D}\alpha = ((\mathrm{D}\beta)\circ g)\mathrm{D}g$. Therefore:

$$\begin{aligned}(V(\mathrm{D}\alpha))^2 &= \det((((\mathrm{D}\beta)\circ g)\mathrm{D}g)^\mathsf{T}(((\mathrm{D}\beta)\circ g)\mathrm{D}g)) \\ &= (\det(\mathrm{D}g))^2\det(((\mathrm{D}\beta)\circ g)^\mathsf{T}((\mathrm{D}\beta)\circ g)) = (\det(\mathrm{D}g))^2(V((\mathrm{D}\beta)\circ g))^2 \blacksquare\end{aligned}$$

Exercise 22.1: Let A be open in $\mathbb{R}^k$, $\alpha : A \to \mathbb{R}^n$ be a C^1 map, and $Y = \alpha(A)$. Suppose $h : \mathbb{R}^n \to \mathbb{R}^n$ is an isometry and let $Z = h(Y)$ and $\beta = h\circ\alpha$. Then Y_α and Z_β have the same volume.

Since h is an isometry, we know h has the form $h(x) = Qx + b$ where Q is an orthogonal matrix and b is a constant vector. Thus by chain rule, we have that $\mathrm{D}\beta = Q\mathrm{D}\alpha$. So:

$$\int_{Z_\beta}\mathrm{d}V = \int_A V(\mathrm{D}\beta) = \int_A V(Q\mathrm{D}\alpha) = \int_A V(\mathrm{D}\alpha) = \int_{Y_\alpha}\mathrm{d}V$$

7/16/2025

Now our previous approach to manifolds is lacking in some respects. For one, it'd be nice if our manifolds were able to include their boundaries. However, requiring the domains of our parametrizations to be strictly open makes that difficult. Also, it'd be nice if we could talk about manifolds that can't be parametrized by a single function (such as the unit sphere S^2 in $\mathbb{R}^3$).

To address the first issue, we extend our notion differentiability. Let S be any subset of $\mathbb{R}^k$ and let $f : S \to \mathbb{R}^n$. We say $f \in C^r(S)$ iff there exists an open set $U \supset S$ and a function $g : U \to \mathbb{R}^n$ in $C^r(U)$ such that $g|_S = f$.

Note that if $f_1, f_2 \in C^r(S)$, then we have that $f_1 + f_2 \in C^r(S)$ and $f_1 f_2 \in C^r(S)$. After all, supposing we can extend f_1 and f_2 to the functions g_1 and g_2 defined respectively on the open sets U and V containing S, then $U \cap V$ is still open and contains S, and $g_1 + g_2, g_1 g_2 \in C^r(U\cap V)$.

Also, if $f_1 \in C^r(S)$ and $f_2 \in C^r(T)$ where $f_1(S) \subseteq T$, then $f_2 \circ f_1 \in C^r(S)$. This is because if g_1 extends f_1 to an open set $U \supseteq S$ and g_2 extends f_2 to an open set $V \supseteq T$, then $g_2 \circ g_1$ is a C^r extension of $f_2 \circ f_1$ to the open set $g_1^{-1}(V) \supseteq S$.

Next, we show that being in $C^r(S)$ is a local property. However, to do that we first prove a result about partitions of unity.

Recall from Math 240B that a partition of unity on a set $E \subseteq X$ (where X is an LCH space such as $\mathbb{R}^n$) is a collection of functions $\{h_\alpha\}_{\alpha \in A} \in C(X, [0,1])$ such that:

- Each $x \in X$ has a neighborhood on which only finitely many h_α are nonzero,

  Although Munkres only requires this to hold for all $x \in E$.

- $\sum_{\alpha \in A} h_\alpha(x) = 1$ for all $x \in E$.

Also, we say $\{h_\alpha\}_{\alpha \in A}$ is subordinate to the open cover $\mathcal{U}$ of E if for each α there exists $U \in \mathcal{U}$ such that $\mathrm{supp}(h_\alpha) \subseteq U$.

A really cool result Munkres spends all of chapter 16 proving is that for any collection $\mathcal{A}$ of open sets in $\mathbb{R}^n$ whose union is A, there is a partition of unity $\{h_m\}_{m \in \mathbb{N}}$ of A consisting of C_c^∞ functions and such that $\{h_m\}_{m \in \mathbb{N}}$ is subordinate to $\mathcal{A}$.

Lemma 16.2: Let $\mathcal{A}$ be a collection of open sets in $\mathbb{R}^n$ whose union is A. Then there exists a countable collection $\{Q_i\}_{i \in \mathbb{N}}$ of rectangles (i.e. Cartesian products of closed intervals) contained in A such that:

1. The sets $\{Q_i^\circ\}_{i \in \mathbb{N}}$ cover A.
2. Each Q_i is contained in an element of $\mathcal{A}$.
3. Each point of A has a neighborhood that intersects only finitely many of the sets Q_i.

Proof:

Step 1: Dividing A into a nicely structured sequence of compact sets:

Because $\mathbb{R}^n$ is σ-compact, we know A is a countable union of compact sets. Then, by taking larger and larger finite unions of those compact sets we get an increasing sequence of compact sets $\{K_i\}_{i \in \mathbb{N}}$ whose union is A.

Next, to get a nicer sequence, we do more finagling. Let $D_1 = K_1$. Then for $i \geq 1$, define D_{i+1} inductively as follows:

We know there exists a precompact open set V such that $(D_i \cup K_i) \subseteq V \subseteq \overline{V} \subseteq A$. Thus, set $D_{i+1} = \overline{V}$.

Thus, $\{D_i\}_{i \in \mathbb{N}}$ is a sequence of compact sets whose union is A and which satisfies that $D_i \subseteq D_{i+1}^\circ$ for all i. Also, for convenience of notation let $D_i = \emptyset$ if $i \leq 0$.

Finally, set $B_i = D_i - D_{i-1}^\circ$ for all i. Then note that $\{B_i\}_{i \in \mathbb{N}}$ is a sequence of compact sets whose union is A and which satisfies that B_i is disjoint from D_{i-2} for all $i \geq 2$. Consequently, this means that any B_i only intercepts B_{i-1} and B_{i+1} in our sequence. Also, $U_i := D_{i+1}^\circ - D_{i-2}$ is an open neighborhood of B_i which intercepts only B_{i-1}, B_i, and B_{i+1}.

Step 2: Making our covering of A:

After fixing i, note that for any $x \in B_i$ we know that there is a set $E \in \mathcal{A}$ such that $x \in E$. So, we can pick a rectangle Q_x such that $x \in Q_x^\circ$ and $Q_x \subseteq E \cap U_i$. Doing this for all $x \in B_i$, we get a collection $\{Q_x\}_{x \in B_i}$ of sets whose interiors give an open covering of B_i. So, because B_i is compact, there is a finite collection $\mathcal{C}_i := \{Q_{x_1}, \ldots, C_{Q_{n_i}}\}$ of rectangles whose interiors cover all of B_i and such that each rectangle is a subset of some element of $\mathcal{A}$ intercepted with U_i.

Repeat this process for all i and let $\mathcal{C} = \bigcup_{i \in \mathbb{N}} \mathcal{C}_i$. Then $\mathcal{C}$ is a countable collection of rectangles such that each $Q \in \mathcal{C}$ is contained in an element of $\mathcal{A}$. Also, for any $x \in A$, we know there is some j with:

$$x \in (B_{j-1} \cup B_j \cup B_{j+1}) - \bigcup_{\substack{i \in \mathbb{N} \\ i \notin \{j-1, j, j+1\}}} B_i$$

In turn, x is in the interior of one the rectangles in $\mathcal{C}_{j-1} \cup \mathcal{C}_j \cup \mathcal{C}_{j+1}$.

Finally, supposing $x \in B_k$, then we know the open neighborhood $U_k \subseteq A$ of x only intercepts U_{k-1} and U_{k+1}, which in turn only intercept B_{k-2} through B_{k+2}. So, U_k is an open neighborhood of x which intercepts at most the rectangles from $\mathcal{C}_{j-2}$ through $\mathcal{C}_{j+2}$. ■

Theorem 16.3: Let $\mathcal{A}$ be a collection of open sets in $\mathbb{R}^n$ whose union is A. Then there is a partition of unity $\{h_m\}_{m \in \mathbb{N}}$ subordinate to $\mathcal{A}$ such that each h_m is in C_c^∞.

Proof:

Construct $\{Q_m\}_{m \in \mathbb{N}}$ like in the previous lemma. Then note that for each Q_m, there is a C_c^∞ function g_m such that $g_m(x) > 0$ if $x \in Q_m^\circ$ and $g_m(x) = 0$ otherwise.

Specifically: If you define $f(x) = e^{1/x}$ when $x > 0$ and $f(x) = 0$ when $x \leq 0$, then $f \in C^\infty(\mathbb{R})$. In turn, $g := f(x)f(1-x) \in C^\infty(\mathbb{R})$ with $g(x) > 0$ when $x \in (0,1)$ and $g(x) = 0$ otherwise.

Finally, if $Q_m = [a_1, b_1] \times \ldots \times [a_n, b_n]$, then define:

$$g_m(x_1, \ldots, x_n) = g(\tfrac{x_1 - a_1}{b_1 - a_1}) \cdots g(\tfrac{x_n - a_n}{b_n - a_n})$$

Thus $g_m \in C^\infty(\mathbb{R}^n)$ with $g_m(x) > 0$ when $x \in Q_m^\circ$ and $g_m(x) = 0$ otherwise.

Having done that, we've now guarenteed that $\{g_m\}_{m \in \mathbb{N}}$ has all the properties we want except that we don't necessarily have that $\sum_{m \in \mathbb{N}} g_m(x) = 1$ for all $x \in A$. To fix that, we normalize our functions.

Let $\lambda(x) := \sum_{m=1}^{\infty} g_m(x)$. Then for any $x \in A$, we know there is an open neighborhood N_x of x that intercepts only finitely many $\operatorname{supp}(g_m)$. It follows then that $\lambda(x) < \infty$ for all $x \in A$ and that λ is infinitely differentiable for all $x \in A$. Meanwhile, since any $x \in A$ is in the interior of at least one Q_m, we know that $\lambda(x) > 0$ for all $x \in A$.

Now for each m define $h_m(x) = g_m(x)/\lambda(x)$ when $x \in A$ and $h_m(x) = 0$ when $x \notin A$. Then it's clear that $\sum_{m\in\mathbb{N}} h_m(x) = 1$ when $x \in A$. Also, we still have that each $h_m \in C^\infty$. To see that, first note that h_m is infinitely differentiable via quotient rule on A. Also, since Q_m is compact, A^{C} is closed, and both are disjoint, we know there is some minimum distance δ between the two sets. So for any $x \in A^{\mathsf{C}}$, we know that h_m is just the zero function while on a ball of radius $\delta/2$ around x. So, all of the derivatives of h_m exist and equal zero at x for any $x \in A^{\mathsf{C}}$. Finally, since $\mathrm{supp}(h_m) = \mathrm{supp}(g_m)$, we have that h_m satisfies our other requirements.

Now returning to our goal of extending the concept of differentiability, we have the following result:

Lemma 23.1: Let S be a subset of $\mathbb{R}^k$ and let $f : S \to \mathbb{R}^n$. If for each $x \in S$ there is a neighborhood U_x of x and a C^r function $g_x : U_x \to \mathbb{R}^n$ which agrees with f on $U_x \cap S$, then $f \in C^r(S)$.

Proof:

For each $x \in S$, pick a set U_x and a C^r function $g_x : U_x \to \mathbb{R}^n$ as allowed by the hypothesis of the lemma. Then set $\mathcal{A} = \{U_x : x \in S\}$ and call the union of that collection of sets A. Via the prior result, there is a partition of unity $\{\phi_m\}_{m\in\mathbb{N}}$ on A consisting of C_c^∞ functions and which is subordinate to $\mathcal{A}$. In turn, for any $m \in \mathbb{N}$ there exists x_m with $\mathrm{supp}(\phi_m) \subseteq U_{x_m}$. It then follows that $h_m \coloneqq \phi_m g_{x_m} \in C^r(U_m)$ and that h_m vanishes outside a compact subset of U_m, meaning we can extend h_m to being in $C^r(\mathbb{R}^k)$ by setting $h_m = 0$ outside U_{x_m}.

Finally, define $g = \sum_{m=1}^\infty h_m$ on A. Then for any $x \in A$, we know that x has a neighborhood on which g is only a sum of finitely many h_m. So, $g \in C^r(A)$. Also, if $x \in S$, then for any m with $\phi_m(x) \neq 0$, $h_m(x) = \phi_m(x)g_m(x) = \phi_m(x)f(x)$. Therefore, for any $x \in A \cap S$:

$$g(x) = \sum_{m\in\mathbb{N}} h_m(x) = f(x)\sum_{m\in\mathbb{N}} \phi_m(x) = f(x). \blacksquare$$

7/18/2025

Let $H^k = \{(x_1, \ldots, x_k) \in \mathbb{R}^k : x_k \geq 0\}$ denote the "upper-Half-space". Also, let $H_+^k = \{(x_1, \ldots, x_k) \in \mathbb{R}^k : x_k > 0\}$ denote the "open upper-Half-space" in $\mathbb{R}^k$.

Lemma 23.2: Let U be open in H^k but not in $\mathbb{R}^k$, and let $\alpha : U \to \mathbb{R}^n$ be in $C^r(U)$. That way, there exists a C^r extension $\beta : U' \to \mathbb{R}^n$ of α defined on an open set U' of $\mathbb{R}^k$. Then for $x \in U$, the derivative $\mathrm{D}\beta(x)$ depends only on the function α and is independent of the extension β. Hence it follows we may denote this derivative by $\mathrm{D}\alpha(x)$ without ambiguity.

Why: We know that $\frac{\partial}{\partial x_k}\beta$ is fully determined by the right-hand limit:

$$\lim_{h\to 0^+} \frac{\beta(x+he_k)-\beta(x)}{h} = \lim_{h\to 0^+} \frac{\alpha(x+he_k)-\alpha(x)}{h}$$

It follows that all first order partial derivatives of β on U are uniquely determined by α. Then proceeding by induction and noting that $\partial^\gamma \beta$ extends $\partial^\gamma \alpha$ from U to U' for all multi-indices γ of degree less than r, we can apply the same reasoning to conclude that all partial derivatives of β with order less than r on U are uniquely determined by α.

With that, we now have the ability to define parametrized manifolds with a boundary by making the domain of our parametrization open with respect to H^k instead of $\mathbb{R}^k$. As for our other issue of wanting to talk about manifolds that can't be parametrized by a single function, we shall deal with that now.

Let $k > 0$. A k-manifold in $\mathbb{R}^n$ of class C^r is a subset M of $\mathbb{R}^n$ having the following property: For each $p \in M$ there is an open set V of M containing p, a set U that is open in either $\mathbb{R}^k$ or H^k, and a continuous bijective map $\alpha : U \to V$ such that:

1. $\alpha \in C^r(U)$
2. $\alpha^{-1} : V \to U$ is continuous
3. $\mathrm{D}\alpha(x)$ has rank k for each $x \in U$.

The map α is called a coordinate patch on M about p.

Also, we call a discrete collection of points in $\mathbb{R}^n$ to be a 0-manifold.

Lemma 23.3: Let M be a manifold in $\mathbb{R}^n$ and $\alpha : U \to V$ be a coordinate patch on M. If U_0 is a subset of U that is open in U, then the restriction of α to U_0 is also a coordinate patch on M.

Proof:
Since α^{-1} is continuous and U_0 is open in U, we know that $V_0 := \alpha(U_0)$ is also open in V and thus also M. Hence, $\alpha|_{U_0}$ is a coordinate patch on M because it carries U_0 onto V_0 in a bijective fashion, and it's a C^r map with a continuous inverse and $\mathrm{D}(\alpha|_{U_0})$ having rank k just because it's the restriction of α which has all of those things.

Exercise 23.3(b): Why is $\alpha : [0, 1) \to \mathbb{R}^2$ defined by $\alpha(t) = (\cos(2\pi t), \sin(2\pi t))$ not a coordinate patch for the unit circle S^1?

In this example, α^{-1} is not continuous at $\alpha(0) = (1, 0)$. One way to see this is that the limit of α^{-1} going one way around the circle towards $(1, 0)$ will be 1, whereas the limit going the other way around the circle will be 0.

It should be noted though that S^1 is still a 1-manifold. It's just that we need to use multiple overlapping coordinate patches that don't individually go all the way around the circle in order to cover it.

In order to prove the next theorem, I actually need to generalize the version of the inverse function theorem that I learned in 140C so that if f is a bijective C^r map with a nonsingular derivative, then I know that $g = f^{-1}$ is also C^r rather than just merely C^1. But to do that, I need to finally learn Cramer's rule.

I'm also realizing right about now that in my original notes where I defined $\det$ (my MITx notes which I just got back from my parents, yay!), while my construction still generalizes to matrices defined on arbitrary scalar fields perfectly well, it does have a slight problem of using the parity of permutations in it's definition. However, in Math 100A we defined the parity of a permutation by taking the determinant of it's matrix representation. So, I might as well deal with that cyclic definition now...

Given a permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$, we can define the function:

$$\mathrm{sgn}(\sigma) = \prod_{i<j} \mathrm{sgn}(\sigma(j) - \sigma(i))$$

Now I claim that sgn is a group homomorphism from S_n to the multiplicative group $\{-1, 1\} \subseteq \mathbb{R}^n$ To see this, suppose $\sigma \in S_n$ and that $\tau_{k_1,k_2} \in S_n$ is the transposition swapping k_1 and k_2. Then it's clear that:

$$\mathrm{sgn}(\tau_{k_1,k_2} \circ \sigma) = -\mathrm{sgn}(\sigma) = \mathrm{sgn}(\tau_{k_1,k_2})\mathrm{sgn}(\sigma).$$

In turn, if $\sigma' \in S_n$ is another arbitrary permutation, then by expressing $\sigma' = \tau_1 \circ \tau_2 \circ \ldots \circ \tau_N$ where all the τ_i are transpositions, we have that:

$$\begin{aligned}
\mathrm{sgn}(\sigma' \circ \sigma) &= \mathrm{sgn}(\tau_1 \circ \tau_2 \circ \ldots \circ \tau_{N-1} \circ \tau_N \circ \sigma) \\
&= \mathrm{sgn}(\tau_1)\mathrm{sgn}(\tau_2 \circ \ldots \circ \tau_{N-1} \circ \tau_N \circ \sigma) \\
&\vdots \\
&= \mathrm{sgn}(\tau_1)\mathrm{sgn}(\tau_2) \ldots \mathrm{sgn}(\tau_{N-1})\mathrm{sgn}(\tau_N)\mathrm{sgn}(\sigma) \\
&= \mathrm{sgn}(\tau_1)\mathrm{sgn}(\tau_2) \ldots \mathrm{sgn}(\tau_{N-1} \circ \tau_N)\mathrm{sgn}(\sigma) \\
&\vdots \\
&= \mathrm{sgn}(\tau_1)\mathrm{sgn}(\tau_2 \circ \cdots \tau_{N-1} \circ \tau_N)\mathrm{sgn}(\sigma) \\
&= \mathrm{sgn}(\tau_1 \circ \tau_2 \circ \cdots \tau_{N-1} \circ \tau_N)\mathrm{sgn}(\sigma) = \mathrm{sgn}(\sigma')\mathrm{sgn}(\sigma)
\end{aligned}$$

Also, it's easily checked that $\mathrm{sgn}(\mathrm{Id}) = 1$. Thus sgn is a group homomorphism. And, since every transposition has a negative sign, we get the following nice interpretation of the sign of a permutation. Specifically: $\mathrm{sgn}(\sigma) = 1$ if σ can only be constructed using an even number of transpositions starting from the identity, and $\mathrm{sgn}(\sigma) = -1$ if σ can only be constructed using an odd number of transpositions starting from the identity.

Next, here are Cramer's rules:

Theorem 2.13: Let $A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$ be an $n \times n$ matrix. Also let:

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ and } c = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

Then if $Ax = c$, we have:

$$\det(A)x_i = \det\left(\begin{bmatrix} a_1 & \cdots & a_{i-1} & c & a_{i+1} & \cdots & a_n \end{bmatrix}\right)$$

Proof:
Define $C = \begin{bmatrix} e_1 & \cdots & e_{i-1} & x & e_{i+1} & \cdots & e_n \end{bmatrix}$.

Then, we have that $AC = \begin{bmatrix} a_1 & \cdots & a_{i-1} & c & a_{i+1} & \cdots & a_n \end{bmatrix}$. Thus, we know:

$$\det(A)\det(C) = \det\left(\begin{bmatrix} a_1 & \cdots & a_{i-1} & c & a_{i+1} & \cdots & a_n \end{bmatrix}\right).$$

Also note that $\det(C) = x_i(-1)^{i+i}\det(I_{n-1}) = x_i(-1)^{2i} = x_i$. The desired conclusion then follows.

Theorem 2.14: Let A be an $n \times n$ matrix of rank n and let $B = A^{-1} = [b_{i,j}]$. Then letting $A_{j,i}$ denote the matrix which results from removing the jth row and ith column of A, we have that:

$$b_{i,j} = \frac{(-1)^{j+i}\det(A_{j,i})}{\det(A)}$$

Proof:
After fixing j, set $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ equal to the jth column of B.

Now since $AB = I_n$, we know $Ax = e_j$. Therefore, by our last theorem we know:

$$\det(A)x_i = \det\left(\begin{bmatrix} a_1 & \cdots & a_{i-1} & e_j & a_{i+1} & \cdots & a_n \end{bmatrix}\right)$$

The latter determinant equals $(-1)^{j+i}\det(A_{j,i})$. Therefore:

$$x_i = b_{i,j} = \frac{(-1)^{j+i}\det(A_{j,i})}{\det(A)}. \blacksquare$$

Now, here's the generalization of the inverse function theorem, that if $f \in C^r$ and satisfies all the other hypotheses of the inverse function theorem, then we can guarentee that our inverse function $g = f^{-1}$ is also C^r.

Proof:
Jumping to where we ended our proof of the inverse function theorem in math 140C, we had shown that $\mathrm{D}g(y) = (\mathrm{D}f(g(y)))^{-1}$ for all y in some open subset of the image of f. If $g(y) = (g_1(y), \ldots, g_n(y))$, then by our prior theorem we know that:

$$\frac{\partial}{\partial y_j}g_i(y) = \frac{(-1)^{j+i}\det([\mathrm{D}f(g(y))]_{j,i})}{\det(\mathrm{D}f(g(y)))}$$

Now we already know that g is C^1. So, when proceeding by induction for $r > 1$, it suffices to assume g is also C^{r-1}. But then note that since $\det(\mathrm{D}f(g(y))) \neq 0$ for all y, and since all the partial derivatives of f are C^{r-1}, our above expression shows that $\frac{\partial}{\partial y_j}g_i(y)$ is also C^{r-1}. And since this works for all partial derivatives of g, we've proven that g is $C^{(r-1)+1}$. $\blacksquare$

And finally to finish off for tonight…

Theorem 24.1 Let M be a k-manifold in $\mathbb{R}^n$ of class C^r. Also let $a_0 : U_0 \to V_0$ and $a_1 : U_1 \to V_1$ be coordinate patches on M with $W = V_0 \cap V_1$ nonempty, and let $W_i = \alpha_i^{-1}(W)$. Then the map $(\alpha_1^{-1} \circ \alpha_0) : W_0 \to W_1$ is a C^r map with a nonsingular derivative.

(Side note: We often call $\alpha_1^{-1} \circ \alpha_0$ the transition function between the coordinate patches α_0 and α_1.)

Proof:
It suffices to show that if $\alpha : U \to V$ is a coordinate patch on M, then $\alpha^{-1} : V \to \mathbb{R}^k$ is in $C^r(V)$.

Why: If α_0 and α_1^{-1} are both of class C^r, then so is their composite $\alpha_1^{-1} \circ \alpha_0$. By similar reasoning, we also know that $\alpha_0^{-1} \circ \alpha_1 : W_1 \to W_0$ is in $C^r(W_0)$. And since $(\alpha_1^{-1} \circ \alpha_0)$ and $\alpha_0^{-1} \circ \alpha_1$ are inverses of each other, we know by chain rule that for any $x \in W_0$ and $y = (\alpha_1^{-1} \circ \alpha_0)(x)$:

$$\mathrm{D}(\alpha_0^{-1} \circ \alpha_1)(y)\mathrm{D}(\alpha_1^{-1} \circ \alpha_0)(x) = \mathbf{1}$$

The only way this is possible is if $\det(\mathrm{D}(\alpha_1^{-1} \circ \alpha_0)) \neq 0$ for all $x \in W_0$.

Next, to prove that α^{-1} is of class C^r, it suffices to show that it is locally of class C^r. So let p_0 be a point of V and set $x_0 = \alpha^{-1}(p_0)$.

First consider the case U is open in H^k but not in $\mathbb{R}^k$. Then, we can extend α to a C^r map β on an open set U' of $\mathbb{R}^k$. Now $\mathrm{D}\alpha(x_0)$ has rank k. So after some suitable permutation of our standard basis vectors, we can assume the first k rows of the matrix $\mathrm{D}\alpha(x_0)$ are linearly independent. If we then define $\pi : \mathbb{R}^n \to \mathbb{R}^k$ to be the projection from $\mathbb{R}^n$ onto those first k basis vector coordinates, we have that the map $g = \pi \circ \beta$ maps U' into $\mathbb{R}^k$ and $\mathrm{D}g(x_0)$ is non-singular. So by the inverse function theorem, we know g is a C^r diffeomorphism on an open set W of $\mathbb{R}^k$ about x_0 (meaning g is C^r, g has an inverse g^{-1}, and g^{-1} is C^r).

Now we claim $h = g^{-1} \circ \pi$ (which is a C^r map) extends α^{-1} to a neighborhood A of p_0. Firstly note $U_0 := W \cap U$ is open in U. Hence, α^{-1} being continuous implies that $V_0 := \alpha(U_0)$ is open in V. This means there is an open set $A \subseteq \mathbb{R}^n$ such that $A \cap V = V_0$. By intercepting A with $\pi^{-1}(g(W))$ ($= \beta(W)$), we can force A to be contained in the domain of h.

Now $h : A \to \mathbb{R}^k$ is of class C^r, and if $p \in A \cap V = V_0$, then when letting $x = \alpha^{-1}(p)$ we have:

$$h(p) = h(\alpha(x)) = g^{-1}(\pi(\alpha(x))) = g^{-1}(g(x)) = x = \alpha^{-1}(p).$$

As for the case where U is open in $\mathbb{R}^k$, then just set $U' = U$ and $\beta = \alpha$ and the prior reasoning still works. ∎

Side note: As a corollary, we now know that if two coordinate patches parametrize the same manifold, then the domains of those two coordinate patches are diffeomorphic. So hopefully that adds to the significance of theorem I wrote at the bottom of page 79.

7/19/2025

Today I shall formalize what I mean by the "boundary" of a manifold and maybe also do some cool exercises. I'll try to get as much done as possible since I'm going to be busy at San Diego pride tomorrow.

Let M be a k-manifold in $\mathbb{R}^n$ and let $p \in M$. If there is a coordinate patch $\alpha : U \to V$ on M about p such that U is open in $\mathbb{R}^k$, we say p is an interior point of M. Otherwise, we say p is a boundary point of M.

We denote the set of boundary points of M as ∂M and call it the boundary of M. Meanwhile, we call $M - \partial M$ the interior of M. Note that these definitions are distinct from the topological definitions of boundaries and interiors.

> Untill I get bored of this and go back to Folland or someone else, I'll be using ∂M to refer to the manifold definition of boundary as opposed to a different definition.

Lemma 24.2: Let M be a k-manifold in $\mathbb{R}^n$ and $\alpha : U \to V$ be a coordinate patch about the point p of M.

(a) If U is open in $\mathbb{R}^k$, then p is an interior point of M.
(b) If U is open in H^k and $p = \alpha(x_0)$ where $x_0 \in H^k_+$, then p is an interior point of M.
(c) If U is open in H^k and $p = \alpha(x_0)$ where $x_0 \in \mathbb{R}^{k-1} \times 0$, then p is a boundary point of M.

Proof:
Parts (a) and (b) are trivial. As for part (c), let $\alpha_0 : U_0 \to V_0$ be the coordinate patch in the hypothesis of the lemma and suppose (for the sake of contradiction) that there is another coordinate patch $\alpha_1 : U_1 \to V_1$ about p with U_1 open in $\mathbb{R}^k$.

Since V_0 and V_1 are open in M, the set $W = V_0 \cap V_1$ is also open in M. Let $W_i = \alpha_i^{-1}(W)$ for $i = 0, 1$. Then W_0 is open in H^k and contains x_0 (which consequently means W_0 isn't open in $\mathbb{R}^k$). Also, W_1 is open in $\mathbb{R}^k$. But now note that by our prior theorem, $\alpha_0^{-1} \circ \alpha_1$ is a C^r map from W_1 to $W_0 \subseteq \mathbb{R}^k$ with a nonsingular derivative matrix. So, by specifically part (A) of the inverse function theorem (as covered in math 140C), we know W_1 maps to an open set in $\mathbb{R}^k$. Yet $\alpha_0^{-1} \circ \alpha_1(W_1) = W_0$ is not open in $\mathbb{R}^k$. Hence, a contradiction. ■

> Side note: Holy fuck I did not realize before now that in math 140C we proved that C^1 functions to $\mathbb{R}^k$ with a nonsingular derivative matrix are open maps.

Note, we trivially have that H^k is a k-manifold of class C^∞ in $\mathbb{R}^k$ (just define the coordinate patch to be the identity map on H^k.) Then, $\partial H^k = \mathbb{R}^{k-1} \times 0$ by the prior lemma.

Theorem 24.3: Let M be a k-manifold in $\mathbb{R}^n$ of class C^r. If ∂M is nonempty, then ∂M is a $k-1$ manifld without boundary in $\mathbb{R}^n$ of class C^r.

Proof:
Let $p \in \partial M$, and then let $\alpha : U \to V$ be a coordinate patch on M about p. Then U is open in H^k and $p = \alpha(x_0)$ for some $x_0 \in \partial H^k$. By the prior lemma, $\alpha(x) \in \partial M$ for all $x \in U \cap \partial H^k$ and $\alpha(x) \notin \partial M$ for all $x \in U - \partial H^k$. Thus, we know that the restriction of α to $U \cap \partial H^k$ is a bijective map onto the open set $V_0 := V \cap \partial M$ of ∂M.

Now let U_0 be the open set in $\mathbb{R}^{k-1}$ such that $U_0 \times 0 = U \cap \partial H^k$. Then for any $x \in U_0$, define $\alpha_0(x) = \alpha(x, 0)$. Thus, $\alpha_0 : U_0 \to V_0$ is a coordinate patch on ∂M about p.

- It is C^r because so is α.
- $\mathrm{D}\alpha_0(x)$ has rank $k-1$ for all x since $\mathrm{D}\alpha_0$ just consists of the first $k-1$ columns of $\mathrm{D}\alpha(x, 0)$.
- Finally, α_0^{-1} is continuous because it equals the composition of α^{-1} restricted to the set V_0 followed by the projection of $\mathbb{R}^k$ onto it's first $(k-1)$-coordinates (and both of those functions are continuous).

This proves ∂M is a manifold. Also, this shows that p is an interior point of ∂M. So, ∂M has no boundary.

Theorem 24.4: Let $\mathcal{O}$ be an open set in $\mathbb{R}^n$, and let $f : \mathcal{O} \to \mathbb{R}$ be a C^r map. Also let M be the set of points for which $f(x) = 0$ and N be the set of points for which $f(x) \geq 0$. If $M \neq \emptyset$ and $\mathrm{D}f(x)$ has rank 1 for all x in M, then N is an n-manifold in $\mathbb{R}^n$ and $M = \partial N$.

Consequently, a level set of f is a manifold so long as f has no critical points in that level set.

Proof:
Firstly, suppose $p \in N$ with $f(p) > 0$. Then if α is the identity map on the set $U := f^{-1}((0, \infty))$, we have that α is a C^∞ bijective map from the open set U in $\mathbb{R}^n$ to itself such that α has a continuous inverse and a full rank derivative matrix. So, α is a coordinate patch on N.

Meanwhile, suppose $f(p) = 0$. Then since $\mathrm{D}f(p) \neq 0$, at least one partial derivative $\frac{\partial}{\partial x_i} f(p)$ is nonzero. By a sufficient permutation of our standard basis vectors, we can assume $i = n$. So, define $F : \mathcal{O} \to \mathbb{R}^n$ by the equation $F(x) = (x_1, \ldots, x_{n-1}, f(x))$. Thus, F is a C^r map with a nonsingular derivative matrix at p since:

$$\mathrm{D}F = \begin{bmatrix} I_{n-1} & 0 \\ * & \frac{\partial}{\partial x_n} f \end{bmatrix}$$

By the inverse function theorem, we know F is a C^r diffeomorphism from an open neighborhood V of p in $\mathbb{R}^n$ to an open set U of $\mathbb{R}^n$. Furthermore, F carries the open set $V \cap N$ of N onto the open set $U \cap H^n$ of H^n. Therefore, $F^{-1}|_{(U \cap H^n)}$ works as our coordinate patch on N about p.

Finally note that $F(p) \in \partial H^n$. This shows that $M = \partial N$. ■

Corollary 24.5: The n-ball $B^n(a) := \{x : \|x\|_2 \leq a\}$ is a C^∞ n-manifold whose boundary is $S^{n-1} := \{x : \|x\|_2 = a\}$.

Proof:
Consider the function $f(x) = a^2 - (\|x\|_2)^2$.

The next exercise gives us an important tool for constructing manifolds (which makes it kinda shocking that Munkres leaves it as an exercise).

Exercise 24.2 Let $f : \mathbb{R}^{n+k} \to \mathbb{R}^n$ be of class C^r. Let M be the set of all x such that $f(x) = 0$. Assume that M is nonempty and $\mathrm{D}f(x)$ has rank n for all $x \in M$. Then M is a k-manifold in $\mathbb{R}^{n+k}$ without boundary. Furthermore, if N is the set of all x such that $f_1(x) = \cdots = f_{n-1}(x) = 0$ and $f_n(x) \geq 0$, and the matrix $\partial(f_1, \ldots, f_{n-1})/\partial x$ has rank $n-1$ at each point of N, then N is a $k+1$ manifold and $M = \partial N$.

Lemma: Let $m \leq n$ and suppose $f : \mathbb{R}^{n+k} \to \mathbb{R}^m$ is a C^r map such that the matrix $\mathrm{D}f$ has rank m at the point p. Then there are open sets $U, V \subseteq \mathbb{R}^{n+k}$ with $p \in V$, as well as a C^r diffeomorphism $G : U \to V$ with rank $n+k$ satisfying that $f \circ G(x) = \pi_m(x)$ where π_m is a projection from $\mathbb{R}^{n+k}$ to m of it's coordinates (and by applying a suitable permutation of our basis vectors, we can assume that those coordinates are the first m coordinates).

Proof:
Since $\mathrm{D}f$ has rank m at p, we know that the derivative matrix has m linearly independent columns at p, and by a permutation of our standard bases, we can assume those m columns are the first m columns. Therefore, it makes sense to adopt the notation of writing $x = (x^{(1)}, x^{(2)})$ in $\mathbb{R}^{n+k}$ where $x^{(1)} \in \mathbb{R}^m$ and $x^{(2)} \in \mathbb{R}^{n+k-m}$. Also, it makes sense to define the projection $\pi_m(x) = x^{(1)}$.

Now define the function $F : \mathbb{R}^{n+k} \to \mathbb{R}^{n+k}$ by:

$$F(x) = (f_1(x), \ldots, f_m(x), x^{(2)}).$$

Then F is a C^r map with the derivative matrix:

$$\mathrm{D}F = \begin{bmatrix} \partial(f_1, \ldots, f_m)/\partial x^{(1)} & \partial(f_1, \ldots, f_m)/\partial x^{(2)} \\ 0 & I_{n+k-m} \end{bmatrix}$$

Since $\mathrm{D}F$ is full rank at p, by the inverse function theorem there are open sets $U, V \subseteq \mathbb{R}^{n+k}$ with $p \in V$ such that F is a C^r diffeomorphism from V to U with full rank. Taking $G := F^{-1}$, we get:

$$f \circ G(x) = (\pi_m \circ F) \circ G(x) = \pi_m \circ (F \circ F^{-1})(x) = \pi_m(x) = x^{(1)}$$

Given $f = (f_1, \dots, f_{n-1}, f_n)$, we'll denote $\tilde{f} = (f_1, \dots, f_{n-1})$. Now, the rest of the exercise just involves applying the prior lemma three times.

1. Suppose $p \in N - M$ (meaning $f_n(p) > 0$ and $\tilde{f}(p) = 0$). Then using our prior lemma, there are open sets $U, V \subseteq \mathbb{R}^{n+k}$ with $p \in V$ as well as a C^r diffeomorphism $\tilde{G} : U \to V$ with full rank such that $\tilde{f} \circ \tilde{G}(x) = \pi_{n-1}(x)$. Using the continuity of f_n, we can assume $f_n(x) > 0$ for all $x \in V$ by making V sufficiently small.

   Now note that $V \cap N$ is an open subset of N containing p such that:
   $$\begin{aligned} V \cap N &= \tilde{G}(U) \cap \tilde{f}^{-1}(\{0\}) \cap \{x : f_n(x) > 0\} \\ &= \tilde{G}(U) \cap (\pi_{n-1} \circ \tilde{G}^{-1})^{-1}(\{0\}) \cap \{x : f_n(x) > 0\} \\ &= \tilde{G}(U) \cap \tilde{G}(\pi_{n-1}^{-1}(\{0\})) \cap \{x : f_n(x) > 0\} \\ &= \tilde{G}\left(U \cap (0^{n-1} \times \mathbb{R}^{k+1})\right) \cap \{x : f_n(x) > 0\} \\ &= \tilde{G}\left(U \cap (0^{n-1} \times \mathbb{R}^{k+1}) \cap \tilde{G}^{-1}(\{x : f_n(x) > 0\})\right) \end{aligned}$$

   Then since $U \cap \tilde{G}^{-1}(\{x : f_n(x) > 0\})$ is open in $\mathbb{R}^{n+k}$, we can deduce that there is an open set $A \subseteq \mathbb{R}^{k+1}$ with $0^{n-1} \times A = U \cap (0^{n-1} \times \mathbb{R}^{k+1}) \cap \tilde{G}^{-1}(\{x : f_n(x) > 0\})$. And by defining $\alpha(x_1, \dots, x_{k+1}) = \tilde{G}(0^{n-1}, x_1, \dots, x_{k+1})$, we have that α is a bijective C^r map from the open set A of $\mathbb{R}^{k+1}$ to $V \cap N$ with rank $k+1$ and a continuous inverse. Hence, α is
   a coordinate patch on N about p.

2. Suppose $p \in M$ and only assume the part of the exercise statement that comes before the word "furthermore". Then let $U, V \subseteq \mathbb{R}^{n+k}$ be open sets with $p \in V$, and let $G : U \to V$ be a C^r diffeomorphism with full rank satisfying that $f \circ G(x) = \pi_n(x)$. Then $V \cap M$ is an open subset of M containing p such that:
   $$\begin{aligned} V \cap M &= G(U) \cap f^{-1}(\{0\}) \\ &= G(U) \cap (\pi_n \circ G^{-1})^{-1}(\{0\}) = G(U) \cap G(\pi_n^{-1}(\{0\})) \\ &= G(U \cap (0^n \times \mathbb{R}^k)) \end{aligned}$$

   Now we know there is some open set $A \subseteq \mathbb{R}^k$ such that $0^n \times A = U \cap (0^n \times \mathbb{R}^k)$. Therefore, by defining $\alpha(x_1, \dots, x_k) = \tilde{G}(0^n, x_1, \dots, x_k)$, we have that α is a bijective C^r map from the open set A of $\mathbb{R}^{k+1}$ to $V \cap M$ with rank k and a continuous inverse. Hence, α is a coordinate patch on M about p.

3. Finally, suppose $p \in M$ and this time assume the entire hypothesis of the exercise. Also let $G : U \to V$ be as in the prior part. Now:
   $$\begin{aligned} V \cap N = G(U) \cap f^{-1}(\{0\}) &= G(U) \cap G(\pi_n^{-1}(0^{n-1} \times [0, \infty))) \\ &= G(U \cap (0^{n-1} \times [0, \infty) \times \mathbb{R}^k)) \end{aligned}$$

   Now if τ is the function permuting the first and $(k+1)$th basis vectors, then we know there is some open set $A \subseteq H^{k+1}$ such that $0^{n-1} \times \tau(A) = U \cap (0^{n-1} \times [0, \infty) \times \mathbb{R}^k)$. So, define $\alpha(x_1, \dots, x_{k+1}) = G(0^{n-1}, x_{k+1}, x_1, \dots, x_k)$. Then α is a bijective C^r map from the open set A of H^{k+1} to $V \cap N$ with rank $k+1$ and a continuous inverse. Hence, α is a coordinate patch on N about p.

Also, if $x \in U$ satisfies that $\alpha(x) = p$, then since $f \circ G = \pi_n$, we know that $f(G(x)) = f(p) = 0$. So $x \in \partial H^{k+1}$. This proves p is on the boundary of N. ■

Note from 8/28/2025: How the fuck did I just realize this is just implicit function theorem.

7/28/2025

Here's a fun application of the prior exercise.

Exercise 24.6: Let $\mathcal{O}(n)$ denote the set of all orthogonal n by n matrices, considered as a subspace of $\mathbb{R}^N$ where $N = n^2$. Show that $\mathcal{O}(n)$ is a compact $\binom{n}{2}$-manifold of class C^∞ in $\mathbb{R}^N$ without boundary.

Firstly, define $F : \mathbb{R}^N \to \mathbb{R}^N$ by $F(\mathbf{A}) = \mathbf{A}\mathbf{A}^T - I_n$. Also define $f : \mathbb{R}^N \to \mathbb{R}^{\frac{1}{2}n(n+1)}$ such that $f(\mathbf{A})$ is the vector containing just the elements of $F(\mathbf{A})$ on or above the main diagonal. Then it is fairly clear that $F(\mathbf{A}) = 0$ if and only if $f(\mathbf{A}) = 0$ if and only if $\mathbf{A} \in \mathcal{O}(n)$.

Next note that if $\mathbf{A} = [x_{i,j}]$, then the (i,j)th. column of the derivative matrix $\mathrm{D}F$ is:

$$\frac{\partial}{\partial x_{i,j}} F(\mathbf{A}) = \left(\frac{\partial}{\partial x_{i,j}}\mathbf{A}\right)\mathbf{A}^T + \mathbf{A}\left(\frac{\partial}{\partial x_{i,j}}\mathbf{A}\right)^T$$

$$= \begin{bmatrix} & \mathbf{0}^{(i-1)\times n} & \\ x_{1,j} & \cdots & x_{n,j} \\ & \mathbf{0}^{(n-i)\times n} & \end{bmatrix} + \begin{bmatrix} & \mathbf{0}^{(i-1)\times n} & \\ x_{1,j} & \cdots & x_{n,j} \\ & \mathbf{0}^{(n-i)\times n} & \end{bmatrix}^T$$

$$= \begin{bmatrix} & & x_{1,j} & & \\ & & \vdots & & \\ x_{1,j} & \cdots & 2x_{i,j} & \cdots & x_{n,j} \\ & & \vdots & & \\ & & x_{n,j} & & \end{bmatrix}$$

Meanwhile, the (i,j)th. column of $\mathrm{D}f$ is just the vector with the elements of $\mathrm{D}F$ on or above the diagonal. From this, it is clear that f and F are C^∞ maps.

Taking a different view, if $k \leq \ell$, then the (k,ℓ)th. row of $\mathrm{D}f$ is:

$$\left(\begin{bmatrix} x_{\ell,1} & \cdots & x_{\ell,n} \end{bmatrix} (k\text{th. row})\right) + \left(\begin{bmatrix} x_{k,1} & \cdots & x_{k,n} \end{bmatrix} (\ell\text{th. row})\right)$$

Now let $v_{k,\ell}$ denote the (k,ℓ)th. row of $\mathrm{D}f$; let r_i denote the ith. row of $\mathbf{A}$; and let $\delta_{i,j}(x,y)$ equal 1 if $(x,y) = (i,j)$ and equal 0 otherwise. Then supposing $k_1 \leq \ell_1$ and $k_2 \leq \ell_2$, we have that:

$$v_{k_1,\ell_1} \cdot v_{k_2,\ell_2} = \delta_{k_1,k_2}(r_{\ell_1} \cdot r_{\ell_2}) + \delta_{k_1,\ell_2}(r_{\ell_1} \cdot r_{k_2}) + \delta_{\ell_1,k_2}(r_{k_1} \cdot r_{\ell_2}) + \delta_{\ell_1,\ell_2}(r_{k_1} \cdot r_{k_2})$$

If $\mathbf{A}$ is orthogonal, then this simplifies to:

$$\begin{aligned} v_{k_1,\ell_1} \cdot v_{k_2,\ell_2} &= \delta_{k_1,k_2}\delta_{\ell_1,\ell_2} + \delta_{k_1,\ell_2}\delta_{\ell_1,k_2} + \delta_{\ell_1,k_2}\delta_{k_1,\ell_2} + \delta_{\ell_1,\ell_2}\delta_{k_1,k_2} \\ &= 2\delta_{k_1,k_2}\delta_{\ell_1,\ell_2} + 2\delta_{k_1,\ell_2}\delta_{\ell_1,k_2} \end{aligned}$$

There are two cases where this dot product will be nonzero. The first case is if $k_1 = k_2$ and $\ell_1 = \ell_2$. Meanwhile, the second case is if $k_1 = \ell_2$ and $\ell_1 = k_2$. However, since we are requiring that $k_1 \leq \ell_1$ and $k_2 \leq \ell_2$, the second case actually implies the first case.

This proves that the rows of $\mathrm{D}f$ actually form an orthogonal set of vectors in $\mathbb{R}^N$. Hence, f has full rank on $\mathcal{O}(n)$.

It now follows from the previous exercise that $\mathcal{O}(n)$ is a manifold without boundary in $\mathbb{R}^N$. It's dimension will be $n^2 - \frac{(n+1)n}{2} = \frac{n^2}{2} - \frac{n}{2} = \binom{n}{2}$. Meanwhile, to see that the manifold is compact, note firstly that it is bounded by the set $\{x \in \mathbb{R}^N : \|x\|_\infty \leq 1\}$. Also, the points in the manifold are given by the set $f^{-1}(\{0\})$, and that set is closed since f continuous and $\{0\}$. ■

7/29/2025

I'm gonna finish the current chapter of Munkres and then switch to a different book to learn about differential forms. For today, my agenda is to define integration of scalar-valued functions on general manifolds in $\mathbb{R}^n$.

Let M be a k-manifold of class C^r in $\mathbb{R}^n$ (with $r \geq 1$ and $k \leq n$). Then recall that we already defined integration on M if M is parametrized by a single coordinate patch from an open set of $\mathbb{R}^k$. Hopefully, it's also clear to see that our previous definition works if our coordinate patch is from an open set of H^k. Also, by theorem 24.1 plus the theorem at the bottom of page 79 of this journal, we now know that our definition of the integral is independent of the parametrization we use.

Note from 7/31/2025: Actually I haven't yet showed that the parametrization is independent if the domain of that parametrization is open in H^k but not $\mathbb{R}^k$.

In general though, M probably can't be parametrized by just one coordinate patch. So, we instead bodge our definition using a partition of unity.

<u>Lemma:</u> There is a countable set $\{\alpha_n : U_n \to V_n\}_{n \in \mathbb{N}}$ of coordinate patches on M such that $\bigcup_{n \in \mathbb{N}} V_n = M$.

Proof:

1. Note that topologically speaking, M is a second countable LCH space.

   The fact that M is second countable and Hausdorff is just a consequence of the fact that $\mathbb{R}^n$ is both of those things and M is equipped with the subspace topology.

To show that M is locally compact, note that if $p \in M$, then there is a homeomorphism α from some open set U in $\mathbb{R}^k$ or H^k to an open set $V \subseteq M$ containing p. Then given the $x \in U$ satisfying that $\alpha(x) = p$, there is a compact set $K \subseteq U$ with $x \in K$.

If U is open in $\mathbb{R}^k$, then it's obvious that K exists. Meanwhile, if U is open in H^k, then consider picking a U' which is open in $\mathbb{R}^k$ and satisfies that $U' \cap H^k = U$. Then we know there is a compact set K' such that $x \in K' \subseteq U'$. And since H^k is closed, we can set $K = K' \cap H^k$ and know K is compact.

Next, by Urysohn's lemma, there is a precompact open set V such that $K \subseteq V \subseteq \overline{V} \subseteq U$. Hence, we have that $\alpha(\overline{V})$ is a compact subset of M containing p. Also $\alpha(V)$ is an open subset of $\alpha(\overline{V})$ which contains p. Hence, $\alpha(\overline{V})$ is a compact neighborhood of p.

2. In turn, we know that M is σ-compact. From that hopefully it is obvious how we can get a countable covering of coordinate patches over M. ■

<u>Lemma:</u> Let $\{\alpha_a : U_a \to V_a\}_{a \in A}$ be a collection of coordinate patches on M such that $\bigcup_{a \in A} V_a = M$. Then there exists a countable collection of C^∞ functions $\{\phi_i : \mathbb{R}^n \to \mathbb{R}\}_{i \in \mathbb{N}}$ satisfying that:

- $\phi_i(x) \geq 0$ for all $x \in \mathbb{R}^n$ and $i \in \mathbb{N}$.
- Each $p \in M$ has a neighborhood in M on which only finitely many ϕ_i are nonzero.
- $\sum_{i \in \mathbb{N}} \phi_i(p) = 1$ for all $p \in M$.
- For each $i \in \mathbb{N}$, there is some $a \in A$ such that $\mathrm{supp}(\phi_i) \cap M \subseteq V_a$.

In the future I will just refer to $\{\phi_i\}_{i \in \mathbb{N}}$ as being a partition of unity on M subordinate to our collection of coordinate patches.

Proof:
For each coordinate patch α_a, choose an open set $V_a' \subseteq \mathbb{R}^n$ such that $V_a' \cap M = V_a$. Then by theorem 16.3, we get our desired partition which is subordinate to $\{V_a'\}_{a \in A}$.

<u>Definition:</u> Let $(\phi_i)_{i \in \mathbb{N}}$ be a partition of unity on M subordinate to the collection of coordinate patches $\{\alpha_i : U_i \to V_i\}_{i \in \mathbb{N}}$ which cover M. Without loss of generality, suppose $\mathrm{supp}(\phi_i) \cap M \subseteq V_i$. Then, we define a Borel measure on M by:

$$V(E) := \sum_{i=1}^{\infty} \int_{E \cap V_i} \phi_i \mathrm{d}V_{\alpha_i} = \sum_{i=1}^{\infty} \int_{\alpha_i^{-1}(E \cap V_i)} (\phi_i \circ \alpha_i) V(\mathrm{D}\alpha_i) \mathrm{d}m$$

From here it's pretty obvious that $V(\emptyset) = 0$ and that V is countably additive. Also, similarly to before we can then deduce that:

$$\int_M f \mathrm{d}V = \sum_{i=1}^{\infty} \int_{V_i} f \phi_i \mathrm{d}V_{\alpha_i} = \sum_{i=1}^{\infty} \int_{U_i} (f \phi_i \circ \alpha_i) V(\mathrm{D}\alpha_i) \mathrm{d}m$$

Now our first challenge is to show that this definition is independent of our choice of coordinate patches and partition of unity.

Let $(\psi_i)_{i\in\mathbb{N}}$ be another partition of unity on M subordinate to another collection of coordinate patches $\{\beta_i : U_i' \to V_i'\}_{i\in\mathbb{N}}$ which cover M (and like before suppose $\mathrm{supp}(\psi_i) \cap M \subseteq V_i'$).

Now importantly, by our prior results about integration on manifolds parametrized by single coordinate patches, we know that $\int_{V_i\cap V_j'} f \mathrm{d}V_{\alpha_i} = \int_{V_i\cap V_j'} f \mathrm{d}V_{\beta_j}$ for all integrable f on $V_i \cap V_j'$. Therefore, if $E \in \mathcal{B}_M$, we have that:

$$\begin{aligned}
\sum_{i=1}^{\infty} \int_{V_i} \phi_i \chi_E \mathrm{d}V_{\alpha_i} &= \sum_{i=1}^{\infty} \int_{V_i} \phi_i \chi_E \left(\sum_{j=1}^{\infty} \psi_j \chi_{V_j'}\right) \mathrm{d}V_{\alpha_i} \\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \int_{V_i\cap V_j'} \phi_i\psi_j\chi_E \mathrm{d}V_{\alpha_i} \\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \int_{V_i\cap V_j'} \phi_i\psi_j\chi_E \mathrm{d}V_{\beta_j} \\
&= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \int_{V_i\cap V_j'} \phi_i\psi_j\chi_E \mathrm{d}V_{\beta_j} \\
&= \sum_{j=1}^{\infty} \int_{V_j'} \psi_j \chi_E \left(\sum_{i=1}^{\infty} \phi_i \chi_{V_i}\right) \mathrm{d}V_{\beta_j} = \sum_{j=1}^{\infty} \int_{V_j'} \psi_j \chi_E \mathrm{d}V_{\beta_j}
\end{aligned}$$

Side note: we can swap the order of the sums in the second to last line by applying Fubini-Tonelli's theorem.

This shows that our definition of the measure of E is independent of our patches and partition of unity.

One other thing to note is that if M can be parametrized by a single coordinate patch, then this result shows that our definition for parametrized manifolds is compatible with this one. After all, just take $\phi_i = \delta_{1,i}$ (where δ is the Kronecker delta function), and let α_i be the same coordinate patch parametrizing all of our manifold for all i.

7/31/2025

Ok. So in the previous journal entry, I was sorta following along with Munkres while only slightly modifying his thoerems. However, after some thought, I actually want to take a completely different approach to defining a measure on manifolds which is more in line with math 240A. So, I'm going to completely depart from Munkres and do a bunch of stuff entirely on my own (i.e. not based on any people's books).

Let M be a k-manifold of class C^r in $\mathbb{R}^n$ (with $r \geq 1$ and $k \leq n$). Then recall from the last journal entry that M is an LCH second countable topological space. Hence, it follows that M is σ-compact, and from there it follows easily that given any arbitrary collection $\{\alpha_a : U_a \to V_a\}_{a \in A}$ of coordinate patches covering $E \subseteq M$, there is a countable subcover over E.

Also, recall that if $\alpha : U \to V$ is a coordinate patch on M, then we can use the formula $\int_{\alpha^{-1}(E)} V(\mathrm{D}\alpha)\mathrm{d}m$ to calculate the "surface-volume" of any Borel set $E \subseteq V$.

Note that if $\alpha_1 : U_1 \to V_1$ and $\alpha_2 : U_2 \to V_2$ are both coordinate patches on M, $E \subseteq V_1 \cap V_2$ is Borel, and U_1, U_2 are open in $\mathbb{R}^k$, then as we already noted, by theorem 24.1 plus the theorem at the bottom of page 79 of this journal:

$$\int_{\alpha_1^{-1}(E)} V(\mathrm{D}\alpha_1)\mathrm{d}m = \int_{\alpha_2^{-1}(E)} V(\mathrm{D}\alpha_2)\mathrm{d}m.$$

That said, before continuing on I want to show that this equivalence still holds if U_1 or U_2 is open in H^k but not $\mathbb{R}^k$.

To start off, by restricting α_1 and α_2 to their preimages of $V_1 \cap V_2$, we can without loss of generality assume that $V_1 = V_2$. This is important because it makes it so that $\alpha_2^{-1} \circ \alpha_1(U_1) = U_2$ and $\alpha_1^{-1} \circ \alpha_2(U_2) = U_1$.

Now, its impossible for U_1 to be open in $\mathbb{R}^k$ but not U_2, or vice versa. After all, if U_1 is open in $\mathbb{R}^k$, we must have that the transition function $\alpha_2^{-1} \circ \alpha_1 : U_1 \to U_2$ maps U_1 to an open set in $\mathbb{R}^k$. Analogous reasoning using the transition function $\alpha_1^{-1} \circ \alpha_2$ works if U_2 is open in $\mathbb{R}^k$.

Now suppose both U_1 and U_2 are open only in H^k. Then, we can easily see that $W_i := U_i - \partial H^k$ is the interior of U_i in $\mathbb{R}^k$ and that $m(U_i - W_i) = 0$ for both i. Also, since the transition functions map open sets of $\mathbb{R}^k$ to open sets of $\mathbb{R}^k$, we have that $\alpha_2^{-1} \circ \alpha_1(W_1) \subseteq W_2$ and $\alpha_1^{-1} \circ \alpha_2(W_2) \subseteq W_1$. This is enough to say that the transition functions restricted to W_1 and W_2 are a diffeomorphism.

Now finally, by applying the lemma at the bottom of page 79, we have for all functions f which are integrable over V:

$$\begin{aligned}\int_{U_1}(f \circ \alpha_1)V(\mathrm{D}\alpha_1)\mathrm{d}m &= 0 + \int_{W_1}(f \circ \alpha_1)V(\mathrm{D}\alpha_1)\mathrm{d}m \\ &= 0 + \int_{W_2}(f \circ \alpha_2)V(\mathrm{D}\alpha_2)\mathrm{d}m = \int_{U_2}(f \circ \alpha_2)V(\mathrm{D}\alpha_2)\mathrm{d}m\end{aligned}$$

Set $f = \chi_E$ and we are done.

We take the following steps to define a measure on M:

(1) Defining an algebra or even a ring of sets is too much to ask for right now. But, we can at least define a collection of sets $\mathcal{A}$ satisfying that if $A \in \mathcal{A}$ and $E \subseteq A$ is a Borel subset of M, then $E \in \mathcal{A}$. Specifically, let $\mathcal{A}$ be the collection of Borel sets $A \subseteq M$ for which there exists a coordinate patch $\alpha : U \to V$ with $A \subseteq V$.

(2) Next, we define a "premeasure" μ_0 on $\mathcal{A}$. Specifically, for each $A \in \mathcal{A}$, define $\mu_0(A) = \int_{\alpha^{-1}(A)} V(\mathrm{D}\alpha)\mathrm{d}m$ where $\alpha : U \to V$ is some coordinate patch with $A \subseteq V$.

Importantly, even though μ_0 isn't a proper premeasure according to Folland's definition since $\mathcal{A}$ isn't actually an algebra, it is still the case that μ_0 and $\mathcal{A}$ are structured enough that the following is easily seen to hold:

- $\mu_0(\emptyset) = 0$
- If $A, B \in \mathcal{A}$ satisfy $A \subseteq B$. then $\mu_0(A) \leq \mu_0(B)$.
- If $(A_j)_{j\in\mathbb{N}}$ is a sequence of disjoint sets in $\mathcal{A}$ with $\bigcup_{j\in\mathbb{N}} A_j \in \mathcal{A}$, then:

$$\mu_0(\bigcup_{j\in\mathbb{N}} A_j) = \sum_{j=1}^{\infty} \mu_0(A_j).$$

(3) Now, we use μ_0 to define an outer measure on $\mathcal{A}$. For any $E \subseteq M$, define:

$$\mu^*(E) = \inf\left\{ \sum_{i=1}^{\infty} \mu_0(A_i) : A_i \in \mathcal{A} \text{ for all } i \text{ and } E \subseteq \bigcup_{i\in\mathbb{N}} A_i \right\}$$

Since any subset of M can be covered by countably many coordinate patches, we know that $\mu^*(E)$ is well defined. The rest of the proof that μ^* is an outer measure is then identical to proposition 1.10 of Folland (which is at the top of page 21 of my Latex math 240a notes).

(4) <u>Lemma:</u> $\mu^*|_{\mathcal{A}} = \mu_0$ and every set in $\mathcal{A}$ is μ^* measurable.

Proof:
To prove the first claim, suppose $E \in \mathcal{A}$ and $(A_m)_{m\in\mathbb{N}}$ is a sequence of sets in $\mathcal{A}$ covering E. It's trivial that $\mu^*(E) \leq \mu_0(E)$. Meanwhile let $B_1 = E \cap A_1$ and $B_m = E \cap A_m - \bigcup_{j=1}^{m-1} A_j$. Then the B_m are each disjoint Borel subsets of $E \in \mathcal{A}$ whose union is all of E. Hence all the B_m are in $\mathcal{A}$ and we have:

$$\mu_0(E) = \mu_0(\bigcup_{m\in\mathbb{N}} B_m) = \sum_{m\in\mathbb{N}} \mu_0(B_m) \leq \sum_{m\in\mathbb{N}} \mu_0(A_m)$$

This shows that $\mu_0(E) \leq \mu^*(E)$, thus proving the first claim.

To show the second claim, suppose $A \in \mathcal{A}$, $E \subseteq X$, and $\varepsilon > 0$. Then there exists a sequence $(B_j)_{j\in\mathbb{N}}$ of sets in $\mathcal{A}$ such that $E \subseteq \bigcup_{j=1}^{\infty} B_j$ and $\sum_{j=1}^{\infty} \mu_0(B_j) \leq \mu^*(E) + \varepsilon$. Importantly, $(B_j \cap A)_{j\in\mathbb{N}}$ and $(B_j - A)_{j\in\mathbb{N}}$ are both sequences of sets in $\mathcal{A}$ covering $E \cap A$ and $E - A$ respectively. Also, $\mu_0(B_j) = \mu_0(B_j \cap A) + \mu_0(B_j - A)$ since $B_j \cap A$ and $B_j - A$ are disjoint sets in $\mathcal{A}$ whose union B_j is also in $\mathcal{A}$. Therefore, we have that:

$$\begin{aligned}\mu^*(E) + \varepsilon &\geq \sum_{j=1}^{\infty} \mu_0(B_j) \\ &= \sum_{j=1}^{\infty} \mu_0(B_j \cap A) + \sum_{j=1}^{\infty} \mu_0(B_j - A) \geq \mu^*(E \cap A) + \mu^*(E - A)\end{aligned}$$

Taking $\varepsilon \to 0$, we have that $\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E - A)$ for all $E \subseteq M$. Thus A is μ^* measurable. ■

(5) We now know by Carathéodory's theorem that if $\mathcal{N}$ is the σ-algebra of μ^*-measurable sets, then $\mu := \mu^*|_{\mathcal{N}}$ is a complete measure on $\mathcal{N}$. Furthermore, $\mathcal{A} \subseteq \mathcal{N}$ with $\mu|_{\mathcal{A}} = \mu_0$.

(6) Lemma: If $\mathcal{B}_M$ is the collection of Borel sets on our manifold M, then $\mathcal{B}_M \subseteq \mathcal{N}$. Hence, we can restrict μ to be a Borel measure if we wanted.

Proof:
If $E \in \mathcal{B}_m$, then let $(A_j)_{j\in\mathbb{N}}$ be a countable covering of E consisting of sets in $\mathcal{A}$. Then $E \cap A_j \in \mathcal{A} \subseteq \mathcal{N}$ for all j and $\bigcup_{j\in\mathbb{N}}(E \cap A_j) = E$. So $E \in \mathcal{N}$.

(7) The next thing we want to do now is show that μ is σ-finite. One reason we want to do this is so that we can apply theorems such as Fubini-Tonelli and Radon-Nikodym. Another reason is so that (as I'll show in the next step) μ is guarenteed to be the unique measure on $(M, \mathcal{N})$ which preserves our definition of the measure of a manifold parametrized by a single coordinate patch. Hence, this construction agrees with what I was doing back when I was loosely following Munkres.

To start, we will show that every point $p \in M$ has a neighborhood $A \in \mathcal{N}$ with finite measure.

Let $\alpha : U \to V$ be a coordinate patch on M about p. Then, given $x \in U$ satisfying that $\alpha(x) = p$, let $K \subseteq U$ be a compact neighborhood of x. Since α is a homeomorphism, it is clear that $A := \alpha(K)$ is a compact set containing p in its interior. Hence, $A \in \mathcal{B}_M \subseteq \mathcal{N}$, A is a neighborhood of p and:

$$\mu(A) = \int_K V(\mathrm{D}\alpha)\mathrm{d}m$$

Now $V(\mathrm{D}\alpha)$ is continuous. So by the extreme value theorem, there exists some $C \geq 0$ such that:

$$\int_K V(\mathrm{D}\alpha)\mathrm{d}m \leq \int_K C\mathrm{d}m = Cm(K)$$

And since $m(K) < \infty$, we've shown that $\mu(A) < \infty$.

Now since M is σ-compact, if we pick a set A_p for each $p \in M$ using the reasoning above, then there is a countable subcovering of the A_p over M. This proves that M is σ-finite. ■

(8) Lemma: If $\nu : \mathcal{N} \to [0, \infty]$ is another measure on $(M, \mathcal{N})$ satisfying that $\nu|_{\mathcal{A}} = \mu_0$, then $\nu = \mu$.

The proof of this is almost identical to that of theorem 1.14 in Folland (bottom of page 24 on my Latex math 240a notes).

The one difference is that if $(A_j)_{j\in\mathbb{N}} \subseteq \mathcal{A}$ and $A = \bigcup_{j\in\mathbb{N}} A_j$, then $\nu(A) = \mu(A)$ because $A_m - \bigcup_{j=1}^{m-1} A_j \in \mathcal{A}$ for all m and hence:

$$\nu(A) = \sum_{m=1}^{\infty} \nu(A_m - \textstyle\bigcup_{j=1}^{m-1} A_j) = \sum_{m=1}^{\infty} \mu(A_m - \textstyle\bigcup_{j=1}^{m-1} A_j) = \mu(A). \blacksquare$$

Why did I take this pivot? The reason is that now all of Munkres theorems from chapter 25 of his book are obvious including the final theorem about how one would in practice calculate an integral on M. Also, I got to show that this construction is universal in a sense.

I'm now going to switch over to following Guillemin's Differential Forms since I was recommended this book by another tutor. Once again, I'm not going to perfectly follow the book. But I will be using the book as a loose guide.

Tensors:
Let V be an n-dimensional vector space over a field F. Let V^k be the set of all k-tuples of elements of V. Then a function $T : V^K \to F$ is said to be linear in its ith. variable if for all $u, v_1, \ldots, v_k \in V$ and $a, b \in F$, we have that:

$$\begin{aligned} &T(v_1, \ldots, v_{i-1}, av_i + bu, v_{i+1}, \ldots, v_k) \\ &\qquad = aT(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_k) + bT(v_1, \ldots, v_{i-1}, u, v_{i+1}, \ldots, v_k). \end{aligned}$$

If T is linear in all it's variables, we say T is k-linear and that T is a k-tensor.

Given $k \geq 1$, we shall denote $\mathcal{L}^k(V)$ to be the set of all k-tensors from V. Also, we shall denote $\mathcal{L}^0(V) \coloneqq F$. Note that we can add together and scale k-linear maps to get more k-linear maps. Thus, $\mathcal{L}^k(V)$ is a vector space.

Side note: $\mathcal{L}^1(V)$ is also called the dual space of V and denoted V^*.

Sorry but my apartment mate sent me down a rabbithole that led me to writing a proof on the following pages. I'll come back to tensors right after I'm done writing the following...

8/5/2025

So for context, my apartment mate was reading through a paper by Joel Spencer (who is famous for using probabilistic methods to study combinatorics and graph theory). This paper, which is titled Balancing Unit Vectors, gives a pretty neat existence proof using probability that given any finite collection $\{u_1, \ldots, u_m\}$ of vectors in $\mathbb{R}^n$ with $\|u_i\|_2 \leq 1$ for all i, there exists coefficients $\varepsilon_1, \ldots, \varepsilon_m \in \{-1, +1\}$ such that: $\|\varepsilon_1 u_1 + \ldots + \varepsilon_m u_m\|_2 \leq \sqrt{n}$. However, my apartment mate was having trouble with the following part of the paper and thus asked for my help with it:

> The theorem quickly follows. A linear algebra argument yields $\alpha_1, \ldots, \alpha_m$ satisfying $\alpha_1 u_1 + \cdots + \alpha_m u_m = 0$ such that all $|\alpha_i| \leqslant 1$ and $\alpha_i = \pm 1$ for all but at most n i's. Reordering vectors for convenience we have

Our issue was figuring out what linear algebra argument Spencer was fucking using. After thinking about it for three days while I was grading tests, I finally came up with a proof:

Theorem: Let $u_1, \ldots, u_m$ be vectors in $\mathbb{R}^n$. Then there are constants $a_1, \ldots, a_m$ satisfying $a_1 u_1 + \ldots + a_m u_m = 0$ such that all $|a_i| \leq 1$ and $a_i = \pm 1$ for all but at most n i's.

Proof:

We'll proceed by an inductive argument. For our base case, let $u_{m+1} := 0$. That way $\sum_{i=1}^m 0 u_i = u_{m+1}$. Next, suppose that for $n+1 \leq k \leq m$, we've shown that there are constants $b_1, \ldots, b_k \in [-1, 1]$ and $\varepsilon_{k+1}, \ldots, \varepsilon_{m+1} \in \{-1, +1\}$ satisfying that:

$$b_1 u_1 + \ldots + b_k u_k = \varepsilon_{k+1} u_{k+1} + \ldots + \varepsilon_{m+1} u_{m+1}.$$

If we consider the matrix $U := [u_1 \quad \cdots \quad u_k]$, then letting $b = (b_1, \ldots, b_k)$ we have that any $a = (a_1, \ldots, a_k) \in b + \ker(U)$ will satisfy that:

$$a_1 u_1 + \ldots + a_k u_k = \varepsilon_{k+1} u_{k+1} + \ldots + \varepsilon_{m+1} u_{m+1}.$$

Since $k > n$, we know that $\ker(U)$ is nontrivial and hence unbounded. At the same time, $\|b\|_\infty \leq 1$. Hence, by the connectedness of $b + \ker(U)$ and the continuity of the ∞-norm, we know there is some $a \in b + \ker(U)$ with $\|a\|_\infty = 1$. After reordering our u_i, this is the same as saying that there exists $a_1, \ldots, a_{k-1} \in [-1, 1]$ and $\varepsilon_k = \pm 1$ such that:

$$a_1 u_1 + \ldots + a_{k-1} u_{k-1} + \varepsilon_k u_k = \varepsilon_{k+1} u_{k+1} + \ldots + \varepsilon_{m+1} u_{m+1}.$$

Subtract both sides by $\varepsilon_k u_k$ to complete the induction step.

After induction, we will eventaully get constants $a_1, \ldots, a_n \in [-1, 1]$ and $\varepsilon_{n+1}, \ldots, \varepsilon_{m+1}$ equal to ± 1 such that:

$$a_1 u_1 + \ldots + a_n u_n = \varepsilon_{n+1} u_{n+1} + \ldots + \varepsilon_m u_m + \varepsilon_{m+1} u_{m+1}.$$

Move everything over to one side of the equation and forget about the u_{m+1} and we have proven what we wanted.

8/6/2025

Now I'm going to go back to studying tensors.

A multi-index of length k is a k-tuple $I = (i_1, \dots, i_k)$ of integers. We say I is a multi-index of n if each i is between 1 and n. Now let $u_1, \dots, u_n$ be a basis of V. For $T \in \mathcal{L}^k(V)$, write $T_I \coloneqq T(u_{i_1}, \dots, u_{i_k})$ for every multi-index I of n of length k.

Proposition 1.3.7: The T_I uniquely determine T.

Proof:
When $k = 1$, T is just a linear map and we've already proven this for linear maps. For $k > 1$, we proceed by induction. For each i, define $T_i \in L^{k-1}(V)$ by:

$$(v_1, \dots, v_{k-1}) \mapsto T(v_1, \dots, v_{k-1}, u_i).$$

Then for $v = c_1 u_1 + \dots + c_n u_n$, we have:

$$T(v_1, \dots, v_{k-1}, v) = \sum_{i=1}^{n} c_i T_i(v_1, \dots, v_{k-1})$$

Also, by induction each T_i is uniquely determined by the coefficients T_I where I is a multi-index of n of length k with a final index equal to i.

Side note: We can see that if $C = (c_{i,j}) \in \mathcal{M}_{n \times k}(F)$ is the matrix satisfying that $v_j = \sum_{i=1}^{n} c_{i,j} u_i$ for each j, then:

$$T(v_1, \dots, v_k) = \sum_{I=(i_1,\dots,i_k)} \left(\prod_{j=1}^{k} c_{i_j,j} \right) T_I$$

Given two tensors: $T_1 \in \mathcal{L}^k(V)$ and $T_2 \in \mathcal{L}^\ell(V)$, we define the tensor product of T_1 and T_2 as:

$$(T_1 \otimes T_2)(v_1, \dots, v_k, v_{k+1}, \dots, v_{k+\ell}) = T_1(v_1, \dots, v_k) T_2(v_{k+1}, \dots, v_{k+\ell})$$

Note that:

- $\otimes$ is associative.
- If T_1 or T_2 is a 0-tensor, then $\otimes$ is just scalar multiplication.
- If $\lambda \in F$, $T_1 \in \mathcal{L}^k(V)$, and $T_2 \in \mathcal{L}^\ell(V)$, then $\lambda(T_1 \otimes T_2) = (\lambda T_1) \otimes T_2 = T_1 \otimes (\lambda T_2)$.
- If $T_1, T_2 \in \mathcal{L}^k(V)$ and $T_3 \in \mathcal{L}^\ell(V)$, then $(T_1 + T_2) \otimes T_3 = (T_1 \otimes T_3) + (T_2 \otimes T_3)$. Also $T_3 \otimes (T_1 + T_2) = (T_3 \otimes T_1) + (T_3 \otimes T_2)$

Suppose $T \in \mathcal{L}^k(V)$ and there exists $\ell_1, \dots, \ell_k \in V^* = \mathcal{L}^1(V)$ such that $T = \ell_1 \otimes \dots \otimes \ell_k$. Then we say T is decomposable.

If $u_1, \dots, u_n$ is a basis for V, then we can define the dual basis: $u_1^*, \dots, u_n^*$ for V^* such that if $v = \sum_{i=1}^{n} c_i u_i$, then $u_j^*(v) = c_j$.

To prove that this is a basis, first note that if $f \in V^*$ and $\lambda_i = f(u_i)$, then we can easily see that for any $v = \sum_{i=1}^{n} c_i u_i$, then:

$$f(v) = \sum_{i=1}^n c_i f(u_i) = \sum_{i=1}^n \lambda_i c_i = \sum_{i=1}^n \lambda_i u_i^*(v).$$

It follow that $u_1^*, \ldots, u_n^*$ span all of V^*. Also, consider if $f = \sum_{i=1}^n \lambda_i u_i^* = 0$ where all the $\lambda_i \in F$. Then we know that $0 = f(u_j) = \lambda_j$ for all j. This shows that $u_1^*, \ldots, u_n^*$ are linearly independent.

Tangent: if V, W are vector fields over F and $A : V \to W$ is a linear map, then we define the transpose $A^\dagger : W^* \to V^*$ of A by $f \mapsto A^\dagger(f) = f \circ A$.

Claim 1.2.15: Suppose $e_1, \ldots, e_m$ is a basis of V and $u_1, \ldots, u_n$ is a basis of W. Then if $A = (a_{i,j})$ is the the matrix of A with respect to the given bases, we have that the matrix of $A^\dagger$ with respect to the dual bases of V and W is given by $(a_{j,i})$.

Proof:
Suppose $(c_{j,i})$ is the matrix representation of $A^\dagger$. Then:

$$A^\dagger(u_i^*)(e_j) = u_i^*(A(e_j)) = u_i^*(\textstyle\sum_{k=1}^n a_{k,j} u_k) = a_{i,j}$$

Simultaneously:

$$A^\dagger(u_i^*)(e_j) = \textstyle\sum_{k=1}^m c_{k,i} e_k^*(e_j) = c_{j,i}$$

Let $u_1, \ldots, u_n$ be a basis of V and $u_1^*, \ldots, u_n^*$ be the corresponding dual basis of V^*. Then for every multi-index $I = (i_1, \ldots, i_k)$ of n of length k, define:

$$u_I^* = u_{i_1}^* \otimes \cdots \otimes u_{i_k}^*.$$

Note that if $J = (j_1, \ldots, j_k)$ is another multi-index of n of length k, then $u_I^*(u_{j_1}, \ldots, u_{j_k}) = \delta_{I,J}$ where δ is the Kronecker delta function.

Theorem 1.3.13: The k-tensors u_I^* form a basis for $\mathcal{L}^k(V)$.

Proof:
Suppose $T \in \mathcal{L}^k(V)$. Then if $T' \coloneqq \sum_I T_I u_I^*$, we have that $T'_J = T_J$ for all multi-indices J of n of length k. Therefore, since T' and T are uniquely determined by the same T_I, this proves that $T = T'$. So, T is in the span of the u_I^*. This proves that the u_I^* span all of $\mathcal{L}^k(V)$.

Next suppose $T' = \sum_I C_I u_I^* = 0$ where each $C_I \in F$. Then if J is a multi-index of n of length k,

$$0 = T'(u_{j_1}, \ldots, u_{j_k}) = \sum_I C_I u_I^*(u_{j_1}, \ldots, u_{j_k}) = C_J$$

This shows that the u_I^* are linearly independent. ■

Corollary 1.3.15: If V is an n-dimensional vector space, then $\mathcal{L}^k(V)$ is an n^k-dimensional vector space.

If V, W are vector fields over F, $A : V \to W$ is a linear map, and $T \in \mathcal{L}^k(W)$, we define $A^\dagger T(v_1, \ldots, v_k) \coloneqq T(Av_1, \ldots, Av_k)$.

We call $A^\dagger T$ the pullback of T by the map A. Note that this is just a generalization of taking the transpose of A.

Proposition 1.3.18: If we denote $A^\dagger : \mathcal{L}^k(W) \to \mathcal{L}^k(V)$ to be the map $T \mapsto A^\dagger T$, then $A^\dagger$ is a linear map.

This should be pretty obvious...

Furthermore, $A^\dagger(T_1 \otimes T_2) = (A^\dagger T_1) \otimes (A^\dagger T_2)$.

Proof:

Suppose $T_1 \in \mathcal{L}^k(V)$ and $T_2 \in \mathcal{L}^\ell(V)$. Then:

$$\begin{aligned} A^\dagger(T_1 \otimes T_2)&(v_1, \ldots, v_k, v_{k+1}, \ldots, v_{k+\ell}) \\ &= (T_1 \otimes T_2)(Av_1, \ldots, Av_k, Av_{k+1}, \ldots, Av_{k+\ell}) \\ &= T_1(Av_1, \ldots, Av_k)T_2(Av_{k+1}, \ldots, Av_{k+\ell}) \\ &= A^\dagger T_1(v_1, \ldots, v_k)A^\dagger T_2(v_{k+1}, \ldots, v_{k+\ell}) \\ &= (A^\dagger T_1 \otimes A^\dagger T_2)(v_1, \ldots, v_k, v_{k+1}, \ldots, v_{k+\ell}). \blacksquare \end{aligned}$$

As a corollary to the above fact, we know that pullbacks map decomposable tensors to decomposable tensors.

Finally, suppose $B : U \to V$ is another linear map. Then $(AB)^\dagger T = B^\dagger(A^\dagger T)$ for all $T \in \mathcal{L}^k(W)$. In other words, $(AB)^\dagger = B^\dagger A^\dagger$.

Proof:

$$\begin{aligned} B^\dagger(A^\dagger T)(v_1, \ldots, v_k) &= A^\dagger T(Bv_1, \ldots, Bv_k) \\ &= T(ABv_1, \ldots, ABv_k) = (AB)^\dagger T(v_1, \ldots, v_k). \blacksquare \end{aligned}$$

Alternating k-Tensors:

Let V be an n-dimensional vector space over a field F with characteristic $\neq 2$ and S_k be the symmetric group over $\{1, \ldots, k\}$. For $\sigma \in S_k$ and $T \in \mathcal{L}^k(V)$, we define $T^\sigma \in \mathcal{L}^k(V)$ by:

$$T^\sigma(v_1, \ldots, v_k) = T(v_{\sigma^{-1}(1)}, \ldots, v_{\sigma^{-1}(k)})$$

Proposition 1.4.14:

(a) If $T = \ell_1 \otimes \cdots \otimes \ell_k$ where each $\ell_i \in V^*$, then $T^\sigma = \ell_{\sigma(1)} \otimes \cdots \otimes \ell_{\sigma(k)}$.

Proof:

$$\begin{aligned} T^\sigma(v_1, \ldots, v_k) = T(v_{\sigma^{-1}(1)}, \ldots, v_{\sigma^{-1}(k)}) &= \ell_1(v_{\sigma^{-1}(1)}) \cdots \ell_k(v_{\sigma^{-1}(k)}) \\ &= \ell_{\sigma(1)}(v_1) \cdots \ell_{\sigma(k)}(v_k) \\ &= (\ell_{\sigma(1)} \otimes \cdots \otimes \ell_{\sigma(k)})(v_1, \ldots, v_k) \end{aligned}$$

(b) If $\sigma \in S_k$, the function $T \mapsto T^\sigma$ is a linear map from $\mathcal{L}^k(V)$ to $\mathcal{L}^k(V)$.

This should be obvious. Also note that this map is invertible via the function $T \mapsto T^{(\sigma^{-1})}$.

(c) If $\sigma, \tau \in S_k$, then $(T^\sigma)^\tau = T^{\sigma\tau}$.

Proof:
Let $u_i \coloneqq v_{\tau^{-1}(i)}$ for all i. Then:

$$\begin{aligned}(T^\sigma)^\tau(v_1, \ldots, v_k) &= T^\sigma(v_{\tau^{-1}(1)}, \ldots, v_{\tau^{-1}(k)}) \\ &= T^\sigma(u_1, \ldots, u_k) = T(u_{\sigma^{-1}(1)}, \ldots, u_{\sigma^{-1}(k)}) \\ &= T(v_{\tau^{-1}(\sigma^{-1}(1))}, \ldots, v_{\tau^{-1}(\sigma^{-1}(k))}) \\ &= T(v_{(\sigma\tau)^{-1}(1)}, \ldots, v_{(\sigma\tau)^{-1}(k)}) \\ &= T^{\sigma\tau}(v_1, \ldots, v_k)\end{aligned}$$

Let V be a vector space and $k \geq 1$ be an integer. Then $T \in \mathcal{L}^k(V)$ is alternating if $T^\sigma = \operatorname{sgn}(\sigma)T$ for all $\sigma \in S_k$. We denote $\mathcal{A}^k(V)$ as the set of all alternating k-tensors on V.

Note:

- If $c_1, c_2 \in F$ and $T_1, T_2 \in \mathcal{A}^k(V)$, then since $T \mapsto T^\sigma$ is a linear map, we have for all $\sigma \in S_k$ that:
$$\begin{aligned}(c_1T_1 + c_2T_2)^\sigma &= c_1T_1^\sigma + c_2T_2^\sigma \\ &= c_1\operatorname{sgn}(\sigma)T_1 + c_2\operatorname{sgn}(\sigma)T_2 = \operatorname{sgn}(\sigma)(c_1T_1 + c_2T_2)\end{aligned}$$

  This proves that $\mathcal{A}^k(V)$ is a subspace of $\mathcal{L}^k(V)$.

- We shall define $\mathcal{A}^0(V) \coloneqq \mathcal{L}^0(V) = F$.

Given an integer $k > 0$, and a tensor $T \in \mathcal{L}^k(V)$, let:

$$\operatorname{Alt}(T) \coloneqq \sum_{\tau \in S_k} \operatorname{sgn}(\tau)T^\tau.$$

Then the alternation operation has the following properties:
Proposition 1.4.17:

(a) Given any $T \in \mathcal{L}^k(V)$ and $\sigma \in S_k$ (where $k > 0$), we have that $(\operatorname{Alt}(T))^\sigma = \operatorname{sgn}(\sigma)\operatorname{Alt}(T)$. I.e., $\operatorname{Alt}(T)$ is an alternating tensor.

Proof:
By proposition 1.4.14 plus the fact that $(\operatorname{sgn}(\sigma))^2 = 1$, we have that:

$$\begin{aligned}(\operatorname{Alt}(T))^\sigma &= \Big(\sum_{\tau \in S_k} \operatorname{sgn}(\tau)T^\tau\Big)^\sigma \\ &= 1 \cdot \sum_{\tau \in S_k} \operatorname{sgn}(\tau)T^{\tau\sigma} = (\operatorname{sgn}(\sigma))^2 \sum_{\tau \in S_k} \operatorname{sgn}(\tau)T^{\tau\sigma} \\ &= \operatorname{sgn}(\sigma) \sum_{\tau \in S_k} \operatorname{sgn}(\tau\sigma)T^{\tau\sigma} = \operatorname{sgn}(\sigma) \sum_{\tau' \in S_k} \operatorname{sgn}(\tau')T^{\tau'} \\ &= \operatorname{sgn}(\sigma)\operatorname{Alt}(T)\end{aligned}$$

(b) If $T \in \mathcal{A}^k(V)$, the $\mathrm{Alt}(T) = k!T$.

Proof:
Since $T^\tau = \mathrm{sgn}(\tau)T$ for all $\tau \in S_k$, we know:
$$\mathrm{Alt}(T) = \sum_{\tau \in S_k} \mathrm{sgn}(\tau)T^\tau = \sum_{\tau \in S_k} (\mathrm{sgn}(\tau))^2 T = \sum_{\tau \in S_k} (1)T = |S_k|T = k!T$$

(c) $\mathrm{Alt}(T^\sigma) = (\mathrm{Alt}(T))^\sigma$.

Proof:
By similar reasoning to in part (a), we have that:
$$\begin{aligned}\mathrm{Alt}(T^\sigma) = 1 \cdot \sum_{\tau \in S_k} \mathrm{sgn}(\tau)T^{\sigma\tau} &= (\mathrm{sgn}(\sigma))^2 \sum_{\tau \in S_k} \mathrm{sgn}(\tau)T^{\sigma\tau} \\ &= \mathrm{sgn}(\sigma) \sum_{\tau \in S_k} \mathrm{sgn}(\sigma\tau)T^{\sigma\tau} \\ &= \mathrm{sgn}(\sigma) \sum_{\tau' \in S_k} \mathrm{sgn}(\tau')T^{\tau'} = \mathrm{sgn}(\sigma)\mathrm{Alt}(T)\end{aligned}$$

And since $\mathrm{sgn}(\sigma)\mathrm{Alt}(T) = (\mathrm{Alt}(T))^\sigma$ by part (a), we know $\mathrm{Alt}(T^\sigma) = (\mathrm{Alt}(T))^\sigma$.

(d) The map $\mathrm{Alt} : \mathcal{L}^k(V) \to \mathcal{L}^k(V)$ defined by $T \mapsto \mathrm{Alt}(T)$ is a linear map. (Also it's onto if F has characteristic 0 or $> k$.)

Proof:
Alt is a linear map because it is a linear combination of a bunch of linear maps. The onto property follows from part (b).

8/7/2025

If $I = (i_1, \ldots, i_k)$ is a multi-index of n of length k, then we write:

- I is repeating if $i_s = i_r$ for some $s \neq r$.
- I is increasing if $i_1 < i_2 < \ldots < i_k$.
- Given $\sigma \in S_k$, we define $I^\sigma = (i_{\sigma(1)}, \ldots, i_{\sigma(k)})$.

Note that if I is not repeating, then there is a unique permutation $\sigma \in S_k$ such that I^σ is increasing.

Let $u_1, \ldots, u_n$ be a basis of the vector space V over a field F of characteristic $\neq 2$, and let $u_1^*, \ldots, u_n^*$ be the corresponding dual basis. Now given the multi-index $I = (i_1, \ldots, i_k)$, set $u_I^* = u_{i_1}^* \otimes \cdots \otimes u_{i_k}^*$. Next define $\Psi_I = \mathrm{Alt}(u_I^*)$.

Proposition 1.4.20: Let $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_k)$ be multi-indices.

(a) $\Psi_{I^\sigma} = \mathrm{sgn}(\sigma)\Psi_I$.

To start off, by the last proposition:
$$\mathrm{sgn}(\sigma)\Psi_I = \mathrm{sgn}(\sigma)\mathrm{Alt}(u_I^*) = (\mathrm{Alt}(u_I^*))^\sigma = \mathrm{Alt}((u_I^*)^\sigma).$$

Next, set $\ell_j = u^*_{i_j}$ for $1 \leq j \leq k$. Then by proposition 1.4.14, we have:

$$(u^*_I)^\sigma = (\ell_1 \otimes \cdots \otimes \ell_k)^\sigma = \ell_{\sigma(1)} \otimes \cdots \otimes \ell_{\sigma(k)} = u^*_{i_{\sigma(1)}} \otimes \cdots \otimes u^*_{i_{\sigma(k)}} = u^*_{I^\sigma}.$$

Thus $\operatorname{sgn}(\sigma)\Psi_I = \operatorname{Alt}((u^*_I)^\sigma) = \operatorname{Alt}(u^*_{I^\sigma}) = \Psi_{I^\sigma}$.

(b) If I is repeating, $\Psi_I = 0$.

If I is repeating, then there exists $r \neq s$ such that $i_r = i_s$. Then in turn, if $\tau_{r,s} \in S_k$ is the transposition of r and s, then $\operatorname{sgn}(\tau_{r,s}) = -1$ and $I^{\tau_{r,s}} = I$. Then by part (a), we have:

$$\Psi_I = \Psi_{I^{\tau_{r,s}}} = \operatorname{sgn}(\tau_{r,s})\Psi_I = -\Psi_I$$

This is only possible if $\Psi_I(v_1, \ldots, v_k) = 0$ for all $v_1, \ldots, v_k \in V$. Hence Ψ_I is the zero map.

(c) If I and J are strictly increasing, then $\Psi_I(u_{j_1}, \ldots, u_{j_k}) = \delta_{I,J}$ where δ is the Kronecker delta function.

To start off, note that:

$$\begin{aligned}
\Psi_I(u_{j_1}, \ldots, u_{k_k}) &= \sum_{\tau \in S_k} \operatorname{sgn}(\tau)(u^*_I)^\tau(u_{j_1}, \ldots, u_{j_k}) \\
&= \sum_{\tau \in S_k} \operatorname{sgn}(\tau)(u^*_{i_1} \otimes \cdots \otimes u^*_{i_k})^\tau(u_{j_1}, \ldots, u_{j_k}) \\
&= \sum_{\tau \in S_k} \operatorname{sgn}(\tau)(u^*_{i_{\tau(1)}} \otimes \cdots \otimes u^*_{i_{\tau(k)}})(u_{j_1}, \ldots, u_{j_k}) \\
&= \sum_{\tau \in S_k} \operatorname{sgn}(\tau) u^*_{i_{\tau(1)}}(u_{j_1}) \cdots u^*_{i_{\tau(k)}}(u_{j_k})
\end{aligned}$$

Now it's clear that $u^*_{i_{\tau(1)}}(u_{j_1}) \cdots u^*_{i_{\tau(k)}}(u_{j_k}) = \delta_{I^\tau, J}$. Also, since both J and I are strictly increasing and also since there is only one permutation such that I^σ is strictly increasing for any nonrepeating I, we know that $\delta_{I^\tau, J} = 1$ iff $\tau = \operatorname{Id}$ and $I = J$. And in that case $\operatorname{sgn}(\tau) = 1$. Hence:

$$\sum_{\tau \in S_k} \operatorname{sgn}(\tau) u^*_{i_{\tau(1)}}(u_{j_1}) \cdots u^*_{i_{\tau(k)}}(u_{j_k}) = \delta_{I,J}.$$

<u>Proposition 1.4.24:</u> Suppose F has characteristic 0 or greater than k. Then $\{\Psi_J : J \text{ is increasing}\}$ is a basis for $\mathcal{A}^k(V)$.

Proof:
Suppose $T \in \mathcal{A}^k(V)$. By theorem 1.3.13, we know there exists $a_I \in F$ such that $T = \sum_I a_I u^*_I$. However, we also know that $k!T = \operatorname{Alt}(T)$. Since Alt is a linear map, we thus know that:

$$T = \frac{1}{k!}\operatorname{Alt}(T) = \sum_I \frac{a_I}{k!}\operatorname{Alt}(u^*_I) = \sum_I \frac{a_I}{k!}\Psi_I$$

If I is repeating, then $\frac{a_I}{k!}\Psi_I^*$ cancels. Otherwise, there is some $\sigma \in S_k$ and some increasing multi-index J such that:

$$\frac{a_I}{k!}\Psi_I = \frac{a_I}{k!}\Psi_{J^\sigma} = \frac{a_I \operatorname{sgn}(\sigma)}{k!}\Psi_J.$$

By collecting terms, we get that $T = \sum_{J \text{ increasing}} c_J \Phi_J$ where each $c_J \in F$.

This shows that the Ψ_J span all of $\mathcal{A}^k(V)$. Next we show that they form a basis. Suppose $T = \sum_{J \text{ increasing}} c_J \Phi_J = 0$.

Then by part (c) of the last proposition, we know that if $I = (i_1, \ldots, i_k)$ is an increasing multi-index, then:

$$0 = T(u_{i_1}, \ldots, u_{i_k}) = C_I$$

So, all the C_I are equal to 0. ■

<u>Corollary:</u> If F has characteristic 0 or greater than k, then $\mathcal{A}^k(V)$ has dimension $\binom{n}{k}$.

<u>Corollary 2:</u> If F has characteristic 0 or greater than $k \geq n$, then any alternating n-tensor on V is a scalar multiple of a determinant function. Also, there are no nontrivial alternating m-tensors where $n < m \leq k$.

<u>Exercise 1.4.ix:</u> Suppose $A : V \to W$ is a linear map. Then if $T \in \mathcal{A}^k(W)$, we have that $A^\dagger T \in \mathcal{A}^k(V)$. Hence, the pullback operation maps alternating tensors to alternating tensors.

Proof:
Suppose $\sigma \in S_k$. Then for any $v_1, \ldots, v_k \in V$, we have that:

$$\begin{aligned}
(A^\dagger T)^\sigma(v_1, \ldots, v_k) &= A^\dagger T(v_{\sigma^{-1}(1)}, \ldots, v_{\sigma^{-1}(k)}) \\
&= T(Av_{\sigma^{-1}(1)}, \ldots, Av_{\sigma^{-1}(k)}) \\
&= T^\sigma(Av_1, \ldots, Av_k) \\
&= \operatorname{sgn}(\sigma) T(Av_1, \ldots, Av_k) = \operatorname{sgn}(\sigma) A^\dagger T(v_1, \ldots, v_k)
\end{aligned}$$

Hence, $(A^\dagger T)^\sigma = \operatorname{sgn}(\sigma) A^\dagger T$. This proves that $A^\dagger T$ is alternating. ■

<u>Exercise 1.4.x:</u> Additionally to the last exercise, we have that if $T \in \mathcal{L}^k(V)$, then $A^\dagger(\operatorname{Alt}(T)) = \operatorname{Alt}(A^\dagger T)$.

Proof:
If $v_1, \ldots, v_k \in V$, then:

$$\begin{aligned}
\operatorname{Alt}(A^\dagger T)(v_1, \ldots, v_k) &= \sum_{\tau \in S_k} \operatorname{sgn}(\tau)(A^\dagger T)^\tau(v_1, \ldots, v_k) \\
&= \sum_{\tau \in S_k} \operatorname{sgn}(\tau) A^\dagger T(v_{\tau^{-1}(1)}, \ldots, v_{\tau^{-1}(k)}) \\
&= \sum_{\tau \in S_k} \operatorname{sgn}(\tau) T(Av_{\tau^{-1}(1)}, \ldots, Av_{\tau^{-1}(k)})
\end{aligned}$$

$$= \sum_{\tau \in S_k} \operatorname{sgn}(\tau) T^\tau(Av_1, \ldots, Av_k)$$

$$= \operatorname{Alt}(T)(Av_1, \ldots, Av_k) = A^\dagger(\operatorname{Alt}(T))(v_1, \ldots, v_k)$$

This shows that $\operatorname{Alt}(A^\dagger T) = A^\dagger(\operatorname{Alt}(T))$. ■

The space $\Lambda^k(V^*)$:

If $k > 1$, a decomposable k-tensor $\ell_1 \otimes \cdots \otimes \ell_k$ with each $\ell_i \in V^*$ is called redundant if $\ell_i = \ell_{i+1}$ for some index i. We let $\mathcal{I}^k(V)$ be the span of all redundant k-tensors.

If $k = 1$, we define $\mathcal{I}^1(V) := \{0\} \subseteq \mathcal{L}^1(V)$.

Also if $k = 0$, we define $\mathcal{I}^0(V) := \{0\} \subseteq F$.

Proposition 1.5.2: Suppose F has characteristic $\neq 2$. If $T \in \mathcal{I}^k(V)$, then $\operatorname{Alt}(T) = 0$. In other words, $\mathcal{I}^k(V) \subseteq \ker(\operatorname{Alt})$.

Proof:

If $T \in \mathcal{I}^k(V)$, then we know there are redundant decomposable k-tensors $T_1, \ldots, T_m$ as well as scalars $c_1, \ldots, c_m \in F$ such that $T = \sum_{j=1}^m c_j T_j$. Then since $\operatorname{Alt}(T) = \sum_{j=1}^m c_j \operatorname{Alt}(T_j)$, all we need to do now is show that $\operatorname{Alt}(T_j) = 0$ for every j.

Since T_j is a redundant decomposable k-tensor, we know that $T_j = \ell_1 \otimes \cdots \otimes \ell_k$ where $\ell_i = \ell_{i+1}$ for some $1 \leq i < k$. In turn, if $\tau_{i,i+1} \in S_k$ is the transposition of i and $i+1$, we have that $(T_j)^{\tau_{i,i+1}} = T_j$ and $\operatorname{sgn}(\tau_{i,i+1}) = -1$. Hence:

$$\operatorname{Alt}(T_j) = \operatorname{Alt}((T_j)^{\tau_{i,i+1}}) = \operatorname{sgn}(\tau_{i,i+1})\operatorname{Alt}(T_j) = -\operatorname{Alt}(T_j)$$

This implies that $\operatorname{Alt}(T_j) = 0$. ■

Proposition 1.5.3: If $T \in \mathcal{I}^r(V)$ and $T' \in \mathcal{L}^s(V)$, then $T \otimes T'$ and $T' \otimes T$ are in $\mathcal{I}^{r+s}(V)$.

Proof:

The argument for $T' \otimes T$ being in $\mathcal{I}^{r+s}(V)$ is mostly identical to the argument for $T \otimes T'$ being in $\mathcal{I}^{r+s}(V)$. So, I'll focus only on proving the latter.

To start off, like before we know that there are redundant decomposable r-tensors $T_1, \ldots, T_m$ as well as scalars $c_1, \ldots, c_m \in F$ such that $T = \sum_{j=1}^m c_j T_j$. Hence, it suffices to show that $T_j \otimes T' \in \mathcal{I}^{r+s}(V)$ for all $1 \leq j \leq m$ since:

$$T \otimes T' = (\textstyle\sum_{j=1}^m c_j T_j) \otimes T' = \sum_{j=1}^m c_j (T_j \otimes T')$$

Fortunately, by writing $T' = \sum_I d_I u_I^*$, we can see that:

$$T_j \otimes T' = T_j \otimes (\sum_I d_I u_I^*) = \sum_I d_I (T_j \otimes u_I^*)$$

Now since both u_I^* and T_j are decomposable and T_j is redundant, we can easily see that $T_j \otimes u_I^*$ is decomposable and redundant. It follows that $T_j \otimes T' \in \mathcal{I}^{r+s}(V)$. ■

Proposition 1.5.4: Suppose F has characteristic $\neq 2$. If $T \in \mathcal{L}^k(V)$ and $\sigma \in S_k$, then $T^\sigma = \operatorname{sgn}(\sigma) T + S$ where $S \in \mathcal{I}^k(V)$.

Proof:
Hopefully you're getting use to this trick. It suffices to assume T is decomposable. After all, after writing $T = \sum_I c_I u_I^*$, if we can show for all multi-indexes I that $(u_I^*)^\sigma = \operatorname{sgn}(\sigma)u_I^* + S_I$ where $S_I \in \mathcal{I}^k(V)$, then we can set $S = \sum_I c_I S_I \in \mathcal{I}^k(V)$ and have that:

$$T^\sigma = \sum_I c_I (u_I^*)^\sigma = \operatorname{sgn}(\sigma)\sum_I c_I u_I^* + \sum_I S_I = \operatorname{sgn}(\sigma)T + S$$

So suppose $T = \ell_1 \otimes \cdots \otimes \ell_k$. Then given $\sigma \in S^k$, we can write $\sigma = \tau_1 \ldots \tau_m$ as the product of m many transpositions of adjacent pairs of numbers in $\{1, \ldots, k\}$.

(by adjacent I mean a pair $\{j, j+1\}$ where $1 \leq j < k$...)

We shall induct on m. First assume $m = 1$. Thus $\sigma = \tau_{j,j+1}$ for some $1 \leq j < k$ and hence $\operatorname{sgn}(\sigma) = -1$. Also:

$$\begin{aligned} T^\sigma - \operatorname{sgn}(\sigma)T &= T^\sigma + T \\ &= (\ell_1 \otimes \cdots \otimes \ell_{j-1} \otimes \ell_{j+1} \otimes \ell_j \otimes \ell_{j+2} \otimes \cdots \otimes \ell_k) + (\ell_1 \otimes \cdots \otimes \ell_k) \\ &= (\ell_1 \otimes \cdots \otimes \ell_{j-1}) \otimes ((\ell_{j+1} \otimes \ell_j) + (\ell_j \otimes \ell_{j+1})) \otimes (\ell_{j+2} \otimes \cdots \otimes \ell_k) \end{aligned}$$

Now note that:

$$(\ell_j + \ell_{j+1}) \otimes (\ell_j + \ell_{j+1}) = (\ell_j \otimes \ell_j) + (\ell_j \otimes \ell_{j+1}) + (\ell_{j+1} \otimes \ell_j) + (\ell_{j+1} \otimes \ell_{j+1}).$$

Therefore:

$$\begin{aligned} T^\sigma - \operatorname{sgn}(\sigma)T = (\ell_1 \otimes \cdots \otimes \ell_{j-1}) \otimes (\ell_j + \ell_{j+1}) \otimes (\ell_j + \ell_{j+1}) \otimes (\ell_{j+1} \otimes \cdots \otimes \ell_k) \\ - (\ell_1 \otimes \cdots \otimes \ell_{j-1}) \otimes \ell_j \otimes \ell_j \otimes (\ell_{j+2} \otimes \cdots \otimes \ell_k) \\ - (\ell_1 \otimes \cdots \otimes \ell_{j-1}) \otimes \ell_{j+1} \otimes \ell_{j+1} \otimes (\ell_{j+2} \otimes \cdots \otimes \ell_k) \end{aligned}$$

Hence $T^\sigma - \operatorname{sgn}(\sigma)T \in \mathcal{I}^k(V)$ and we are done with this case.

Now suppose $m > 1$. Then $\sigma = \tau_{j,j+1}\sigma'$ where σ' is the product of $m-1$ transpositions. By induction, we know that there exists $S_1 \in \mathcal{I}^k(V)$ such that:

$$T^\sigma = (T^{\tau_{j,j+1}})^{\sigma'} = \operatorname{sgn}(\sigma')T^{\tau_{j,j+1}} + S_1$$

Also by our base case, there is $S_2 \in \mathcal{I}^k(V)$ such that $T^{\tau_{j,j+1}} = \operatorname{sgn}(\tau_{j,j+1})T + S_2$. Then setting $S = \operatorname{sgn}(\tau_{j,j+1})S_2 + S_1$, we have that $S \in \mathcal{I}^K(V)$ and:

$$T^\sigma = \operatorname{sgn}(\sigma')(\operatorname{sgn}(\tau_{j,j+1})T + S_2) + S_1 = \operatorname{sgn}(\sigma)T + S. \blacksquare$$

Corollary 1.5.6: Suppose F has characteristic $\neq 2$. If $T \in \mathcal{L}^k(V)$, then $\operatorname{Alt}(T) = k!T + S$ where $S \in \mathcal{I}^k(V)$.

Proof:
Given any $\tau \in S_k$, let $S_\tau \in \mathcal{I}^k(V)$ be such that $T^\tau = \operatorname{sgn}(\tau)T + S_\tau$. Then $S := \sum_{\tau \in S_k} \operatorname{sgn}(\tau)S_\tau \in \mathcal{I}^k(V)$ and:

$$\operatorname{Alt}(T) = \sum_{\tau \in S_k} \operatorname{sgn}(\tau)T^\tau = \sum_{\tau \in S_k} (\operatorname{sgn}(\tau))^2 T + \sum_{\tau \in S_k} \operatorname{sgn}(\tau)S_\tau = k!T + S. \blacksquare$$

Corollary 1.5.8: Let $k \geq 1$. Then let V be a vector space over a field F of characteristic 0 or $> \max(k, 2)$. Then:

$$\mathcal{I}^k(V) = \ker(\mathrm{Alt} : \mathcal{L}^k(V) \to \mathcal{A}^k(V))$$

Proof:
We already know from proposition 1.5.2 that $\mathcal{I}^k(V) \subseteq \ker(\mathrm{Alt})$. To prove the reverse relation, suppose $T \in \mathcal{L}^k(V)$ satisfies that $\mathrm{Alt}(T) = 0$. Then based on the previous corollary, we know there exists $S \in \mathcal{I}^k(V)$ such that $-\frac{1}{k!}S = T - \mathrm{Alt}(T)$. Hence $T \in \mathcal{I}^k(V)$. ■

Theorem 1.5.9: Suppose F is a field of characteristic 0 or $> \max(k, 2)$. Then any element $T \in \mathcal{L}^k(V)$ can be written uniquely as a sum $T_1 + T_2$ where $T_1 \in \mathcal{A}^k(V)$ and $T_2 \in \mathcal{I}^k(V)$. I.e, $\mathcal{L}^k(V) = \mathcal{A}^k(V) \oplus \mathcal{I}^k(V)$.

Proof:
Let $W \in \mathcal{I}^k(V)$ satisfy that $\mathrm{Alt}(T) = k!T + W$. Then set $T_1 = \frac{1}{k!}\mathrm{Alt}(T)$ and $T_2 = -\frac{1}{k!}W$. Then clearly $T = T_1 + T_2$ with $T_1 \in \mathcal{A}^k(V)$ and $T_2 \in \mathcal{I}^k(V)$.

Next, to prove uniqueness suppose $T_1' + T' = T$ with $T_1' \in \mathcal{A}^k(V)$ and $T_2' \in \mathcal{I}^k(V)$. Then $T_1 - T_1' \in \mathcal{A}^k(V)$, $T_2 - T_2' \in \mathcal{I}^k(V)$, and $(T_1 - T_1') + (T_2 - T_2') = 0$. So:

$$0 = \mathrm{Alt}(0) = \mathrm{Alt}((T_1 - T_1') + (T_2 - T_2')) = k!(T_1 - T_1')$$

Hence $T_1 = T_1'$ and it easily follows $T_2 = T_2'$. ■

Let $k \geq 0$. Let V be a finite dimensional vector space over a field F of characteristic 0 or $> \max(k, 2)$. Then we define:

$$\Lambda^k(V^*) \coloneqq \mathcal{L}^k(V)/\mathcal{I}^k(V)$$

By the first isomorphism theorem along with the previous theorem, we have that $\Lambda^k(V^*) \cong \mathcal{A}^k(V)$.

8/8/2025

Here is a tangent about symmetric tensors. For this section, suppose V is an n-dimensional vector space over a field F with characteristic $\neq 2$.

A tensor $T \in \mathcal{L}^k(V)$ is symmetric if $T^\sigma = T$ for all $\sigma \in S_k$. We denote the space of symmetric tensors $\mathcal{S}^k(V)$.

You can show by the same reasoning as with $\mathcal{A}^k(V)$ that $\mathcal{S}^k(V)$ is a vector subspace.

Exercise 1.5.iii: Suppose F has characteristic 0 or $> k$. Then if T is a symmetric k-tensor and $k \geq 2$, we have that $T \in \mathcal{I}^k(V)$.

Proof:
Let $\sigma \in S_k$ be an odd permutation. Then by proposition 1.4.17:

$$\mathrm{Alt}(T) = \mathrm{Alt}(T^\sigma) = \mathrm{sgn}(\sigma)\mathrm{Alt}(T) = -\mathrm{Alt}(T).$$

The only way this is possible is if $\mathrm{Alt}(T) = 0$. Hence $T \in \ker(\mathrm{Alt})$, and by theorem 1.5.8 that means that $T \in \mathcal{I}^k(V)$. ■

We define a symmetrization operator as follows. Given $T \in \mathcal{L}^k(V)$, define:

$$\mathrm{Sym}(T) := \sum_{\sigma \in S_k} T^\sigma$$

Then like in proposition 1.4.17, we can show that given any $T \in \mathcal{L}^k(V)$ and $\sigma \in S_k$:

(a) $(\mathrm{Sym}(T))^\sigma = \mathrm{Sym}(T)$ (i.e. $\mathrm{Sym}(T) \in \mathcal{S}^k(V)$...)

(b) If $T \in \mathcal{S}^k(V)$, then $\mathrm{Sym}(T) = k!T$

(c) $\mathrm{Sym}(T^\sigma) = \mathrm{Sym}(T)$

(d) $\mathrm{Sym} : \mathcal{L}^k(V) \to \mathcal{A}^k(V)$ is an linear map (which is surjective so long as $k! \neq 0$ in the field F...).

Supposing $k! \neq 0$ in F, then by following a process very similar to what we did with the alternation operation, we can construct a basis for the symmetric tensors:

$$\{\Phi_I^* : I \text{ is a non-decreasing multi-index}\}.$$

Note: By non-decreasing the textbook means that $I = (i_1, \ldots, i_k)$ satisfies that $i_1 \leq i_2 \leq \cdots \leq i_k$.

I'm bored and won't do that construction here. But the important point is that this means $\mathcal{S}^k(V)$ has the same number of dimensions as there are non-decreasing multi-indexes of n of length k. And since there are $\binom{n+k-1}{k}$ ways of picking k elements of the set $\{1, \ldots, n\}$ when you allow yourself to pick the same element multiple times, this means that $\dim(\mathcal{S}^k(V)) = \binom{n+k-1}{k}$.

Side note: If $k! \neq 0$ in F, then we have already shown that $\dim(\mathcal{I}^k(V)) = n^k - \binom{n}{k}$. Since $\mathcal{S}^k(V) \subseteq \mathcal{I}^k(V)$, this shows that $\mathcal{S}^2(V) = \mathcal{I}^2(V)$. That said, we don't in general have that $\dim(\mathcal{I}^k(V)) = \dim(\mathcal{S}^k(V))$ when $k > 2$.

Next, here's some other miscellaneous results.

Exercise 1.5.vii: Suppose F has characteristic 0 or $> \max(k, 2)$. Then if $T \in \mathcal{I}^k(V)$, we have that $T^\sigma \in \mathcal{I}^k(V)$ for all $\sigma \in S_k$.

Proof:
Since $T \in \mathcal{I}^k(V)$, we know that: $\mathrm{Alt}(T^\sigma) = \mathrm{sgn}(\sigma)\mathrm{Alt}(T) = 0$. Therefore, $T^\sigma \in \ker(\mathrm{Alt})$, and by corollary 1.5.8 we know that $\ker(\mathrm{Alt}) = \mathcal{I}^k(V)$. ■

Corollary / Exercise 1.5.v: Let $k \geq 2$ and suppose F has characteristic 0 or $> k$. Then if $T \in \mathcal{L}^{k-2}(V)$ and $\ell \in V^*$, we have that $\ell \otimes T \otimes \ell \in \mathcal{I}^k(V)$.

Proof:
There is a permutation $\sigma \in S_k$ satisfying that $(\ell \otimes T \otimes \ell)^\sigma = (\ell \otimes \ell) \otimes T$. Then by proposition 1.5.3, we have that $(\ell \otimes \ell) \otimes T \in \mathcal{I}^k(V)$. And finally, by applying the last exercise we have that $\ell \otimes T \otimes \ell = ((\ell \otimes \ell) \otimes T)^{\sigma^{-1}} \in \mathcal{I}^k(V)$. ■

Corollary / Exercise 1.5.vi: Let $k \geq 2$ and suppose F has characteristic 0 or $> k$. Then if $T \in \mathcal{L}^{k-2}(V)$ and $\ell_1, \ell_2 \in V^*$, we have that $(\ell_1 \otimes T \otimes \ell_2) + (\ell_2 \otimes T \otimes \ell_1) \in \mathcal{I}^k(V)$.

Proof:

Apply the last exercise plus the fact that:

$$(\ell_1 \otimes T \otimes \ell_2) + (\ell_2 \otimes T \otimes \ell_1) = ((\ell_1 + \ell_2) \otimes T \otimes (\ell_1 + \ell_2)) - (\ell_1 \otimes T \otimes \ell_1) - (\ell_2 \otimes T \otimes \ell_2). \blacksquare$$

8/9/2025

The Wedge Product:

In this section, we'll suppose V is an n-dimensional vector space over a field F of characteristic 0. Also, we shall for each k define the map $\pi : \mathcal{L}^k(V) \to \Lambda^k(V^*)$ such that $\pi(T) = T + \mathcal{I}^k(V)$ for all $T \in \mathcal{L}^k(V)$.

Suppose for each $i \in \{1, 2\}$ we have $\omega_i \in \Lambda^{k_i}(V^*)$. Then if for each i we are given $T_i \in \mathcal{L}^{k_i}(V)$ satisfying that $\pi(T_i) = \omega_i$, we define:

$$\omega_1 \wedge \omega_2 \coloneqq \pi(T_1 \otimes T_2).$$

Claim 1.6.3: The wedge product is well defined.

Proof:

Suppose for each $i \in \{1, 2\}$ that we also have $T_i' \in \mathcal{L}^{k_i}(V)$ satisfying that $\pi(T_i') = \pi(T_i) = \omega_i$. Then for each i there exists $W_i \in \mathcal{I}^k(V)$ such that $T_i' = T_i + W_i$. Hence:

$$\pi(T_1' \otimes T_2') = \pi((T_1 \otimes T_2) + (T_1 \otimes W_2) + (W_1 \otimes T_2) + (W_1 \otimes W_2))$$

Then by applying proposition 1.5.3, we know that:

$$(T_1 \otimes W_2) + (W_1 \otimes T_2) + (W_1 \otimes W_2) \in \mathcal{I}^k(V)$$

Hence, $\pi(T_1' \otimes T_2') = \pi(T_1 \otimes T_2)$. $\blacksquare$

More generally, if for each $i \in \{1, \ldots, m\}$ we have $\omega_i \in \Lambda^{k_i}(V^*)$ and $T_i \in \mathcal{L}^{k_i}(V)$ satsifying that $\pi(T_i) = \omega_i$, then we define:

$$\omega_1 \wedge \cdots \wedge \omega_m \coloneqq \pi(T_1 \otimes \cdots \otimes T_m)$$

This is well defined for basically the same reasoning as before, although to avoid some overly long expressions, it suffices to replace only one tensor at a time.

Claim: Given any $m \geq 3$, we have that:

$$\omega_1 \wedge (\omega_2 \wedge \cdots \wedge \omega_m) = \omega_1 \wedge \cdots \wedge \omega_m = (\omega_1 \wedge \cdots \wedge \omega_{m-1}) \wedge \omega_m$$

Proof:

If for each $i \in \{1, \ldots, m\}$ we have some $T_i \in \mathcal{L}^{k_i}(V)$ satisfying that $\pi(T_i) = \omega_i$, then $\pi(T_2 \otimes \cdots \otimes T_m) = \omega_2 \wedge \cdots \wedge \omega_m$ and $\pi(T_1 \otimes \cdots \otimes T_{m-1}) = \omega_1 \wedge \cdots \wedge \omega_{m-1}$. In turn:

$$\begin{aligned}\omega_1 \wedge \cdots \wedge \omega_m &= \pi(T_1 \otimes \cdots \otimes T_m) \\ &= \pi(T_1 \otimes (T_2 \otimes \cdots \otimes T_m)) = \omega_1 \wedge (\omega_2 \wedge \cdots \wedge \omega_m) \\ &= \pi((T_1 \otimes \cdots \otimes T_{m-1}) \otimes T_m) = (\omega_1 \wedge \cdots \wedge \omega_{m-1}) \wedge \omega_m\end{aligned}$$

Corollary: The wedge product is associative and we get the same result no matter how we use parentheses to group together the ω_i.

Proof:

Suppose we have $\omega_1, \ldots, \omega_m$. If $m = 3$, then we're already done by the last claim. Meanwhile, for $m > 3$ is suffices due to the strong inductive hypothesis on m to show that for any $1 \leq s < r \leq m$ with not both $s = 1$ and $r = m$:

$$\omega_1 \wedge \cdots \wedge \omega_m = \omega_1 \wedge \cdots \wedge \omega_{s-1} \wedge (\omega_s \wedge \cdots \wedge \omega_r) \wedge \omega_{r+1} \wedge \cdots \wedge \omega_m$$

Luckily, when focusing on the case that $r \neq m$, note that by the previous claim as well as the strong inductive hypothesis:

$$\begin{aligned}\omega_1 \wedge \cdots \wedge \omega_{s-1} &\wedge (\omega_s \wedge \cdots \wedge \omega_r) \wedge \omega_{r+1} \wedge \cdots \wedge \omega_{m-1} \wedge \omega_m \\ &= (\omega_1 \wedge \cdots \wedge \omega_{s-1} \wedge (\omega_s \wedge \cdots \wedge \omega_r) \wedge \omega_{r+1} \wedge \cdots \wedge \omega_{m-1}) \wedge \omega_m \\ &= (\omega_1 \wedge \cdots \wedge \omega_{m-1}) \wedge \omega_m = \omega_1 \wedge \cdots \wedge \omega_m\end{aligned}$$

The other case is analogous.

Here are some other properties of the wedge product which I'm too bored to properly write proofs for:

- If $\lambda \in F$, then $\lambda(\omega_1 \wedge \omega_2) = (\lambda\omega_1) \wedge \omega_2 = \omega_1 \wedge (\lambda\omega_2)$.
- $(\omega_1 + \omega_2) \wedge \omega_3 = (\omega_1 \wedge \omega_3) + (\omega_2 \wedge \omega_3)$
- $\omega_1 \wedge (\omega_2 + \omega_3) = (\omega_1 \wedge \omega_2) + (\omega_1 \wedge \omega_3)$

Side note: if we were instead writing the definition of the wedge product in terms of alternating tensors, we'd be defining:

$$T_1 \wedge \cdots \wedge T_m = \frac{1}{(k_1 + \ldots + k_m)!}\mathrm{Alt}(T_1 \otimes \cdots \otimes T_m).$$

Hopefully its obvious why this definition is inferior.

Note that since $\mathcal{I}^1(V) = \{0\}$, we can just identify $\Lambda^1(V^*) = V^*$. Then given $\ell_1, \ldots, \ell_k \in V^* = \Lambda^1(V^*)$, we say that $\omega = \pi(\ell_1 \otimes \cdots \otimes \ell_k) = \ell_1 \wedge \cdots \wedge \ell_k$ is a decomposable element of $\Lambda^k(V^*)$.

Claim: For any $\sigma \in S_k$, we have: $\ell_{\sigma(1)} \wedge \cdots \wedge \ell_{\sigma(k)} = \mathrm{sgn}(\sigma)\ell_1 \wedge \cdots \wedge \ell_k$.

Proof:

Lemma: If $T \in \mathcal{L}^k(V)$ and $\sigma \in S_k$, then $\pi(T^\sigma) = \mathrm{sgn}(\sigma)\pi(T)$.

This is just a consequence of proposition 1.5.4.

As a result of that lemma:

$$\begin{aligned}\ell_{\sigma(1)} \wedge \cdots \wedge \ell_{\sigma(k)} &= \pi(\ell_{\sigma(1)} \otimes \cdots \otimes \ell_{\sigma(k)}) \\ &= \pi((\ell_1 \otimes \cdots \otimes \ell_k)^\sigma) \\ &= \mathrm{sgn}(\sigma)\pi(\ell_1 \otimes \cdots \otimes \ell_k) \\ &= \mathrm{sgn}(\sigma)\ell_1 \wedge \cdots \wedge \ell_k\end{aligned}$$

As a corollary, given any $\ell_1, \ell_2 \in V^*$ we have that $\ell_1 \wedge \ell_2 = -\ell_2 \wedge \ell_1$. Also, given $\ell_1, \ell_2, \ell_3 \in V^*$, we have:

$$\begin{aligned}\ell_1 \wedge \ell_2 \wedge \ell_3 &= -\ell_2 \wedge \ell_1 \wedge \ell_3 = \ell_2 \wedge \ell_3 \wedge \ell_1 \\ &= -\ell_1 \wedge \ell_3 \wedge \ell_2 = \ell_3 \wedge \ell_1 \wedge \ell_2\end{aligned}$$

Let $u_1, \ldots, u_n$ be a basis for V and let $u_1^*, \ldots, u_n^*$ be the corresponding dual basis. Then the collection of $u_{i_1}^* \wedge \cdots \wedge u_{i_k}^*$ such that $I = (i_1, \ldots, i_k)$ is an increasing multi-index forms a basis for $\Lambda^k(V^*)$.

Proof:
Recall that when defining $u_I^* = u_{i_1}^* \otimes \cdots \otimes u_{i_k}$ for a multi-index $I = (i_1, \ldots, i_k)$, we then have that the $\Psi_I := \mathrm{Alt}(u_I^*)$ where I is increasing form a basis of $\mathcal{A}^k(V)$. It follows that each $\pi(\Psi_I)$ where I is increasing is a basis vector of $\Lambda^k(V^*)$. But note that:

$$\pi(\Psi_I) = \pi\left(\sum_{\tau \in S_k} \mathrm{sgn}(\tau)(u_I^*)\tau\right) = \sum_{\tau \in S_k} \mathrm{sgn}(\tau)\pi((u_I^*)^\tau) = \sum_{\tau \in S_k} (\mathrm{sgn}(\tau))^2 \pi(u_I^*) = k!\pi(u_I^*)$$

So, the $\pi(u_I^*) = u_{i_1} \wedge \cdots \wedge u_{i_k}$ also form a basis for $\Lambda^k(V^*)$.

This now let's us prove the following general result:

Theorem 1.6.10: If $\omega_1 \in \Lambda^r(V^*)$ and $\omega_2 \in \Lambda^s(V^*)$, then $\omega_1 \wedge \omega_2 = (-1)^{rs}\omega_2 \wedge \omega_1$.

Proof:
Express $\omega_1 = \sum_I c_I u_{i_1}^* \wedge \cdots \wedge u_{i_r}^*$ and $\omega_2 = \sum_J d_J u_{j_1}^* \wedge \cdots \wedge u_{j_s}^*$.

Then we have that:

$$\begin{aligned}\omega_1 \wedge \omega_2 &= \sum_{I,J} c_I d_J (u_{i_1}^* \wedge \cdots \wedge u_{i_r}^* \wedge u_{j_1}^* \wedge \cdots \wedge u_{j_s}^*) \\ &= \sum_{I,J} c_I d_J (-1)^{rs} (u_{j_1}^* \wedge \cdots \wedge u_{j_s}^* \wedge u_{i_1}^* \wedge \cdots \wedge u_{i_r}^*) \\ &= (-1)^{rs} \sum_{I,J} d_J c_I (u_{j_1}^* \wedge \cdots \wedge u_{j_s}^* \wedge u_{i_1}^* \wedge \cdots \wedge u_{i_r}^*) = (-1)^{rs}\omega_2 \wedge \omega_1.\end{aligned}$$

One more note I'd like to make is that we can identify $\Lambda^0(V^*)$ and F. Then if $\omega \in \Lambda^k(V^*)$ and $\lambda \in F$, we have that $\lambda \wedge \omega = \lambda\omega = \omega \wedge \lambda$.

8/10/2025

Before moving onto the next section of the book, I'm going to do a few of the exercises.

Exercise 1.6.iii: Given $\omega \in \Lambda^r(V^*)$, we define $\omega^1 := \omega$ and $\omega^k := \omega \wedge \omega^{k-1} \in \Lambda^{rk}(V^*)$ for all $k > 1$. In other words, ω^k is the k-fold wedge product of ω with itself.

(A) If r is odd, then $\omega^k = 0$ for all $k > 1$.

Proof:
By an easy application of theorem 1.6.10, we have that:

$$\omega^k = \omega \wedge \omega^{k-1} = (-1)^{r \cdot r^{k-1}}\omega^{k-1} \wedge \omega = (-1)^{r^k}\omega^k$$

But r^k is odd if r is odd. Then in turn, $\omega^k = -\omega^k$. The only way this is possible is if $\omega^k = 0$.

(B) If ω is decomposable, then $\omega^k = 0$ for all $k > 1$.

Proof:
For the ease of notation we'll $\omega^0 = 1 \in F$. Now if $\omega = \ell_1 \wedge \cdots \wedge \ell_r$, then by just swapping two occurences of ℓ_1, we have that:

$$\begin{aligned}\omega^k &= \ell_1 \wedge \cdots \wedge \ell_r \wedge \ell_1 \wedge \cdots \wedge \ell_r \wedge \omega^{k-2} \\ &= (-1)\ell_1 \wedge \cdots \wedge \ell_r \wedge \ell_1 \wedge \cdots \wedge \ell_r \wedge \omega^{k-2} = -\omega^k\end{aligned}$$

This implies $\omega^k = 0$.

Exercise 1.6.iv: If $\omega, \mu \in \Lambda^r(V^*)$, then:

$$(\omega + \mu)^k = \sum_{i=0}^{k} \binom{k}{i} \omega^i \wedge \mu^{k-i}.$$

This is obvious so I'm skipping this problem. I just wanted to write out the result.

The interior Product:
All the assumptions about V and F made yesterday still apply and you should keep assuming them until I tell you to stop (cause I don't want to keep writing this shtick).

Given $T \in \mathcal{L}^k(V)$ where $k > 1$ and $v \in V$, we define the $(k-1)$-tensor:

$$\iota_v T(v_1, \ldots, v_{k-1}) \coloneqq \sum_{r=1}^{k} (-1)^{r-1} T(v_1, \ldots, v_{r-1}, v, v_r, \ldots, v_{k-1})$$

Also if $\lambda \in \mathcal{L}^0(V) = F$, we define $\iota_v \lambda = 0$ for all $v \in V$.

Note that if $v = c_1 v_1 + c_2 v_2$ and $T = d_1 T_1 + d_2 T_2$, then:

$\iota_v T = c_1 \iota_{v_1} T + c_2 \iota_{v_2} T$ and $\iota_v T = d_1 \iota_v T_1 + d_2 \iota_v T_2$.

Also, if $T = \ell_1 \otimes \cdots \otimes \ell_k$ where each $\ell_i \in V^*$, then when writing $\hat{\ell_r}$ to mean that we are deleting ℓ_r from that term of the expression, we have that:

$$\iota_v T = \sum_{r=1}^{k} (-1)^{r-1} \ell_r(v) \ell_1 \otimes \cdots \otimes \hat{\ell_r} \otimes \cdots \otimes \ell_k$$

Slightly less obviously, if $T_1 \in \mathcal{L}^p(V)$ and $T_2 \in \mathcal{L}^q(V)$, we have that:

$$\iota_v(T_1 \otimes T_2) = (\iota_v T_1) \otimes T_2 + (-1)^p T_1 \otimes (\iota_v T_2).$$

Lemma 1.7.8: Let V be a vector space and $T \in \mathcal{L}^k(V)$ where $k \geq 1$. Then for all $v \in V$, $\iota_v(\iota_v(T)) = 0$.

Proof:
By linearity it suffices to prove this statement for decomposable T. Also, this statement is trivial when $k = 1$. So, we can proceed by induction, assuming that the theorem holds for $T \in \mathcal{L}^r(V)$ where $r < k$. Then after expressing $T = T' \otimes \ell$ where $T' \in \mathcal{L}^{k-1}(V)$ and $\ell \in V^*$, we have that:

$$\begin{aligned}\iota_v T = \iota_v(T' \otimes \ell) &= (\iota_v T') \otimes \ell + (-1)^{k-1} T' \otimes (\iota_v T') \\ &= (\iota_v T') \otimes \ell + (-1)^{k-1} \ell(v) T'\end{aligned}$$

By induction $\iota_v(\iota_v T') = 0$. Combining that with the above reasoning shows:

$$\begin{aligned}\iota_v(\iota_v T) &= \iota_v((\iota_v T') \otimes \ell + (-1)^{k-1}\ell(v)T') \\ &= \iota_v((\iota_v T') \otimes \ell) + (-1)^{k-1}\ell(v)\iota_v(T') \\ &= \big(\iota_v(\iota_v T') \otimes \ell + (-1)^{k-2}(\iota_v T') \otimes (\iota_v \ell)\big) + (-1)^{k-1}\ell(v)\iota_v(T') \\ &= 0 + (-1)^{k-2}\ell(v)(\iota_v T') + (-1)^{k-1}\ell(v)\iota_v(T') = 0. \blacksquare\end{aligned}$$

Corollary: If $v_1, v_2 \in V$ and $T \in \mathcal{L}^k(V)$, then $\iota_{v_1}(\iota_{v_2} T) = -\iota_{v_2}(\iota_{v_1} T)$.

Proof:

We know from the prior lemma that:

$$\iota_{v_1+v_2}(\iota_{v_1+v_2} T) = 0$$

Therefore:

$$0 + \iota_{v_1}(\iota_{v_2} T) = \iota_{v_1}(\iota_{v_1+v_2} T) = -\iota_{v_2}(\iota_{v_1+v_2} T) = -\iota_{v_2}(\iota_{v_1} T) - 0$$

Lemma 1.7.11: If $T \in \mathcal{L}^k(V)$ is redundant, then so is $\iota_v T$.

Proof:

Write $T = T_1 \otimes \ell \otimes \ell \otimes T_2$ where $\ell \in V^*$, $T_1 \in \mathcal{L}^p(V)$, and $T_2 \in \mathcal{L}^q(V)$. Then:

$$\iota_v T = \iota_v(T_1) \otimes \ell \otimes \ell \otimes T_2 + (-1)^p T_1 \otimes \iota_v(\ell \otimes \ell) \otimes T_2 + (-1)^{p+2} T_1 \otimes \ell \otimes \ell \otimes \iota_v(T_2)$$

Now the first and third terms are obvious redundant. Meanwhile, the second term cancels because $\iota_v(\ell \otimes \ell) = \ell(v)\ell - \ell(v)\ell = 0$. $\blacksquare$

Corollary: If $T \in \mathcal{I}^k(V)$, then $\iota_v T \in \mathcal{I}^{k-1}(V)$.

Now we define the interior product operator ι_v on $\Lambda^k(V^*)$. If π is the projection of $\mathcal{L}^k(V)$ onto $\Lambda^k(V^*)$ and $\omega = \pi(T) \in \Lambda^k(V^*)$, then we define:

$$\iota_v \omega \coloneqq \pi(\iota_v T) \in \Lambda^{k-1}(V^*).$$

This is well defined since by the previous corollary, if both T and T' satisfy that $\pi(T) = \pi(T') = \omega$, then there is some tensor $S \in \mathcal{I}^{k-1}(V)$ such that $\iota_v T = \iota_v T' + S$.

It is easily shown then that if $v_1, v_2, v \in V$, $\omega, \omega_1 \in \Lambda^p(V^*)$, and $\omega_2 \in \Lambda^q(V^*)$, then:

- $\iota_{v_1+v_2}\omega = \iota_{v_1}\omega + \iota_{v_2}\omega$;
- $\iota_v(\lambda_1\omega_1 + \lambda_2\omega_2) = \lambda_1\iota_v\omega_1 + \lambda_2\iota_v\omega_2$ (where $\lambda_1, \lambda_2 \in F$);
- $\iota_v(\omega_1 \wedge \omega_2) = (\iota_v\omega_1) \wedge \omega_2 + (-1)^p\omega_1 \wedge (\iota_v\omega_2)$.

Also, if you squint you can see that $\iota_v(\iota_v\omega) = \pi(\iota_v(\iota_v T))$ where T satisfies that $\pi(T) = \omega$. Hence, we have that $\iota_v(\iota_v\omega) = 0$, and from there we can show that $\iota_{v_1}(\iota_{v_2}\omega) = -\iota_{v_2}(\iota_{v_1}\omega)$ just like before.

8/13/2025

I'm going to take a break from Guillemin's book and instead try to learn some algebraic topology. To do this I'm going to start following Munkres' Topology.

If $f_1, f_2 : X \to Y$ are continuous maps, we say f_1 is homotopic to f_2 if there is a continuous map $F : X \times [0,1] \to Y$ such that $F(x,0) = f_1(x)$ and $F(x,1) = f_2(x)$. F is called a homotopy between f_1 and f_2. And if f_1 and f_2 are homotopic, we write $f_1 \simeq f_2$. If $f_1 \simeq f_2$ and f_2 is a constant map, then we say f_1 is nulhomotopic.

An important special case is when f_1 and f_2 are paths (i.e. continuous maps from $[0,1]$ to a topological space X). In this case, it can be helpful to make the following stricter distinction. We say f_1 and f_2 are path homotopic if they have the same initial point x_0 and final point x_1, and there is a homotopy F between the two paths such that $F(0,t) = x_0$ and $F(1,t) = x_1$ for all t. Also, we call F a path homotopy and say $f_1 \simeq_p f_2$. Note in other contexts that f_1 and f_2 are allowed to have different intervals as their domain.

Lemma 51.1: $\simeq$ and $\simeq_p$ are equivalence relations.

Proof:

It's clear that any f is homotopic to itself. Also, if $F(x,t)$ is a homotopy showing that $f_1 \sim f_2$, then $G(x,t) = F(x, 1-t)$ is a homotopy showing that $f_2 \sim f_1$.

Finally, suppose $f_1 \simeq f_2$ and $f_2 \simeq f_3$. Then there exists two homotopy's $F^{(1)}$ between f_1 and f_2 and $F^{(2)}$ between f_2 and f_3. So, define:

$$G(x,t) = \begin{cases} F^{(1)}(x, 2t) & \text{for } t \in [0, 1/2] \\ F^{(2)}(x, 2t-1) & \text{for } t \in [1/2, 1] \end{cases}$$

Then G is a homotopy between f_1 and f_3, meaning $f_1 \simeq f_3$.

We know G is continuous by the pasting lemma.

The added stuff needed to $\simeq_p$ is an equivalence relation is obvious.

If f is a path in X from x_0 to x_1 and g is a path in X from x_1 to x_2, we define the product $f * g$ to be the path h given by the equation:

$$h(s) = \begin{cases} f(2s) & \text{for } s \in [0, 1/2] \\ g(2s-1) & \text{for } s \in [1/2, 1] \end{cases}$$

By the pasting lemma, h is a well-defined path in X from x_0 to x_2.

If $f : [0,1] \to X$ is a path, let $[f]$ denote the path homotopy class of f. Then the product operation induces a well-defined operation on path-homotopy classes. Specifically, given a class $[f]$ from x_0 to x_1 and a class $[g]$ from x_1 to x_2, define $[f] * [g] = [f * g]$.

To verify that this is well defined, suppose $f \simeq_p f'$ and $g \simeq_p g'$. Then if F is a homotopy from f to f' and G is a homotopy from g to g', we can define a homotopy H from $f * g$ to $f' * g'$ by the formula:
$$H(s,t) = \begin{cases} F(2s,t) & \text{for } s \in [0, 1/2] \\ G(2s-1,t) & \text{for } s \in [1/2, 1] \end{cases}$$

H is well-defined and continuous by pasting lemma.

Recall that a groupoid is a category in which every morphism is an isomorphism (look at my old Allufi notes to see what a category is...). Using the product operation of path-homotopy classes, we can define a groupoid as follows:

Consider the space X as a collection of objects, and for any $x_0, x_1 \in X$, let $\mathrm{Hom}_X(x_0, x_1)$ be the collection of path homotopy classes from x_0 to x_1. For the law of composition, say that if $[f] \in \mathrm{Hom}_X(x_0, x_1)$ and $[g] \in \mathrm{Hom}_X(x_1, x_2)$, then $[g][f] = [f * g] \in \mathrm{Hom}_X(x_0, x_2)$.

We claim:

- Every point has an identity morphism (namely the homotopy class of the constant map).
- For any $[f] \in \mathrm{Hom}(x_0, x_1)$, you can reverse the path f (i.e. define $\bar{f}(s) \coloneqq f(1-s)$) in order to get an inverse morphism in $\mathrm{Hom}(x_1, x_0)$.
- Finally, if $[f] \in \mathrm{Hom}(x_0, x_1)$, $[g] \in \mathrm{Hom}(x_1, x_2)$, and $[h] \in \mathrm{Hom}(x_2, x_3)$, then:
$$[f] * ([g] * [h]) = ([f] * [g]) * [h].$$

Proof:
We start with two lemmas:

1. If $k : X \to Y$ is a continuous map and F is a path homotopy in X between the paths f and f', then $k \circ F$ is a path homotopy in Y between the paths $k \circ f$ and $k \circ f'$.
2. If $k : X \to Y$ is a continuous map and f and g are paths in X with $f(1) = g(0)$, then $k \circ (f * g) = (k \circ f) * (k \circ g)$.

To prove the first bullet point, let $e_0 : [0,1] \to [0,1]$ be the constant function equal to 0 and $i : [0,1] \to [0,1]$ be the identity map. Then when considering both of those as paths in $[0,1]$, we can fairly easily find a path homotopy G from $e_0 * i$ to i.

One path homotopy that works is to define
$$F(s,t) = t(e_0 * i)(s) + (1-t)i(s).$$

Now suppose $e_{x_0} : [0,1] \to X$ is constant at x_0 and $f : [0,1] \to X$ is a path from x_0 to x_1. Then $e_{x_0} = f \circ e_0$, $f = f \circ i$, and by our two lemmas, $f \circ G$ is a path homotopy from $f = f \circ i$ to $f \circ (e_0 * i) = (f \circ e_0) * (f \circ i) = e_{x_0} * f$. Similar reasoning shows that if $e_{x_1} : [0,1] \to X$ is constant at x_1, then $f \simeq_p f * e_{x_1}$. This proves bullet point 1.

To prove the second bullet point, let $\bar{i} = i(1-s)$. Then we can find a homotopy G from e_0 to $i * \bar{i}$ (one that works is $G(s,t) = t((i * \bar{i})(s))$).

Then for any path f from x_0 to x_1, we can easily see that $e_{x_0} = f \circ e_0$, $f = f \circ i$, and $\bar{f} = f \circ \bar{i}$. Hence by our two lemmas, $f \circ G$ is a homotopy between $e_{x_0} = f \circ (e_0)$ and $f * \bar{f} = (f \circ i) * (f \circ \bar{i}) = f \circ (i * \bar{i})$. Also, once again similar reasoning shows that $e_{x_1} \simeq_p \bar{f} * f$. This proves bullet point 2.

I'm bored. So tldr: to prove the third bullet point just note that we can apply a continuous reparametrization $k(s)$ to $((f * g) * h)(s)$ to get $(f * (g * h))(s)$. Hence, we can define a homotopy:

$$G(s,t) := ((f * g) * h)((1-t)s + tk(s)). \blacksquare$$

Now given a point $x_0 \in X$, define $\pi_1(X, x_0) := \mathrm{End}(x_0)$. This is the fundamental group of X relative to x_0, and it is in fact a group with respect to our product operation since X was a groupoid. I'm going to state the next proposition as abstractly as I can cause why the hell not.

Proposition: Let **C** be a groupoid and let $A, B \in \mathrm{Obj}(\mathbf{C})$. If there exists $g \in \mathrm{Hom}(A, B)$, then $\mathrm{End}(A) \cong \mathrm{End}(B)$.

Proof:
If $f \in \mathrm{End}(A)$, then define $\phi(f) = gfg^{-1}$. Then it's clear that ϕ is a group homomorphism from $\mathrm{End}(A)$ to $\mathrm{End}(B)$. To show that ϕ is injective, suppose $\phi(f) = e_B$ where e_B is the identity morphism on B. Then $f = g^{-1}e_B g = g^{-1}g = e_A$ where e_A is the identity morphism on A. Next, to show that ϕ is surjective, suppose $h \in \mathrm{End}(B)$. Then $f := g^{-1}hg$ satisfies that $\phi(f) = h$.

Corollary: If X is path connected, then $\pi_1(X, x_0) \cong \pi_1(X, x_1)$ for all $x_0, x_1 \in X$.

We say a space X is simply connected if X is path connected and for some $x_0 \in X$, $\pi_1(X, x_0) = \{1\}$.

Lemma 52.3: Let X be a path-connected topological space. Then X is simply connected iff every pair of paths f and f' with the same initial and final point are path homotopic.

$(\Longrightarrow)$
Suppose f and f' are paths from x_0 to x_1. Then if we set $\bar{f}$ and $\bar{f}'$ to be the reversed paths, we know that $[f' * \bar{f}], [f * \bar{f}'] \in \pi_1(X, x_0)$. But now since $\pi_1(X, x_0)$ is trivial, we know that $[f' * \bar{f}] = 1 = [f * \bar{f}']$. So, there is a path homotopy F between $f' * \bar{f}$ and $f * \bar{f}$. Now just define $G(s,t) = F(\frac{1}{2}s, t)$ and we have shown that f and f' are path homotopic.

$(\Longleftarrow)$
Suppose $f \in \pi_1(X, x_0)$ for some $x_0 \in X$. Also suppose g is a path in X from x_0 to x_1 where $x_1 \neq x_0$. (Note, this lemma is trivial if X has only one point. So, we can without loss of generality assume X has more than one point.)

Then since both g and $f * g$ are paths from x_0 to x_1, we know there is a homotopy F between the two paths. So, $[g] = [f * g] = [g] * [f]$. If we apply on the left side the class $[\bar{g}]$, then this means that $1 = [f]$. Hence, $\pi_1(X, x_0)$ is trivial. ■

A consequence of this lemma is that all convex subsets of $\mathbb{R}^n$ are simply connected.

Recall the two lemmas stated on page 118. Those lemmas will let us define an important thing.

Suppose $h : X \to Y$ is a continuous map such that $h(x_0) = y_0$. We will denote this by writing $h : (X, x_0) \to (Y, y_0)$. Then we define the homomorphism induced by h, $h_* : \pi_1(X, x_0) \to \pi_2(X, y_0)$, by the map:

$$h_*([f]) \coloneqq [h \circ f]$$

This is well defined since if F is a homotopy between f and f', then $h \circ F$ is a homotopy between $h \circ f$ and $h \circ f'$. Also, this is indeed a group homomorphism since:

$$h_*([f] * [g]) = [h \circ (f * g)] = [h \circ f] * [h \circ g] = h_*([f]) * h_*([g])$$

Theorem 52.4: If $h : (X, x_0) \to (Y, y_0)$ and $k : (Y, y_0) \to (Z, z_0)$ are continuous maps, then $(k \circ h)_* = k_* \circ h_*$. Also if $i : (X, x_0) \to (X, x_0)$ is the identity map on X, then i_* is the identity map on $\pi_1(X, x_0)$.

Hopefully this is self-explanatory.

Corollary 52.5: If $h : (X, x_0) \to (Y, y_0)$ is a homeomorphism from X to Y, then h_* is an isomorphism from $\pi_1(X, x_0)$ to $\pi_1(Y, y_0)$.

Proof:
Let $k = h^{-1}$. Then from the last theorem we can easily see that k_* and h_* are inverses of each other. Hence, h_* is invertible and thus a group isomorphism. ■

Consequently, we know that the fundamental group of a path connected space is a topological property. So, if two path connected spaces do not have the same fundamental group, they can't be homeomorphic.

8/15/2025

If α is a path from x_0 to x_1 in X, we shall denote $\bar{\alpha}$ to be the reversed path from x_1 to x_0. Also we shall denote $\hat{\alpha} : \pi_1(X, x_0) \to \pi_1(X, x_1)$ to be the isomorphism $[f] \mapsto [\bar{\alpha} * f * \alpha]$.

Exercise 52.3: Let $x_0 \to x_1$ be points of the path-connected space X. Show that $\pi_1(X, x_0)$ is abelian iff for every pair α and β of paths from x_0 to x_1, we have $\hat{\alpha} = \hat{\beta}$.

$(\Longrightarrow)$
Suppose $\pi_1(X, x_0)$ is abelian and we have two paths α, β from x_0 to x_1. Then for any $[f] \in \pi_1(X, x_0)$, we have that $[f] * [\alpha * \bar{\beta}] = [\alpha * \bar{b}] * [f]$. Therefore:

$$\hat{\alpha}(f) = [\bar{\alpha} * f * \alpha] = [\bar{\beta} * f * \beta] = \hat{\beta}(f).$$

$(\Longleftarrow)$
Suppose $[f] \in \pi_1(X, x_0)$ and α, β are paths from x_0 to x_1. Then since $\hat{\alpha}([f * \alpha * \bar{\beta}]) = \hat{\beta}([f * \alpha * \bar{\beta}])$, we have that:
$$[\bar{\alpha} * f * \alpha * \bar{\beta} * \alpha] = [\bar{\beta} * f * \alpha * \bar{\beta} * \beta] = [\bar{\beta} * f * \alpha]$$

Hence $[f] * [\alpha * \bar{\beta}] = [\alpha * \bar{\beta}] * [f]$, and this proves that the group operation of $\pi_1(X, x_0)$ is commutative so long as one of the arguments passes through x_1.

Now to prove general commutativity, suppose $[f], [g] \in \pi_1(X, x_0)$ and α is a path from x_0 to x_1. Then from before we know that:
$$[f] * [g] = [f] * [g * \alpha * \bar{\alpha}] = [g * \alpha * \bar{\alpha}] * [f] = [g] * [f].$$

Exercise 52.4: Let $A \subseteq X$, suppose $r : X \to A$ is a continuous map such that $r(a) = a$ for each $a \in A$. (The map r is called a retraction of X onto A and we call A a retract of X.) If $a_0 \in A$, then show that $r_* : \pi_1(X, a_0) \to \pi_1(A, a_0)$ is surjective.

Suppose $[f]_A \in \pi_1(A, a_0)$. Then f is also a loop in X, meaning it has a class $[f]_X \in \pi_1(X, a_0)$. And $r_*([f]_X) = [r \circ f]_A = [f]_A$. Hence r_* is surjective.

Exercise 52.5: Let A be a subspace of a simply connected space X, and let $h : (A, a_0) \to (Y, y_0)$ be a continuous map. If h is extendable to a continuous map of X into Y, then h_* is the trivial homomorphism (i.e. the homomorphism mapping everything to the identity element).

Let h' be a continuous extension of h to all of X. Now importantly, for any loop $[f]_A \in \pi_1(A, a_0)$, we know f also defines a class $[f]_X \in \pi_1(X, a_0)$. Also importantly:
$$h'_*([f]_X) = [h' \circ f]_Y = [h \circ f]_Y = h_*([f]_A).$$

However, since X is simply connected, we know $[f]_X = 1$ and thus $h'_*([f]_X) = 1$. So, we've shown that $h_*([f]_A) = 1$ for all $[f]_A \in \pi_1(A, a_0)$.

Note: Munkres specifically sets $X = \mathbb{R}^n$ in his statement of the exercise.

Exercise 52.6: Suppose $h : X \to Y$ is a continuous map, α is a path from x_0 to x_1 in X, and $\beta := h \circ \alpha$. Then $h_* \circ \hat{\alpha} = \hat{\beta} \circ h_*$.

Suppose $[f] \in \pi_1(X, x_0)$. Then:
$$\begin{aligned} h_*(\hat{\alpha}([f])) = [h \circ (\bar{\alpha} * f * \alpha)] &= [(h \circ \bar{\alpha}) * (h \circ f) * (h \circ \alpha)] \\ &= [\bar{\beta} * h_*([f]) * \beta] = \hat{\beta}(h_*([f])). \end{aligned}$$

Let $p : E \to B$ be a continuous surjective map. Then an open set $U \subseteq B$ is said to be evenly covered by p if $p^{-1}(U)$ is a union of disjoint open sets $V_\alpha \subseteq E$ satisfying that $p|_{V_\alpha}$ is a homeomorphism from V_α to U. The collection $\{V_\alpha\}_{\alpha \in A}$ will be called a partition of $p^{-1}(U)$ into slices.

Note that if $W \subseteq U$ is also open, then W is also openly covered by p with there being a partition of slices $\{V_\alpha \cap p^{-1}(W)\}_{\alpha \in A}$.

If every point $b \in B$ has an open neighborhood $U \subseteq B$ that is evenly covered by p, we call p a covering map and E a covering space of B.

Claim: If $p : E \to B$ is a covering map of B, then p is an open map.

Proof:

Suppose $A \subseteq E$ is open. Then for any $y \in f(A)$, there exists an open neighborhood U of y that is evenly covered by p. In turn there exists $x \in A$ with $p(x) = y$ and an open neighborhood V_α of x such that $p|_{V_\alpha}$ is a homeomorphism from V_α to U_y. And hence, $p(A \cap V_\alpha)$ is an open neighborhood of y in $U \subseteq B$ that is also a subset of $p(A)$. It follows that $p(A)$ is an open subset of B. ■

Corollary: If $p : E \to B$ is a covering map of B, then p is a local homeomorphism, meaning each point $e \in E$ has an open neighborhood that is mapped homeomorphically by p onto an open subset of B.

Proof:

If $x \in E$, then the reasoning from the prior proof lets us pick an open set $A \cap V_\alpha$ containing x such that, $p|_{A \cap V_\alpha}$ is a homeomorphism to an open set $p(A \cap V_\alpha)$ in B.

Theorem 53.1: The map $p : \mathbb{R} \to S^1$ given by $p(x) = (\cos(2\pi x), \sin(2\pi x))$ is a covering map of S^1.

Hopefully this is obvious. We can cover S^1 with four open sets gotten by intersecting S^1 with the top, bottom, right, and left open halfs respectively of the coordinate plane. Then it's easy to find a partition of each preimage into slices.

As a side note, if $H^1 = \{x \in \mathbb{R} : x \geq 0\}$, then $p|_{H^1}$ is surjective and a local homeomorphism. That said, it is not a covering map since the point $(1, 0) \in S^1$ has no open neighborhood U that is evenly covered by $p|_{H^1}$.

(The specific issue we run into is that for any open set U we pick, $p|_{H^1}$ will not be a surjective map from the slice of the preimage containing 0 to U...)

Theorem 53.2: Let $p : E \to B$ be a covering map. If B_0 is a subspace of B and $E_0 := p^{-1}(B_0)$, then the map $p_0 : E_0 \to B_0$ obtained by restricting p is a covering map.

Proof:

Given $y \in B_0$, let $U \subseteq B$ be an open neighborhood of y that is evenly covered and let $\{V_\alpha\}_{\alpha \in A}$ be a partition of $p^{-1}(U)$ into slices. Then $U \cap B_0$ is an open neighborhood of y in B_0 that is evenly covered by p_0 via the partition $\{V_\alpha \cap E_0\}_{\alpha \in A}$ of $p^{-1}(U \cap B_0)$ into slices. ■

Theorem 53.3: If $p : E \to B$ are covering maps and $p' : E' \to B'$ are covering maps, then the map $p \times p' : E \times E' \to B \times B'$ defined by $(e, e') \mapsto (p(e), p'(e'))$ is a covering map.

Proof:

Given $b \in B$ and $b' \in B'$, let U and U' be open neighborhoods of b and b' that are evenly covered by p and p' respectively. Next let $\{V_\alpha\}_{\alpha \in A}$ and $\{V'_\gamma\}_{\gamma \in C}$ be partitions of $p^{-1}(U)$ and $(p')^{-1}(U')$ respectively into slices. Then:

$$(p \times p')^{-1}(U \times U') = \bigcup_{\alpha \in A} \Big(\bigcup_{\gamma \in C} (V_\alpha \times V'_\gamma) \Big).$$

Also $U \times U'$ is open in $B \times B'$; $V_\alpha \times V'_\gamma$ is open in $E \times E'$ for all α and γ; the $V_\alpha \times V'_\gamma$ are all disjoint; and each $V_\alpha \times V'_\gamma$ is mapped homeomorphically onto $U \times U'$. ■

Exercise 53.2: Let $p : E \to B$ be continuous and surjective, and suppose $U \subseteq B$ is an open set that is evenly covered by p. If U is connected, then the partition of $p^{-1}(U)$ into slices is unique.

Proof:
For the sake of contradiction, suppose $\{V_\alpha\}_{\alpha \in A}$ and $\{V'_\beta\}_{\beta \in B}$ are two different partitions of $p^{-1}(U)$ into slices. Then we know there exists $V_{\alpha_0}, V'_{\beta_0}$ in those two partitions satisfying that $V_{\alpha_0} \cap V'_{\beta_0} \neq \emptyset$ and $V_{\alpha_0} \neq V'_{\beta_0}$. Without loss of generality suppose $V_{\alpha_0} - V'_{\beta_0} \neq \emptyset$. Then since p is a homeomorphism from V_{α_0} to U, we know $p(V_{\alpha_0} \cap V'_{\beta_0})$ is open in U. Also, since $V'_{\beta_0} = p^{-1}(U) - (\bigcup_{\beta \neq \beta_0} V'_\beta)$, we know V'_{β_0} is closed in $p^{-1}(U)$. Hence $V_{\alpha_0} - V'_{\beta_0}$ is open in V_{α_0} and so $p(V_{\alpha_0} - V'_{\beta_0})$ is also open in U.

But now $p(V_{\alpha_0} - V'_{\beta_0})$ and $p(V_{\alpha_0} \cap V'_{\beta_0})$ are two disjoint nonempty open subsets of U. This contradicts that U is connected. ■

Exercise 53.3: Let $p : E \to B$ be a covering map and let B be connected. If $p^{-1}(\{b_0\})$ has k elements for some $b_0 \in B$, then $p^{-1}(\{b\})$ has k elements for all $b \in B$. In such a case E is called a k-fold covering of B.

Proof:
Let $B_k := \{b \in B : |p^{-1}(\{b\})| = k\}$. Now it's very clear that if $b \in B$ and $U \subseteq B$ is an open neighborhood of b that is evenly covered by p, then $U \subseteq B_k$. Hence, B_k is open. Also note that if $b \in (B_k)^\mathsf{C}$ and $U \subseteq B$ is an open neighborhood of b that is evenly covered by p, then $U \subseteq (B_k)^\mathsf{C}$. Hence $(B_k)^\mathsf{C}$ is open. Since B is connected, this means that B_k or $(B_k)^\mathsf{C}$ must be empty. and since $b_0 \in B_k$, we know the empty set isn't B_k. ■

Let $p : E \to B$ be a map. If $f : X \to B$ is a continuous map, a lifting of f is a map $\tilde{f} : X \to E$ such that $p \circ \tilde{f} = f$. Or in other words, the following diagram commutes:

$$\begin{tikzcd} & E \arrow[d, "p"] \\ X \arrow[ur, "\tilde{f}"] \arrow[r, "f"'] & B \end{tikzcd}$$

Lemma 54.1: Let $p : E \to B$ be a covering map and let $e_0 \in p^{-1}(\{b_0\})$. Then if $f : [0,1] \to B$ is a path beginning at b_0, there is a unique lifting of f to a path $\tilde{f} \in E$ beginning at e_0.

Proof:
Let $\{U_\alpha\}_{\alpha \in A}$ be an open covering of B consisting of open sets that are evenly covered by p. Then in turn $\{f^{-1}(U_\alpha)\}_{\alpha \in A}$ is an open covering of $[0,1]$. So, we can invoke the following useful lemma.

Lebesgue Number Lemma: If (X, d) is a compact metric space and an open cover $\mathcal{U}$ of X is given, then the cover admits some Lebesgue number $\delta > 0$. That is, for some $\delta > 0$ we have that: $\operatorname{diam}(E) < \delta \implies \exists U \in \mathcal{U} \ s.t. \ E \subseteq U$.

Proof:

Let $\{U_1, \ldots, U_n\} \subseteq \mathcal{U}$ be a finite subcover of X. If any $U_i = X$, then trivially any $\delta > 0$ works. So assume $U_i \neq X$ for all i. Then since $F_i := X - U_i$ is nonempty for all i, we know that $d(x, F_i) = \inf\{d(x, y) : y \in F_i\}$ is well defined for all $x \in X$ and $i \in \{1, \ldots, n\}$. Furthermore, it is easy to see that $d(x, F_i)$ is a continuous function of x. And since F_i is closed, we have that $d(x, F_i) = 0$ iff $x \in F_i$.

Now define $f : X \to \mathbb{R}$ by $f(x) = \sum\limits_{i=1}^{n} d(x, F_i)$.

Since f is a continuous map from a compact set, we know by the extreme value theorem that f attains a minimum $\alpha \in \mathbb{R}$. Also, since the U_i cover X and each $x \in U_i$ is some ε from F_i, we know $f(x) \neq 0$ for all $x \in X$. Hence, $\alpha > 0$.

Now set $\delta = {}^{\alpha}/_{n}$. Then for any $E \subseteq X$ with $\operatorname{diam}(E) < \delta$, if we pick $x_0 \in E$ we have that $E \subseteq B_\delta(x_0)$. Importantly, since $f(x_0) > \alpha = n\delta$, we must have that $d(x, F_i) > \delta$ for some i. It follows then that $E \subseteq U_i$. This proves that δ works as a Lebesgue number. ■

By the above lemma, we know that there exists $0 = s_0 < s_1 < \ldots < s_n = 1$ satisfying for each $0 \leq i < n$ that $[s_i, s_{i+1}] \subseteq f^{-1}(U_{\alpha_i})$ for some $\alpha_i \in A$. Or in other words, $f([s_i, s_{i+1}]) \subseteq U_{\alpha_i}$ for some $\alpha_i \in A$. After setting $\tilde{f}(0) = e_0$, we may proceed by induction as follows in order to show a suitable $\tilde{f}$ exists.

Assume that $\tilde{f}$ was already defined for all $s \in [0, s_i]$ where $i < n$. Since U_{α_i} is evenly covered by p, we know there is a unique open set $V \subseteq E$ such that $\tilde{f}(s_i) \in V$ and $p|_V$ is a homeomorphism onto U_{α_i}. Then since $f([s_i, s_{i+1}]) \subseteq U_{\alpha_i}$, we know it is well defined to set $\tilde{f}(s) = (p|_V)^{-1}(f(s))$ for all $s \in [s_i, s_{i+1}]$. Also, it is clear then that $\tilde{f}|_{[s_i, s_{i+1}]}$ is continuous since it is the composition of two continuous functions. In turn, we can see by the pasting lemma that $\tilde{f}$ will be continuous on the domain $[0, s_{i+1}]$

To finish off, we now need to show that $\tilde{f}$ is unique. So suppose $g : [0, 1] \to E$ also satisfies that $p \circ g = E$ and $g(0) = e_0$. Then we trivially have that $g(s_0) = \tilde{f}(s_0)$. So, we may proceed by induction as follows in order to show $\tilde{f} = g$.

Assume we've already shown that $\tilde{f}(s) = g(s)$ for all $s \in [0, s_i]$ where $0 \leq i < n$. Now letting V be as in the prior reasoning, we know that $g(s)$ must be in V for all $s \in [s_i, s_{i+1}]$. After all, if $\{V_\gamma\}_{\gamma \in C}$ is a partition of $p^{-1}(U_{\alpha_i})$ into slices, we must have that:

$$g([s_i, s_{i+1}]) \subseteq p^{-1}(U_{\alpha_i}) = \bigcup_{\gamma \in C} V_\gamma.$$

But since all the V_γ are disjoint, nonempty, and open; and $g([s_i, s_{i+1}])$ is connected, it must be the case that $g([s_i, s_{i+1}])$ only intercepts one V_γ. Specifically, that one V_γ is V since $g(s_i) \in V$.

But now we must have for each $s \in [s_i, s_{i+1}]$ that $g(s)$ satisfies that $p(g(s)) = f(s)$. Yet the only point in V which satisfies that is $\tilde{f}(s)$. Hence, $\tilde{f}(s) = g(s)$ for all $[0, s_{i+1}]$. ■

Lemma 54.2: Let $p : E \to B$ be a covering map and let $e_0 \in p^{-1}(\{b_0\})$. Then if $F : [0,1]^2 \to B$ is a continuous map satisfying that $F(0,0) = b_0$, there is a unique lifting of F to a continuous function $\tilde{F} : [0,1]^2 \to E$ such that $\tilde{F}(0,0) = e_0$.

Proof:

We'll start by showing uniqueness. Suppose $\tilde{F}$ and G both are continuous functions from $[0,1]^2$ to E satisfying our lemma. By an easy application of the last lemma, if $(x_0, y_0) \in [0,1]^2$, then we must have $\tilde{f}(s) := \tilde{F}(sx_0, sy_0)$ and $g(s) := G(sx_0, sy_0)$ are equal for all $s \in [0,1]$ since both are the unique lifting of the path $f(s) := F(sx_0, sy_0)$ to a path in E starting at e_0. But then we've shown that $\tilde{F}(x_0, y_0) = \tilde{f}(1) = g(1) = G(x_0, y_0)$. And since (x_0, y_0) was arbitrary, we have shown that $\tilde{F}$ is unique if it exists.

Next, we need to show that a sufficient $\tilde{F}$ exists in the first place. So start by setting $\tilde{F}(0,0) = e_0$. Now for any fixed $(x_0, y_0) \in [0,1]^2$, if we define $f(s) = F(sx_0, sy_0)$, then we know by the prior lemma that there is a continuous map $\tilde{f} : [0,1] \to E$ such that $p \circ \tilde{f}(s) = f(s)$ for all $s \in [0,1]$. Hence, we may define $\tilde{F}(x_0, y_0) = \tilde{f}(1)$. After doing this for all choices of (x_0, y_0), we will have constructed a function $\tilde{F} : [0,1]^2 \to E$ such that $p \circ \tilde{F} = F$.

What's still not clear is that $\tilde{F}$ is continuous. To show this, we will need the observation that if $f(s) = F(sx_0, sy_0)$ and $\tilde{f}$ is the unique lifting of f to path in E such that $p \circ \tilde{f} = f$ and $\tilde{f}(0) = e_0$, then by an easy application of the last lemma we can show that $\tilde{F}(sx_0, sy_0) = \tilde{f}(s)$. Hence, for any $(x_0, y_0) \in [0,1]^2$ we have that $\tilde{F}$ varies continuously as one moves along the straight line between $(0,0)$ to (x_0, y_0).

Now use the Lebesgue number lemma to pick $0 = s_0 < s_1 < \ldots < s_n = 1$ such that for all $0 \leq i, j < n$, $F([s_i, s_{i+1}] \times [s_j, s_{j+1}])$ is contained in some open set U that is evenly covered by p. Then make the inductive hypotheses that there exists $0 \leq i, j < n$ for which we've already shown that $\tilde{F}$ is continuous on:

$$([0, s_i] \times [0,1]) \cup ([s_i, s_{i+1}] \times [0, s_j]).$$

Note that if $i = 0$ or $j = 0$, then the fact that $\tilde{F}$ is continuous on the line connecting $(0,0)$ to $(0,1)$ and on the line connecting $(0,0)$ to $(1,0)$ proves this base case.

Now let U be an open set that contains $F([s_i, s_{i+1}] \times [s_j, s_{j+1}])$ and is evenly covered by p. Then let $V \subseteq E$ be an open set disjoint from the rest of the preimage of U such that $p|_V$ is a homeomorphism from V to U and $\tilde{F}(s_i, s_j) \in V$. Since we already know by induction that F varies continuously along the straight lines from (s_i, s_j)

to (s_i, s_{j+1}), and from (s_i, s_j) to (s_{i+1}, s_j), we know that $F(x, s_j)$ and $F(s_i, y)$ are in V for all $x \in [s_i, s_{i+1}]$ and $y \in [s_j, s_{j+1}]$. This is important because we know that for any $(x, y) \in [s_i, s_{i+1}] \times [s_j, s_{j+1}]$, the straight line from $(0, 0)$ to (x, y) must cross one of those two borders of the rectangle. And since F varies continuously along the straight line from $(0, 0)$ to (x, y), this proves that $F(x, y) \in V$ for all $(x, y) \in [s_i, s_{i+1}] \times [s_j, s_{j+1}]$.

In turn, we have for all $(x, y) \in [s_i, s_{i+1}] \times [s_j, s_{j+1}]$ that $\tilde{F} = (p|_V)^{-1}(F(x, y))$. Thus $\tilde{F}$ is continuous on $[s_i, s_{i+1}] \times [s_j, s_{j+1}]$ since it is the composition of two continuous functions. Also by pasting lemma, we can thus say that $\tilde{F}$ is continuous on $([0, s_i] \times [0, 1]) \cup ([s_i, s_{i+1}] \times [0, s_{j+1}])$. ■

Corollary: If F is a path homotopy then the lifting in the prior lemma: $\tilde{F}$, is a path homotopy.

Proof:
If $F(0, t) = b_0$ for all $t \in [0, 1]$, then we know that $\tilde{F}(\{0\} \times [0, 1]) \subseteq p^{-1}(b_0)$. But the latter set will have the discrete topology as a subspace of E. Since $\tilde{F}(\{0\} \times [0, 1])$ is connected, this must mean that $\tilde{F}$ is constant on $0 \times [0, 1]$. Similar reasoning also shows that $\tilde{F}(\{1\} \times [0, 1])$ has one element. ■

Note: $p^{-1}(b_1)$ as a subspace of E has the discrete topology because p being a covering map means that each of the elements of $p^{-1}(b_1)$ are contained in distinct disjoint open sets. Also, a subset of a space with the discrete topology is connected iff it has one element.

Theorem 54.3: Let $p : E \to B$ be a covering map, and let $e_0 \in p^{-1}(b_0)$. Next let f and g be paths in B from b_0 to b_1 and let $\tilde{f}$ and $\tilde{g}$ be their respective liftings to a path in E starting at e_0. If f and g are path homotopic, then $\tilde{f}$ and $\tilde{g}$ end at the same point and are path homotopic.

Proof:
Let $F : [0, 1]^2 \to B$ be a homotopy between f and g. Then $F(0, 0) = b_0$, meaning there is a unique continuous lifting $\tilde{F} : [0, 1]^2 \to E$ of F to E satisfying that $\tilde{F}(0, 0) = e_0$. By our prior corollary, we know that $\tilde{F}$ will be a homotopy, meaning that $\tilde{f}(\{0\} \times [0, 1]) = \{e_0\}$ and $\tilde{f}(\{1\} \times [0, 1]) = \{e_1\}$ where $e_1 \in p^{-1}(b_1)$. Also, due to the uniqueness we proved in lemma 54.1, it's easy to see that $\tilde{F}(s, 0) = \tilde{f}(s)$ and $\tilde{F}(s, 1) = \tilde{g}(s)$ for all s. ■

Let $p : E \to B$ be a covering map and let $b_0 \in B$. Choose an $e_0 \in E$ such that $p(b_0) = b_0$. Then given an element $[f] \in \pi_1(B, b_0)$, define $\phi([f]) := \tilde{f}(1)$ where $\tilde{f}$ is the unique lifting of f to path in E starting at e_0.

By the last theorem, $\phi : \pi_1(B, b_0) \to p^{-1}(b_0)$ is a well-defined map which we call the lifting correspondence derived from p. Note, ϕ is dependent on our choice of e_0.

Theorem 54.4: Let $p : E \to B$ be a covering map and for some $b_0 \in B$, choose some $e_0 \in p^{-1}(b_0)$. If E is path connected, then the lifting correspondence $\phi : \pi_1(B, b_0) \to p^{-1}(b_0)$ is surjective. Furthermore, if E is simply connected, then ϕ is bijective.

Proof:
If E is path connected, then given any $e_1 \in p^{-1}(b_0)$ there exists a path $\tilde{f}$ from e_0 to e_1. In turn, $f := p \circ \tilde{f}$ is a continuous loop from b_0 to itself in B such that $\phi([f]) = \tilde{f}(1) = e_1$.

Next suppose E is simply connected. Then suppose $[f], [g] \in \pi_1(B, b_0)$ satisfy that $\phi([f]) = \phi([g])$. By letting $\tilde{f}$ and $\tilde{g}$ be the liftings of f and g respectively, we know that $\tilde{f}$ and $\tilde{g}$ are both paths from e_0 to some e_1. Therefore, by lemma 52.3 plus the fact that E is simply connected, we know that $\tilde{f}$ and $\tilde{g}$ are path homotopic via a homotopy $\tilde{F}$. And since p is a continuous map, we have that $F := p \circ \tilde{F}$ is a path homotopy of $f = p \circ \tilde{f}$ and $g = p \circ \tilde{g}$. This proves that $[f] = [g]$. ■

Theorem 54.5: The fundamental group of $S^1 := \{x \in \mathbb{R}^2 : \|x\|_2 = 1\}$ is isomorphic to $\mathbb{Z}$.

Proof:
Let $p : \mathbb{R} \to S^1$ be the covering map $p(t) = (\cos(2\pi t), \sin(2\pi t))$. Also pick $e_0 = 0$ and let $b_0 = p(e_0) = (1, 0)$. Then $p^1(\{b_0\}) = \mathbb{Z}$. And since $\mathbb{R}$ is simply connected, we have that the lifting correspondence $\phi : \pi_1(S^1, b_0) \to \mathbb{Z}$ is bijective.

To show that $\pi_1(S^1, b_0) \cong \mathbb{Z}$, we shall now show that ϕ is a group homomorphism.
Suppose $[f], [g] \in \pi_1(S^1, b_0)$. Then let $\tilde{f}$ and $\tilde{g}$ be the respective liftings of f and g to $\mathbb{R}$ to paths starting at 0. Also let $n := \phi([f])$ and $m := \phi([g])$. If we define $\tilde{h}(s) = \tilde{g}(s) + n$, then because $p(x + n) = p(x)$ for all $x \in \mathbb{R}$, we know $\tilde{h}$ is another lifting of g. However $\tilde{h}$ starts at n instead of 0. It follows that $\tilde{f} * \tilde{h}$ is a well-defined product of paths, and it is lifting of $f * g$ to a path in $\mathbb{R}$ starting at 0. Hence:

$$\phi([f] * [g]) = \phi([f * g]) = (\tilde{f} * \tilde{h})(1) = n + m = \phi([f]) + \phi([g]). \blacksquare$$

The intution for this result is that the fundamental group of S^1 categorizes all loops in S^1 by the net number of complete revolutions of the path around S^1.

Theorem 54.6: Let $p : E \to B$ be a covering map and let $p(e_0) = b_0$.

(a) The induced homomorphism $p_* : \pi_1(E, e_0) \to \pi_1(B, b_0)$ is injective.

Proof:
Suppose $[\tilde{h}] \in \pi_1(E, e_0)$ such that $p_*([\tilde{h}])$ is the identity element of $\pi_1(B, b_0)$. Then there is a homotopy F between $p \circ \tilde{h}$ and the constant loop based at b_0. If $\tilde{F}$ is the lifting of F to E satisfying that $\tilde{F}(0, 0) = e_0$, then $\tilde{F}$ will be a homotopy between $\tilde{h}$ and the constant loop based at e_0.

(b) Let $H = p_*(\pi_1(E, e_0))$. Then if $\pi_1(B, b_0)/H$ is collection of right cosets of H, then the lifting correspondance of ϕ induces an injective map:

$$\Phi : \pi_1(B, b_0)/H \to p^{-1}(b_0)$$

Furthermore, Φ is bijective if E is path connected.

Proof:
Let f and g be loops in B based at b_0, and let $\tilde{f}$ and $\tilde{g}$ be the liftings of those loops in E to paths starting at e_0. Since H is a subgroup of $\pi_1(B, b_0)$, we know that $[g] \in H * [f]$ iff $H * [g] = H * [f]$. Thus, to show that Φ is well-defined, we just need to show that $[g] \in H * [f] \Longrightarrow \phi([g]) = \phi([f])$. Meanwhile, to show that Φ is injective, we just need to show that $[g] \in H * [f] \Longleftarrow \phi([g]) = \phi([f])$.

$(\Longrightarrow)$
If $[f] \in H * [g]$, then there exists $[h] \in H$ with $[f] = [h * g]$. And due to how we defined H, the lifting $\tilde{h}$ of h to E starting e_0 will be a loop.

Why: We know there exists a loop $\tilde{h}'$ in E based at e_0 such that $[h] = [p \circ \tilde{h}']$. But now by theorem 54.3, we have that the liftings $\tilde{h}$ and $\tilde{h}'$ of h and $p \circ \tilde{h}'$ respectively are path homotopic. So, $\tilde{h}$ is also a loop based at e_0.

It follows that the product $\tilde{h} * \tilde{g}$ is well-defined and is a lifting of $h * g$. That plus the fact that $[f] = [h * g]$ means that by theorem 54.3, $\tilde{f}$ and $\tilde{h} * \tilde{g}$ have the same end point. So:

$$\phi([f]) = \tilde{f}(1) = \tilde{h} * \tilde{g}(1) = \tilde{g}(1) = \phi([g])$$

$(\Longleftarrow)$
Next suppose $\phi([f]) = \phi([g])$. Then $\tilde{f}$ and $\tilde{g}$ end at the same point of E. So, we can define the loop $\tilde{h}$ in E based at e_0 as the product of $\tilde{f}$ and the reverse of $\tilde{g}$. Importantly, $[\tilde{h} * \tilde{g}] = [\tilde{f}]$. So, if $\tilde{F}$ is a path homotopy between $\tilde{h} * \tilde{g}$ and $\tilde{f}$, then $F := p \circ \tilde{F}$ is a path homotopy between $f = p \circ \tilde{f}$ and $p \circ (\tilde{h} * \tilde{g}) = (p \circ \tilde{h}) * g$. Also, $p \circ \tilde{h} \in H$. So $[f] \in H * [g]$.

If E is path connected, then we know ϕ is surjective. In turn Φ will also be surjective.

(c) If f is a loop in B based at b_0, then $[f] \in H$ iff f lifts to a loop in E based at e_0.

Proof:
Since Φ is injective and the constant loop about b_0 is mapped to e_0 by ϕ, we know that $\phi([f]) = e_0$ iff $[f] \in H$. But $\phi([f]) = e_0$ precisely iff f lifts to a loop in E based at e_0. ■

8/17/2025

Lemma 55.1: If A is a retract of X, then the homomorphism of fundamental groups induced by the inclusion map $j : A \to X$ is injective.

Proof:
If $r : X \to A$ is a retraction, then $r \circ j$ is just the identity map on A. It follows that $(r \circ j)_* = r_* \circ j_*$ is the identiy map on $\pi_1(A, a_0)$ for any $a_0 \in A$. This is only possible if r_* is surjective (which we admittedly already proved on page 121) and j_* is injective. ■

We'll denote $D^2 := \{x \in \mathbb{R}^2 : \|x\|_2 \leq 1\}$.

Theorem 55.2: There is no retraction from D^2 to S^1.

Proof:
If such a retraction did exist, then we would know by the previous lemma that there exists an injective group homomorphism from the fundamental group of S^1 to the fundamental group of D^2. However, the fundamental group of S^1 has strictly greater cardinality than that of D^2. So, no such injection exists. ■

Lemma 55.3: Suppose $h : S^1 \to X$ is a continuous map. Then the following are equivalent.

(1) h is nulhomotopic (meaning there is a homotopy from h to a constant function on X).

(2) h extends to a continuous map $k : D^2 \to X$.

(3) h_* is the trivial homomorphism of fundamental groups.

$(1 \implies 2)$
Let $H : S^1 \times [0, 1] \to X$ be a homotopy from h to some constant map. Then if we define $\pi : S^1 \times [0, 1] \to D^2$ by $\pi(x, t) = x(1-t)$, we have that π defines a quotient map.

It's obvious that π is continuous and surjective. Also, it's clear that π is an open function since any closed set of $S^1 \times [0, 1]$ is compact and thus maps to another compact set of D^2 which must be closed.

Now since H is constant over the preimage with respect to π of any given singleton, we know that H induces a continuous map $k : D^2 \to X$ satisfying that $H = k \circ \pi$. Also, it's clear that for $x \in S^1$, $k(x) = H(x, 0) = h(x)$.

$(2 \implies 3)$
Let $k : D^2 \to X$ be an extension of h and let $j : S^1 \to D^2$ be the inclusion map. Then $h = k \circ j$, meaning $h_* = k_* \circ j_*$. But now since the fundamental group of D^2 is trivial, we know that h_* must be the trivial homomorphism.

$(3 \implies 1)$
Let $p : \mathbb{R} \to S^1$ be the standard covering map (the one defined in theorem 53.1). Then $p_0 := p|_{[0,1]}$ is a generator for $\pi_1(S^1, 0)$. If we let $x_0 = h(b_0)$, then $f = h \circ p_0$ is a loop in X. And since h_* is trivial, we know there exists a path homotopy F from $h \circ p_0$ to the constant map x_0.

Now the map $(p_0 \times i) : [0,1]^2 \to S^1 \times [0,1]$ defined by $(p_0 \times i)(x,t) = (p_0(x), t)$ is a quotient map.

It's continuous and surjective. Also, once again it maps any closed set to another closed set since all closed subsets of $[0,1]^2$ are compact and $S^1 \times [0,1]$ is Hausdorff.

Since F is constant on $(p_0 \times i)^{-1}(y)$ for any $y \in S^1 \times [0,1]$, we have that F induces a continuous map $H : S^1 \times [0,1] \to X$ satisfying that $H \circ (p_0 \times i) = F$. Also, $x_0 = F(s,1) = H(p_0(s), 1)$ for all $s \in [0,1]$. This shows that H is our desired homotopy. ■

Corollary 55.4: The inclusion map $j : S^1 \to \mathbb{R} - \{0\}$ is not nulhomotopic. Also the identity map $i : S^1 \to S^1$ is not nulhomotopic.

Proof:
There is a retraction $k : \mathbb{R} - \{0\} \to S^1$ given by $k(x) = x/\|x\|_2$. It follows by lemma 55.1 that j_* is injective. And since the fundamental group of S^1 is not trivial, that means that j_* is not the trivial homomorphism. Hence, j isn't nulhomotopic by the last lemma.

Similarly, i_* is the identity homomorphism and thus isn't trivial.So i isn't nulhomotopic by the last lemma. ■

8/18/2025

Today I want to start learning about general manifolds. So, I will switch over to following John Lee's book Introduction to Smooth Manifolds. Soon I might switch back either to Munkres or Guillemin.

Suppose M is a topological space. Then we say M is a topological manifold of dimension n (a.k.a a topological n-manifold) if:

- M is second countable and Hausdorff.
- M is locally Euclidean of dimension n, meaning that for any $p \in M$, there exists an open neighborhood $U \subseteq M$ of p and a homeorphism φ from U to an open set $\hat{U}$ of H^n or $\mathbb{R}^n$.

If $n = 0$, we say M is a topological 0-manifold if M is countable and equipped with the discrete topology.

While it's clear that each pair (U, φ) define local coordinates on M, one weird thing notation-wise is that those coordinates are commonly written as:

$\varphi(p) = (x^1(p), \ldots, x^n(p))$ (as opposed to using subscripts).

This notation also extends to trivial Euclidean manifolds where (for example) John Lee denotes $x \in \mathbb{R}$ as $x = (x^1, \ldots, x^n)$. I don't know why this is apparently common notation in this field of math. (Also, to be clear my differential geometry professor before I dropped the course also used that notation. So, I'm not making it up that the notation is common....)

One manifold I haven't seen before is the n-dimensional real projective space, denoted $\mathbb{RP}^n$. It is defined as the set of 1-dimensional linear subspaces of $\mathbb{R}^{n+1}$ equipped with the quotient topology determined by the map $\pi : \mathbb{R}^{n+1} \to \mathbb{RP}^n$ where $x \mapsto \mathrm{span}\{x\}$.

Note: Given $x \in \mathbb{R}^{n+1}$, denote $[x] \coloneqq \pi(x)$.

To prove that $\mathbb{RP}^n$ is in fact a manifold, for each $i \in \{1, \ldots, n+1\}$ let $\tilde{U}_i = \{x \in \mathbb{R}^{n+1} : x^i \neq 0\}$. Then let $U = \pi(\tilde{U}_i)$. Now $\tilde{U}_i$ is a saturated set with respect to π (meaning $\pi^{-1}(\pi(U_i)) = U_i$) for each i. Hence $\pi(U_i)$ is open $\mathbb{RP}^n$ for each i.

Next, define a map $\varphi_i : U_i \to \mathbb{R}^n$ by:

$$\varphi([(x^1, \ldots, x^{n+1})]) \coloneqq (\tfrac{x^1}{x^i}, \ldots, \tfrac{x^{i-1}}{x^i}, \tfrac{x^{i+1}}{x^i}, \ldots, \tfrac{x^{n+1}}{x^i}).$$

If we scale $x \in \mathbb{R}^{n+1}$ by a nonzero constant, $\varphi_i([x])$ doesn't change. It follows that φ_i is a well-defined map. Also, φ_i is continuous since it's clear that $\varphi_i \circ \pi$ is continuous and π is a quotient map. Finally, to show that φ_i has a continuous inverse, note that:

$$\varphi_i^{-1}(u^1, \ldots, u^n) = [(u^1, \ldots, u^{i-1}, 1, u^{i+1}, \ldots, u^{n+1})].$$

Since φ_i^{-1} is the composition of a continuous function from $\mathbb{R}^n$ to $\mathbb{R}^{n+1}$ and π which is also continuous, we have that φ_i^{-1} is continuous. So φ_i is a homeomorphism.

Since the U_i cover $\mathbb{RP}^n$ we've thus shown that $\mathbb{RP}^n$ is locally Euclidean of dimension n. The proof that $\mathbb{RP}^n$ is second countable and Hausdorff is left as an exercise by Lee.

We know $\mathbb{RP}^n$ is second countable since it is the union of a finite number of open second countable sets.

To show that $\mathbb{RP}^n$ is Hausdorff, suppose that $[x_1], [x_2] \in \mathbb{RP}^n$ with $[x_1] \neq [x_2]$, and without loss of generality suppose x_1 and x_2 are unit vectors. Then x_1 and x_2 are not collinear. So, using the Hausdorffness of S^n, there is an open neighborhood $U_1 \subseteq S^n$ of x_1 such that $x_2, -x_2 \notin U_1$, and similarly there is an open neighborhood $U_2 \subseteq S^n$ of x_2 such that $x_1, -x_1 \notin U_2$.

Now define $V_i = \{cu : c \in \mathbb{R} - \{0\}, u \in U_i\}$ for each i. Then each V_i is easily checked to be a saturated open set with respect to π. Also, we have that $[x_2] \notin \pi(V_1)$ and $[x_1] \notin \pi(V_2)$. So $\pi(V_1)$ and $\pi(V_2)$ are disjoint open sets in $\mathbb{RP}^n$ separating $[x_1]$ and $[x_2]$.

One more note is that if we restrict π to just S^n, then we get a continuous surjective map from a compact set to $\mathbb{RP}^n$. This says that $\mathbb{RP}^n$ is compact. ■

Lemma 1.10: Every topological manifold M without boundary has a countable basis of precompact coordinate balls (i.e. sets that are homeomorphic to an open ball in $\mathbb{R}^n$) and coordinate half-balls (i.e. sets that are homeomorphic to $B_r(x) \cap H^n$ where $B_r(x)$ is a ball centered at a point in ∂H^n).

Proof:

Firstly, since every point in M has a coordinate patch (Lee uses the word "chart") about it, there exists an open covering $\{U_\alpha\}_{\alpha \in A}$ of M consisting of domains of coordinate patches $\varphi_\alpha : U_\alpha \to \hat{U}_\alpha \subseteq \mathbb{R}$.

Lemma: If X is a second countable space and $\{U_\alpha\}_{\alpha \in A}$ is an open cover of X, then there exists a countable subcover $\{U_{\alpha_n}\}_{n \in \mathbb{N}}$ of X.

Proof:

Let $\mathcal{B} = \{B_n\}_{n \in \mathbb{N}}$ be a countable basis for the topology of X. Then for each n we know there exists U_{α_n} with $B_n \subseteq U_{\alpha_n}$. Then in turn, since $\mathcal{B}$ is an open cover of X, we have that $\{U_{\alpha_n}\}_{n \in \mathbb{N}}$ also covers X.

It follows from that lemma there is a countable collection of coordinate charts $\varphi_{\alpha_n} : U_{\alpha_n} \to \hat{U}_{\alpha_n}$ covering M. Then, it's obvious how each φ_{α_n} defines a countable basis for U_{α_n} consisting of precompact coordinate balls and half balls, and if we take the union of all those bases for each n, we get a countable basis for all of M. ■

Corollary: Every topological manifold M is locally path connected (meaning there exists a basis of M consisting of path connected sets).

This is cause every coordinate ball is path connected on account of being homemorphic to a path connected set.

Now due to my patchy background, I'm only sorta familiar with how topological connectedness works. So, here is my attempt at reviewing / teaching some stuff to myself:

Here's what I already know / have proven in math 240B (Folland exercise 4.10):

- A topological space X is connected if there doesn't exist two nonempty open sets U and V in X such that $U \cap V = \emptyset$ and $U \cup V = X$.
- Equivalently, X is connected if the only two clopen sets are $\emptyset$ and X.

- If $\{E_\alpha\}_{\alpha \in A}$ is a collection of connected sets with nonempty intersection, then $E := \bigcup_{\alpha \in A} E_\alpha$ is connected.
- If $E \subseteq X$ is connected, then so is $\overline{E}$.
- For every $x \in X$ there is a maximal connected set $E \subseteq X$ containing x. This set is called a component of X. Additionally, we know that E is closed.
- A topological space X is path connected if for any $x, y \in X$, there is a continuous map $f : [0,1] \to X$ satisfying that $f(0) = x$ and $f(1) = y$.
- Any convex set in a topological vector space is path connected.

Here's some stuff I've used but not ever gotten around to proving before.

- If X is path connected, then X is connected.

  Proof:
  If X has only 1 element, then X is trivially both path-connected and connected.

  For the sake of contradiction, suppose X is path connected and that there exists nonempty open sets $U, V \subseteq X$ such that $U \cap V = \emptyset$ and $U \cup V = X$. Then pick $x \in U - V$ and $y \in V - U$. Since X is path connected, we know there is a path $f : [0,1] \to X$ from x to y.

  If we define $U' = f^{-1}(U)$ and $V' = f^{-1}(V)$, then the continuity of f guarentees that U' and V' are open subsets of $[0,1]$. Also, since U and V are disjoint and contain the entire range of f, we know that U' and V' are disjoint and their union is all of $[0,1]$. And since $0 \in U'$ and $1 \in V'$, neither sets are empty. All of this would points towards $[0,1]$ being a disconnected set.

  However, $[0,1]$ is connected.

  > Let U and V be two open sets partitioning $[0,1]$, and without loss of generality suppose $0 \in U$. Then set $\alpha := \sup\{s > 0 : [0,s) \subseteq U\}$. If $\alpha < 1$, then it'd be clear that $\alpha \notin U$. But it'd also be clear that α is not an interior point of $V = [0,1] - U$, which would contradict that V is open. So, we know that $\alpha = 1$. But now this requires that $[0,1] - U$ equals either $\{1\}$ or $\emptyset$, and only the latter is open in $[0,1]$. So there does not exist two open nonempty sets which form a partition of $[0,1]$.

  Hence, we have a contradiction and conclude that it is impossible for X to be path connected but not connected. ■

- If $f : X \to Y$ is continuous and X is connected, then $f(X)$ is connected. Similarly, if X is path connected, then $f(X)$ is path connected.

  Proof:
  If X is connected, let U and V be open sets in the subspace topology which partition $f(X)$. Then $f^{-1}(U)$ and $f^{-1}(V)$ are disjoint open subsets of X whose union is X. Since X is connected, it follows that either $f^{-1}(U)$ or $f^{-1}(V)$ is empty. And since $f(f^{-1}(U)) = U$ and $f(f^{-1}(V)) = V$ on account of the fact that both $U, V \subseteq f(X)$, we know either U or V is empty. So, $f(X)$ is connected.

If X is path connected, consider any $y_1, y_2 \in f(X)$. Then $\exists x_1, x_2 \in X$ with $f(x_i) = y_i$ for $i = 1, 2$, as well as a path $g : [0, 1] \to X$ going from x_1 to x_2. It follows that $f \circ g$ is a path going from y_1 to y_2 in $f(X)$. ■

- If $\{E_\alpha\}_{\alpha \in A}$ is a collection of path connected sets with nonempty intersection, then $E := \bigcup_{\alpha \in A} E_\alpha$ is path connected.

  Proof:
  Consider any $x, y \in E$, and let $z \in \bigcap_{\alpha \in A} E_\alpha$. Then there is path contained in one of the E_α going from x to z, and there is another path in another E_α going from z to y. Combining those paths gives us a path from x to y. ■

- E being path-connected does not necessarily mean that $\overline{E}$ is.

  Proof:
  Let $S = \{(t, \sin(1/t)) \in \mathbb{R}^2 : t > 0\}$. Then S is clearly path connected. Also, we can see that $\overline{S} = S \cup (\{0\} \times [-1, 1])$. But, there is no continuous path going from any point in $\{0\} \times [-1, 1]$ to any point in S. Hence $\overline{S}$ is not connected. ■

- For every $x \in X$ there is a maximal path connected set $E \subseteq X$ containing x. This set is called a path component of X.

  Proof:
  Let $\mathcal{E} = \{E_\alpha\}_{\alpha \in A}$ be the collection of every path connected set in X containing x. Note that $\mathcal{E}$ is not empty since $\{x\} \in \mathcal{E}$. Then $E = \bigcup_{\alpha \in A} E_\alpha$ is path connected. Also, clearly it is a maximal path connected set containing x. ■

- Let X be a topological space, and given $x \in X$ let E_x be the maximal connected component of X containing x. Then $\{E_x\}_{x \in X}$ forms a partition of X, meaning that if $E_x \cap E_y \neq \emptyset$, then $E_x = E_y$.

  Proof:
  Suppose $x, y \in X$ satisfy that $E_x \cap E_y \neq \emptyset$. Then $E_x \cup E_y$ is a connected subset containing both x and y. So, since E_x and E_y are maximal, we have that $E_x \cup E_y \subseteq E_x, E_y$. It follows that $E_x = E_x \cup E_y = E_y$. ■

- The previous bullet point also holds if we replace "maximal connected component" with "maximal path connected component".

- Two sets A and B are separated if $A \cap \overline{B} = \emptyset$ and $\overline{A} \cap B = \emptyset$. A topological space X is connected if and only if no two nonempty sets whose union is all of X are disconnected. (This is the definition in math 140a...)

  $(\Longrightarrow)$
  Let A and B be two nonempty separated sets satisfying that $A \cup B = X$. Then since $\overline{A} \cup B = X$ and $\overline{A} \cap B = \emptyset$, it's clear that $B = (\overline{A})^{\mathsf{C}}$. Similarly, it's clear that $A = (\overline{B})^{\mathsf{C}}$. So, both A and B are open. But now A and B are disjoint nonempty open sets partitioning X. This means that X is not connected.

$(\Longleftarrow)$
Let U and V be two disjoint nonempty open sets whose union is X. Then $V = U^\mathsf{C}$ and $U = V^\mathsf{C}$. This means U and V are closed and so $U = \overline{U}$, $V = \overline{V}$, and clearly both U and V are separated. ■

- If X is a topological space and $E \subseteq X$ is clopen, then E is a union of connected components.

  Proof:
  It suffices to show that if $x \in E$ and A_x is the connected component of X containing x, then $A_x \subseteq E$. Fortunately, since A_x is closed and E is clopen, we know that $A_x \cap E$ and $A_x - E$ are both closed subsets of X. In turn, we know that $A_x \cap E$ and $A_x - E$ are disjoint open sets in the relative topology of A_x whose union is all of A_x. Since A_x is connected, it follows that either $A_x \cap E = \emptyset$ or $A_x - E = \emptyset$. But the former case is not true since $x \in A_x \cap E$. So, we know that $A_x - E = \emptyset$. This shows that $A_x \subseteq E$. ■

Here's some nicher proofs:

- A space X is locally connected if X has a basis consisting of sets which are connected. Similarly, X is locally path connected if X has a basis consisting of sets which are path connected.

- If X is locally connected, then every connected component of X is open.

  Proof:
  Let E be a connected component of X. Then for all $y \in E$, there exists a connected open set U_y containing y. And since E is the maximal connected set containing y, we know $U_y \subseteq E$. So $E = \bigcup_{y \in E} U_y$ is open. ■

- By identical reasoning to the last bullet point, if X is locally path connected, then every path component of X is open.
- Suppose X is locally path connected. Then X is connected if and only if X is path connected.

  $(\Longrightarrow)$
  It suffices to show that X has only one path component. Luckily, if X had more than one, then since every path component is open, we'd be able to take the union of all but one in order to get two disjoint nonempty open sets whose union is all of X. But this contradicts that X is connected.

  $(\Longleftarrow)$
  We proved this direction several bullet points ago. ■

- If $\{X_\alpha\}_{\alpha \in A}$ is a family of connected topological spaces, then the product space $X = \prod_{\alpha \in A} X_\alpha$ is also connected.

  Proof:
  Let $<$ be a well-ordering of A, and for any $\beta \in A$ let $S_\beta = \{\alpha \in A : \alpha < \beta\}$ and $\overline{S_\beta} = S_\beta \cup \{\beta\}$. We shall proceed via transfinite induction.

Let $\langle x_\alpha \rangle_{\alpha \in A} \in \prod_{\alpha \in A} X_\alpha$. Then suppose that $\beta \in A$ satisfies that $\left(\prod_{\alpha \in \overline{S_\gamma}} X_\alpha\right) \times \left(\prod_{\alpha \in (\overline{S_\gamma})^\complement} \{x_\alpha\}\right)$ is connected for every $\gamma \in S_\beta$.

We claim that $\left(\prod_{\alpha \in \overline{S_\beta}} X_\alpha\right) \times \left(\prod_{\alpha \in (\overline{S_\beta})^\complement} \{x_\alpha\}\right)$ is connected.

To prove this, first note that if $\beta \in A$, then $X_\beta \times \prod_{\alpha \in A - \{\beta\}} \{x_\alpha\}$ is connected.

> Proof:
> Let U and V be disjoint open sets which partition $X_\beta \times \prod_{\alpha \in A - \{\beta\}} \{x_\alpha\}$, and let π_β be the projection of X onto X_β. Then $\pi_\beta(U)$ and $\pi_\beta(V)$ are disjoint open sets which partition X_β. It follows that one of those two sets is empty, and the only way that is possible is if either U or V is empty.

If $\overline{S_\beta} = \{\beta\}$, then this already proves our claim. Otherwise, notice that:

$$\left(\prod_{\alpha \in \overline{S_\beta}} X_\alpha\right) \times \left(\prod_{\alpha \in (\overline{S_\beta})^\complement} \{x_\alpha\}\right) = \bigcup_{\gamma \in S_\beta} \left(X_\beta \times \left(\prod_{\alpha \in \overline{S_\gamma}} X_\alpha\right) \times \left(\prod_{\alpha \in (\overline{S_\gamma})^\complement - \{\beta\}} \{x_\alpha\}\right)\right)$$

Now each of the sets in that union contain $\langle x_\alpha \rangle$. Also, we claim that each of them are connected. After all, note that:

$$X_\beta \times \left(\prod_{\alpha \in \overline{S_\gamma}} X_\alpha\right) \times \left(\prod_{\alpha \in (\overline{S_\gamma})^\complement - \{\beta\}} \{x_\alpha\}\right) = \bigcup_{y \in X_\beta} \left(\{y\} \times \left(\prod_{\alpha \in \overline{S_\gamma}} X_\alpha\right) \times \left(\prod_{\alpha \in (\overline{S_\gamma})^\complement - \{\beta\}} \{x_\alpha\}\right)\right)$$

We already know that $\{x_\beta\} \times \left(\prod_{\alpha \in \overline{S_\gamma}} X_\alpha\right) \times \left(\prod_{\alpha \in (\overline{S_\gamma})^\complement - \{\beta\}} \{x_\alpha\}\right)$ is connected.

Also, for any $y \in X_\beta$, we can find a continuous map f from X to itself which sets the βth coordinate of a point to y and otherwise acts as the identity for all other coordinates. (We know this map is continuous because $\pi_\alpha \circ f$ is trivially continuous for all $\alpha \in A$...) This in turn shows that:

$$\{y\} \times \left(\prod_{\alpha \in \overline{S_\gamma}} X_\alpha\right) \times \left(\prod_{\alpha \in (\overline{S_\gamma})^\complement - \{\beta\}} \{x_\alpha\}\right) \text{ is connected for all } y.$$

And since $X_\beta \times \prod_{\alpha \in A - \{\beta\}} \{x_\alpha\}$ is a connected set intersecting the set in the previous paragraph for all y and is a subset of the union:

$$\bigcup_{y \in X_\beta} \left(\{y\} \times \left(\prod_{\alpha \in \overline{S_\gamma}} X_\alpha\right) \times \left(\prod_{\alpha \in (\overline{S_\gamma})^\complement - \{\beta\}} \{x_\alpha\}\right)\right),$$

we've thus shown that the entire union is connected. This finishes proving our claim at the top of this page.

By transfinite induction, we can now conclude that:

$$\left(\prod_{\alpha \in \overline{S_\beta}} X_\alpha\right) \times \left(\prod_{\alpha \in (\overline{S_\beta})^\complement} \{x_\alpha\}\right) \text{ is connected for all } \beta \in A.$$

Finally, to finish our proof we can just note that:

$$X = \bigcup_{\beta \in A} \left(\left(\prod_{\alpha \in \overline{S_\beta}} X_\alpha\right) \times \left(\prod_{\alpha \in (\overline{S_\beta})^\complement} \{x_\alpha\}\right)\right)$$

Also, every set in the above union is connected and contains $\langle x_\alpha \rangle_{\alpha\in A}$. Hence X is connected.

- If $\{X_\alpha\}_{\alpha\in A}$ is a family of path connected topological spaces, then the product space $X = \prod_{\alpha\in A} X_\alpha$ is also path connected.

  Proof:
  Let $\langle x_\alpha \rangle_{\alpha\in A}$ and $\langle y_\alpha \rangle_{\alpha\in A}$ be elements of X. Then we know for each $\alpha \in A$ that there is a path $f_\alpha : [0,1] \to X_\alpha$ such that $f_\alpha(0) = x_\alpha$ and $f_\alpha(1) = y_\alpha$. If we now define $f : [0,1] \to X$ by $f(t) = \langle f_\alpha(t) \rangle_{\alpha\in A}$, we will have that f is a continuous path from $\langle x_\alpha \rangle_{\alpha\in A}$ to $\langle y_\alpha \rangle_{\alpha\in A}$.
  (It is continuous because $\pi_\alpha \circ f = f_\alpha$ is continuous for all $\alpha \in A$.) ■

- Every quotient space of a connected space is connected. Also, every quotient space of a path connected space is path connected.

  Proof:
  Let X^* be a partition of X and let $f : X \to X^*$ be the function mapping every element to the set in X^* containing it. If we equip X^* with the quotient topology with resepct to f, then f will be continuous. Hence since X is connected, so will $f(X) = X^*$.

  Similar reasoning works when X is path connected. ■

- If X is a locally connected space, then every open set in X is locally connected. Similarly, if X is a locally path connected space, then so is every open set in X.

  This should be obvious.

- If $(X, \mathcal{T})$ is a topological space and $\mathcal{T}'$ is a coarser topology on $\mathcal{T}$, then $(X, \mathcal{T})$ being connected implies that $(X, \mathcal{T}')$ is connected. Similarly, $(X, \mathcal{T})$ being path connected implies that $(X, \mathcal{T}')$ is path connected. Another way of thinking about this is that adding sets to a topology only makes your space more disconnected.

  Proof:
  If $(X, \mathcal{T}')$ weren't connected, then the disjoint sets in $\mathcal{T}'$ partitioning X would also be an open partition in $(X, \mathcal{T})$.

  Next, if $f : [0,1] \to X$ is continuous with respect to $\mathcal{T}$, then we know it is also continuous with respect to $\mathcal{T}'$. This shows that a path with respect to $\mathcal{T}$ is also a path with respect to $\mathcal{T}'$. ■

8/22/2025

To start off today, I'm going to do an exercise from Folland.

Exercise 4.57: A collection $\mathcal{U}$ of open sets in X is called locally finite if each $x \in X$ has a neighborhood that intersects only finitely many members of $\mathcal{U}$. If $\mathcal{U}$ and $\mathcal{V}$ are open covers of X, $\mathcal{V}$ is a refinement of $\mathcal{U}$ if for each $V \in \mathcal{V}$ there exists $U \in \mathcal{U}$ with $V \subseteq U$. X is called paracompact if every open cover of X has a locally finite refinement.

Clearly any compact set is automatically paracompact since a finite subcover of an open cover will automatically be a locally finite refinement of that cover. However, beware that a refinement of a cover doesn't need to be a subset of the original cover, and it is possible for a cover to be locally finite without being finite.

The following exercise generalizes the theorems written on pages 81 and 82 of this journal.

(a) If X is a σ-compact LCH space, then X is paracompact. In fact, every open cover $\mathcal{U}$ has locally finite refinements $\{V_\alpha\}_{\alpha\in A}$ and $\{W_\alpha\}_{\alpha\in A}$ such that $\overline{V_\alpha}$ is compact and $\overline{W}_\alpha \subseteq V_\alpha$ for all $\alpha \in A$.

Let $(U_n)_{n\in\mathbb{N}}$ be an increasing sequence of precompact open sets such that $\overline{U_n} \subseteq U_{n+1}$ and $X = \bigcup_{n\in\mathbb{N}} U_n$. (Since X is a σ-compact LCH space, we proved in math 240b that such a sequence must exist...) Also, for ease of notation take $U_n = \emptyset$ whenever $n \leq 0$

Now, the collection $\{E \cap (U_{n+2} - \overline{U_{n-1}}) : E \in \mathcal{U}\}$ is an open cover of $\overline{U_{n+1}} - U_n$. And since $\overline{U_{n+1}} - U_n$ is compact, we can choose a finite subcover $\mathcal{V}_n$ from that collection. Doing this for all n and then setting $\mathcal{V} = \bigcup_{n\in\mathbb{N}} \mathcal{V}_n$, we have that $\mathcal{V}$ is a locally finite refinement of $\mathcal{U}$. After all, each x has a neighborhood $U_{n+2} - \overline{U_{n-1}}$ which only the finitely many open sets in $\mathcal{V}_n, \mathcal{V}_{n+1}, \mathcal{V}_{n+2}$, $\mathcal{V}_{n-1}$ and $\mathcal{V}_{n-2}$ can intercept. Also, each set in $\mathcal{V}$ is contained in some set of $\mathcal{U}$. And thirdly, we claim that if $V_\alpha \in \mathcal{V}$, then V_α is precompact. This is because V_α will be a closed subset of $\overline{U_{n+2}}$ for some n and the latter is compact.

Having constructed our first refinement $\mathcal{V} = \{V_\alpha\}_{\alpha\in A}$, we're now ready to construct our second. Fix n and note that each $x \in X$ has a compact neighborhood $N_x \subseteq V_\alpha$ for some $V_\alpha \in \mathcal{V}_n$. The N_x° form an open cover of $\overline{U_{n+1}} - U_n$. Hence, there is a finite collection $\{x_1, \ldots, x_m\}$ of points in $\overline{U_{n+1}} - U_n$ such that $\overline{U_{n+1}} - U_n \subseteq \bigcup_{j=1}^m N_{x_j}^\circ$. So, for each $\alpha \in A$ with $V_\alpha \in \mathcal{V}_n$ let W_α be the union of all the $N_{x_j}^\circ$ such that $N_{x_j} \subseteq V_\alpha$, and let $\mathcal{W}_n$ be the collection of W_α defined in this sentence. It's clear that $\mathcal{W}_n$ is also an open cover of $\overline{U_{n+1}} - U_n$, and that $\overline{W_\alpha} \subseteq V_\alpha$ for all $V_\alpha \in \mathcal{V}_n$. Doing this for all n, we then have that $\mathcal{W} = \bigcup_{n\in\mathbb{N}} \mathcal{W}_n$ is our second desired refinement.

(b) If X is a σ-compact LCH space, for any open cover $\mathcal{U}$ of X there is a partition of unity on X subordinate to $\mathcal{U}$ and consisting of compactly supported functions.

Let $\{V_n\}_{n\in\mathbb{N}}$ and $\{W_n\}_{n\in\mathbb{N}}$ be refinements of $\mathcal{U}$ constructed as in part (a). Note that our refinements in the last part were countable. So, there is no issue just taking the indexing set A to be $\mathbb{N}$.

For each n there exists by Urysohn's lemma a function $f_n \in C_c(X, [0,1])$ such that $f_n(x) = 1$ for all $x \in \overline{W_n}$ and $f_n(x) = 0$ for all $x \in V_n^{\mathsf{C}}$. Setting $f = \sum_{n=1}^\infty f_n$, we then have that f is continuous and finite everywhere in X on account of the fact that every $x \in X$ has a neighborhood where f is only a sum of finitely many continuous functions on that neighborhood. Also, since every $x \in X$ is contained in W_n, we know that $f(x) \geq 1$ for all x. Hence, if we define $g_n = f_n/f$ for each n, we have that g_n is well-defined and still in $C_c(X, [0,1])$.

We claim $(g_n)_{n\in\mathbb{N}}$ is a partition of unity subordinate to $\mathcal{U}$. After all, $\sum_{n\in\mathbb{N}} g_n = \frac{1}{f}\sum_{n\in\mathbb{N}} f_n = 1$. Also, each g_n satisfies that $\operatorname{supp}(g_n) \subseteq V_n$ and $V_n \subseteq U$ for some $U \in \mathcal{U}$. And finally, it is still the case that every $x \in X$ has a neighborhood on which only finitely many of the g_n are nonzero.

As a side note, we can extend this result to proving theorem 16.3 on page 82 of this journal by just noting that $\mathbb{R}^n$ is a σ-compact LCH space and that the Urysohn lemma on $\mathbb{R}^n$ specifies that we can choose each of our f_n to be in $C_c^\infty(X, [0,1])$.

Now I shall go over some more of John Lee's book.

Since manifolds are locally path connected, we know that a manifold is connected if and only if it is path connected. Also, it is clear that the path components of a manifold are identical to it's components. One more proposition is as follows:

Proposition 1.11.d: If M is a topological manifold, then M has countably many components, each of which are open subsets of M and a topological manifold by themselves.

Proof:
Since M is locally path connected, we know every single component is an open set. It follows that the components form an open cover $\mathcal{U}$ of M. And since M is second countable, this means that there is a countable subcover of $\mathcal{U}$. Yet, because all the sets of $\mathcal{U}$ are disjoint, the only way this is possible is if $\mathcal{U}$ was countable to begin with.

Also, it's clear that every component equipped with the subspace topology will still be second countable and Hausdorff. And to show that each component is locally Euclidean of dimension n, just restrict the domain and codomain of the coordinate patches covering that component. ■

Another consequence of every manifold having a countable basis of precompact coordinate balls and half balls is that manifolds are locally compact and σ-compact. And by the exercise I did earlier today, that means that every manifold is paracompact.

In fact, by slightly modifying our construction of $\mathcal{W}$ in the exercise I did before (namely by picking each N_x to be the closure of a precompact coordinate ball and then letting $\mathcal{W}_n$ consist of the $N^\circ_{x_j}$ without bothering to take any unions), we can construct a locally finite refinement consisting of coordinate balls and half balls for any open cover of M.

Here is a proposition about locally finite collections of sets.

Exercise 1.14: Suppose $\mathcal{X}$ is a locally finite collection of subsets of a topological space M.

(a) The collection $\{\overline{X} : X \in \mathcal{X}\}$ is also locally finite.

Proof:
Let $x \in M$ and let $U \subseteq M$ be an open set containing x that intersects only the elements of a finite subset $\mathcal{Y}$ of $\mathcal{X}$. Now we want to show that if $X \in \mathcal{X}$ satisfies $\overline{X} \cap U \neq \emptyset$, then $X \in \mathcal{Y}$. Luckily, if $\overline{X} \cap U \neq \emptyset$ but $X \cap U = \emptyset$ so that $X \notin \mathcal{Y}$, then it must be that any $x \in \overline{X} \cap U$ is an accumulation point of X. However, that is immediately contradicted by the fact that U is a neighborhood of x which does not intercept X anywhere. ■

(b) $\overline{\bigcup_{X \in \mathcal{X}} X} = \bigcup_{X \in \mathcal{X}} \overline{X}$.

Proof:
It is always true that $\bigcup_{X \in \mathcal{X}} \overline{X} \subseteq \overline{\bigcup_{X \in \mathcal{X}} X}$.

Meanwhile, to show the other inclusion, suppose x is an accumulation point of $\bigcup_{X \in \mathcal{X}} X$ but not in any individual $\overline{X}$. Then it must be the case that every neighborhood of x intercepts some set in $\mathcal{X}$, yet it must also be the case that for every $X \in \mathcal{X}$ there is a neighborhood of x with doesn't intercept X. The only way this is possible is if every neighborhood intercepts infinitely many $X \in \mathcal{X}$.

Let N be any neighborhood of x. Then we can construct an infinite sequence of distinct sets $(X_n)_{n \in \mathbb{N}}$ in $\mathcal{X}$ intercepting N as follows.

For the ease of notation set $U_0 = M$. Now at step n, choose any $X_n \in \mathcal{X}$ that intersects $N \cap \bigcap_{k=0}^{n-1} U_k$. Note that we can do this since a finite intersection of neighborhoods of x is still a neighborhood of x. But now we can choose some neighborhood U_n of x such that $X_n \cap U_n = \emptyset$. And now repeat this reasoning.

It's clear that all the chosen X_n intercept $\bigcap_{k=0}^{n-1} U_k \subseteq N$. Additionally, all the chosen X_n are distinct since $X_n \cap \bigcap_{k=0}^{N} U_k \subseteq X_n \cap U_n = \emptyset$ for all $N \geq n$.

But that contradicts that $\mathcal{X}$ is locally finite. ■

<u>Proposition 1.16:</u> The fundamental group of a topological manifold M is countable.

Proof:
Let $\mathcal{B}$ be a countable collection of coordinate balls and half balls covering M. For any $B, B' \in \mathcal{B}$, the intersection $B \cap B'$ has at most countably many components each of which is path connected.

Why? $B \cap B'$ is second countable and locally path connected on account of being an open subset of M. In turn, by the same reasoning as in proposition 1.11.d we know that $B \cap B'$ has only countably many components.

Let $\mathcal{X}$ be a countable set containing a point from each component of $B \cap B'$ for every $B, B' \in \mathcal{B}$ (including $B = B'$). Also, for each $B \in \mathcal{B}$ and $x, x' \in \mathcal{X}$ satisfying that $x, x' \in B$, let $h^B_{x,x'}$ be some path from x to x' in B. Since $\mathcal{X}$ intercepts every component of M, it suffices when calculating the fundamental group to take our base point p to be in $\mathcal{X}$. Then, we define a *special loop* to be a loop based at p that is a finite product of paths $h^B_{x,x'}$.

Now there are only countably many special loops. Therefore, in order to prove that $\pi_1(M,p)$ is countable, it suffices to show that if $f:[0,1]\to M$ is a loop based at p, then f is homotopic to some special loop.

Fortunately, the collection of the components of the sets $f^{-1}(B)$ with $B\in\mathcal{B}$ is an open cover of $[0,1]$. So, there exists $0=a_0<a_1<\cdots<a_k=1$ such that for each $i\geq 1$, $[a_{i-1},a_i]\subseteq f^{-1}(B)$ for some $B\in\mathcal{B}$. Now for each i, let f_i be the restriction of f to interval $[a_{i-1},a_i]$ and then reparametrized so that its domain is $[0,1]$, and also let $B_i\in\mathcal{B}$ be a coordinate ball containing the image of f_i. For each $1\leq i<k$, $f(a_i)\in B_i\cap B_{i+1}$. Also there is some $x_i\in\mathcal{X}$ that lies in the same component of $B_i\cap B_{i+1}$ as $f(a_i)$. So, let g_i be a path in $B_i\cap B_{i+1}$ from x_i to $f(a_i)$.

Note, we'll also write $x_0=x_k=p$ and $g_0=g_k$ are the constant paths based at p. Also, like in Munkres, we'll denote $\bar{g}_i$ to be the reverse path.

Now if we denote $\tilde{f}_i\coloneqq g_{i-1}*f_i*\bar{g}_i$, we have that:

$$\begin{aligned} f &\simeq_p f_1*\cdots*f_k \\ &\simeq_p g_0*f_1*\bar{g}_1*g_1*f_2*\bar{g}_2*g_2*\cdots*\bar{g}_{k-1}*g_{k-1}*f_k*\bar{g}_k \\ &\simeq_p \tilde{f}_1*\cdots*\tilde{f}_k. \end{aligned}$$

But now since each B_i is simply connected, we know that $\tilde{f}_i$ is path homotopic to $h^{B_i}_{x_{i-1},x_i}$. In turn, we have that f is path homotopic to a special loop. ■

A question I've had for a while is how smoothness can even be defined for manifolds which aren't subsets of $\mathbb{R}^n$. After all, differentiability as a concept relies on the topological structure of the reals. It turns out that the answer to this problem is to reframe some prior theorems as defining axioms which must be met.

Let M be a topological n-manifold. Two charts (U,φ) and (V,ψ) on M are said to be C^r compatible if either $U\cap V=\emptyset$ or the transition map $\psi\circ\varphi^{-1}$ from $\varphi(U\cap V)$ to $\psi(U\cap V)$ is a C^r diffeomorphism.

Note that the transition map is a function from an open open subset of either H^n or $\mathbb{R}^n$ to another open subset of either H^n or $\mathbb{R}^n$. So, it makes sense to talk about the transition map as being differentiable.

An atlas $\mathcal{A}$ for M is a collection of charts covering M. $\mathcal{A}$ is called a C^r atlas if any two charts in $\mathcal{A}$ are C^r compatible.

Very roughly speaking, if $1\leq s,r$, we want to define that a manifold M is C^r smooth if M has a C^r atlas $\mathcal{A}$. Additionally, we want to then say that a function $f:M\to\mathbb{R}$ is C^s differentiable if and only if $f\circ\varphi^{-1}$ is C^s differentiable for all charts $(U,\varphi)\in\mathcal{A}$.

A snag we need to work out though is that a space can have many C^r atlases. Also, some of those atlases may have charts which aren't C^r compatible with the charts in the other atlases. In a sense, that would mean they define different "smooth structures". At the same time, it could be the case that two atlases contain charts which are all C^r compatible with the charts in the other atlas. In that case, the two atlases could be said to define the same "smooth structure". We'll get around this issue as follows:

> We define a C^r atlas $\mathcal{A}$ on a topological manifold M to be maximal or complete if it is not contained in a larger C^r atlas. Or in other words, if a chart (U, φ) is C^r compatible with every chart in $\mathcal{A}$, then $(U, \varphi) \in \mathcal{A}$.
>
> If M is a topological manifold, then a C^r structure on M is a maximal C^r atlas.
>
> A C^r manifold is a pair $(M, \mathcal{A})$ where M is a topological manifold and $\mathcal{A}$ is a C^r structure on M.

Proposition 1.17: Let M be a topological manifold.

(a) Every C^r atlas $\mathcal{A}$ for M is contained in a unique maximal C^r atlas called the C^r structure determined by $\mathcal{A}$.

Proof:
Let $\mathcal{A}$ be a C^r atlas for M and define $\overline{\mathcal{A}}$ as the set of all charts which are C^r compatible with every chart in $\mathcal{A}$. If $\overline{\mathcal{A}}$ were a C^r atlas, it would be obvious that it is maximal. After all, if a chart was C^r compatible with every chart in $\overline{\mathcal{A}}$, then it would also be compatible with every chart in $\mathcal{A}$ on account of the fact that $\mathcal{A} \subseteq \overline{\mathcal{A}}$. But that would mean that that chart is also in $\overline{\mathcal{A}}$.

Hence, we proceed by trying to prove that $\overline{\mathcal{A}}$ is a C^r atlas on M. Or in other words, we want to show that for any $(U, \varphi), (V, \psi) \in \overline{\mathcal{A}}$, with $U \cap V \neq \emptyset$, the map $\psi \circ \varphi^{-1} : \varphi(U \cap V) \to \psi(U \cap V)$ is smooth.

Choose $x = \varphi(p) \in \varphi(U \cap V)$. Then there is some chart $(W, \theta) \in \mathcal{A}$ with $p \in W$. And since every chart in $\overline{\mathcal{A}}$ is C^r compatible with (W, θ), we know that $\theta \circ \varphi^{-1}$ and $\psi \circ \theta^{-1}$ are both C^r maps. It follows that $\psi \circ \varphi^{-1} = (\psi \circ \theta^{-1}) \circ (\theta \circ \varphi^{-1})$ is a C^r map on the neighborhood $\varphi(U \cap V \cap W)$ of x.

This proves that $\phi \circ \varphi^{-1}$ is locally C^r. Hence it is C^r in general.

With that, we've prove that there exists a maximal C^r atlas $\overline{\mathcal{A}}$ containing $\mathcal{A}$. To finish off we show uniqueness. Suppose $\mathcal{B}$ is another maximal C^r atlas containg $\mathcal{A}$. Then every chart in $\mathcal{B}$ must be C^r compatible with every char in $\mathcal{A}$. But that then implies that $\mathcal{B} \subseteq \overline{\mathcal{A}}$. And since $\mathcal{B}$ is maximal, we have that $\mathcal{B} = \overline{\mathcal{A}}$. ■

(b) Two C^r atlases determine the same C^r structure iff their union is is a C^r atlas.

Proof:
Let $\mathcal{A}$ and $\mathcal{B}$ be C^r atlases and let $\overline{\mathcal{A}}$ and $\overline{\mathcal{B}}$ denote the smooth structures determined by $\mathcal{A}$ and $\mathcal{B}$ respectively.

$(\Longrightarrow)$
If $\overline{\mathcal{A}} = \overline{\mathcal{B}}$, then we know $\mathcal{B} \subseteq \overline{\mathcal{A}}$. So, every chart of $\mathcal{B}$ is C^r compatible with every chart of $\mathcal{A}$. It follows that any two charts in $\mathcal{A} \cup \mathcal{B}$ are smoothly compatible.

$(\Longleftarrow)$
If $\mathcal{A} \cup \mathcal{B}$ is a C^r atlas, then we know that $\mathcal{A} \cup \mathcal{B} \subseteq \overline{\mathcal{A}}$ and that $\mathcal{A} \cup \mathcal{B} \subseteq \overline{\mathcal{B}}$. It follows that both $\overline{\mathcal{A}}$ and $\overline{\mathcal{B}}$ are the unique smooth structure determined by $\mathcal{A} \cup \mathcal{B}$. So, $\mathcal{A} = \mathcal{B}$. ■

Note that in the prior reasoning, we took $r \in \mathbb{Z}_{>0} \cup \{\infty\}$. If $r = \infty$, we call a C^r manifold a smooth manifold.

Before doing the following exercise, I want to establish a useful result.

Proposition: Let M be a topological manifold. Also let $\mathcal{A}$ be a C^r atlas for M and $\overline{\mathcal{A}}$ be the C^r structure determined by $\mathcal{A}$.

(a) If (U, φ) is a chart in $\mathcal{A}$, then given any C^r diffeomorphism h acting on $\varphi(U)$, we have that the chart $(U, h \circ \varphi) \in \overline{\mathcal{A}}$.

Proof:
By the prior proposition it suffices to show that given another chart (V, ψ) in $\mathcal{A}$ such that $V \cap U \neq \emptyset$, $\psi \circ (h \circ \varphi)^{-1}$ and $(h \circ \varphi) \circ \psi^{-1}$ are C^r maps. Luckily, since h, h^{-1}, $\varphi \circ \psi^{-1}$, and $\psi \circ \varphi^{-1}$ are all C^r maps, this is obvious. ■

(b) If (U, φ) is a chart in $\mathcal{A}$, then given any open set $V \subseteq U$, $(V, \varphi|_V)$ is a chart in $\overline{\mathcal{A}}$.

Hopefully this is obvious.

The significance of the above lemma is that we can cut up and smoothly reparametrize a coordinate chart and we'll still get a chart which is in the C^r structure we've equipped our manifold with.

Problem 1-6: Let M be a nonempty topological manifold of dimension $n \geq 1$. If M has a C^r (or smooth) structure, then show that it has uncountably many distinct ones.

Proof:
We start by proving the following lemma...

Lemma: If $s > 0$, then $F_s(x) := \|x\|_2^{s-1}x$ defines a homeomorphism from B^n to itself and from $H^n \cap B^n$ to itself (where B^n is the open unit ball in $\mathbb{R}^n$). Also, F_s is a C^∞ diffeomorphism on $\mathbb{R}^n - \{0\}$, and F_s is also a C^k diffeomorphism (for any k) on $\mathbb{R}^n$ if and only if $s = 1$.

Proof:
It's clear that F_s is a continuous function on $\mathbb{R}^n$. Also if $x \neq 0$ satisfies that $\|x\|_2^{s-1}x = y$, then $x = \|x\|_2^{1-s}y$ and $\|y\|_2 = \|x\|_2^s$. In turn $\|y\|_2^{(1-s)/s} = \|x\|_2^{1-s}$ and we have thus derived the following inverse function for F_s:

$$F_s^{-1}(x) = \begin{cases} 0 & \text{if } x = 0 \\ \|x\|_2^{(1-s)/s}x & \text{if } x \neq 0 \end{cases}$$

We claim that F_s^{-1} is continuous. This is clearly true on $\mathbb{R}^n - \{0\}$. Meanwhile, to show that $\|x\|_2^{(1-s)/s}x \to 0$ as $x \to 0$, note that:

$$\left\| \|x\|_2^{(1-s)/s}x \right\|_2 = \|x\|_2^{1+\frac{1-s}{s}} = \|x\|_2^{1/s}.$$

The latter clearly goes to 0 as x goes to 0. So F_s^{-1} is continuous at 0. This proves that F_s is a homeomorphism from $\mathbb{R}^n$ to itself. To show that restricting F_s defines a homeomorphism on B^n or $H^n \cap B^n$ just requires noting that both F_s and F_s^{-1} map B^n into B^n and H^n into H^n. Also, since we have formulas for F_s and F_s^{-1}, we can now clearly see that F_s and F_s^{-1} are smooth on $\mathbb{R}^n - \{0\}$.

Finally, note that if $s = 1$, then $F_s(x) = x$ is clearly a smooth diffeomorphism. Meanwhile, if $s \neq 1$, then we either have that $s - 1 < 0$ or $(1-s)/s < 0$. In the former case, we know that F_s is not differentiable at 0. After all, given any $A \in L(\mathbb{R}^n, \mathbb{R}^n)$, we have that:

$$\frac{\|F_s(h) - F_s(0) - Ah\|_2}{\|h\|_2} = \frac{\|\|h\|^{s-1}h - Ah\|_2}{\|h\|_2} = \left\|(\|h\|^{s-1}I - A)\frac{h}{\|h\|_2}\right\|_2$$

It follows then that there is some sequence $(h_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^n$ converging to 0 such that:

$$\left\|(\|h_n\|^{s-1}I - A)\frac{h_n}{\|h_n\|_2}\right\|_2 = \|(\|h_n\|^{s-1}I - A)\|_{\text{op}} \text{ for all } n.$$

I'll also note that by taking the negative of any neccesary h_n, we can force $(h_n)_{n\in\mathbb{N}}$ to be a sequence in H^n. I'm not actually sure if this is strictly necessary but who cares.

But now $\|(\|h_n\|^{s-1}I - A)\|_{\text{op}} \geq \left|\|h_n\|_{\text{op}}^{s-1} - \|A\|_{\text{op}}\right| \to \infty$ as $n \to \infty$ since $\|h_n\|^{s-1} \to \infty$ as $h_n \to 0$. This proves that the derivative of F_s at 0 doesn't exist when $s - 1 < 0$. Analogous reasoning shows that the derivative of F_s^{-1} at 0 doesn't exist when $(1-s)/s < 0$. So, F_s is not a diffeomorphism at 0 when $s \neq 1$.

Now let $\mathcal{A}$ be an atlas contained in our C^r structure for M. Then choose $p \in M$ and let (U, φ) be a chart in M containing p. By using the proposition I showed before doing this exercise, we can define another C^r atlas $\mathcal{A}'$ in the same structure on M such that (U, φ) is the only chart containing p. Specifically, define:

$$\mathcal{A}' := \{(V - \{p\}, \psi|_{V-\{p\}}) : (V, \psi) \in \mathcal{A}\} \cup \{(U, \varphi)\}.$$

Next, by restricting φ to a coordinate ball or half ball W centered at p and then reparametrizing, we can say that there is a chart (W, φ') in our structure on M satisfying that $p \in W$, $\varphi'(p) = 0$, and $\varphi'(W)$ equals either B^n or $H^n \cap B^n$. Hence, we define the C^r atlas:

$$\mathcal{A}'' := \{(V - \{p\}, \psi|_{V-\{p\}}) : (V, \psi) \in \mathcal{A}\} \cup \{(W, \varphi')\}$$

And now we're in a position to use our lemma. Given $s > 0$, let F_s be as in our lemma and define the atlas:

$$\mathcal{B}_s := \{(V - \{p\}, \psi|_{V-\{p\}}) : (V, \psi) \in \mathcal{A}\} \cup \{(W, F_s \circ \varphi')\}$$

Note that $\mathcal{B}_s$ is in fact an atlas for every s since F_s is a homeomorphism, meaning that $(W, F_s \circ \varphi')$ is a well-defined chart in M. Also, we can see that $\mathcal{B}_s$ is actually a C^r atlas. After all, we know from before that every pair of charts in $\mathcal{B}_s$ not including $(W, F_s \circ \varphi')$ are C^r compatible. Meanwhile, note that if $(V, \psi) \in \mathcal{A}$ satisfies that $(V - \{p\}) \cap W \neq \emptyset$, then F_s is a diffeomorphism on the set $\varphi'(V \cap W - \{p\})$ on account of $0 = \varphi'(p)$ not being in that set. It follows easily that $(F_s \circ \varphi') \circ \psi^{-1}$ defined on $\psi(V \cap W - \{p\})$ is a C^r diffeomorphism.

But now note that if $s \neq t$, then $\mathcal{B}_s$ and $\mathcal{B}_t$ do not generate the same C^r structure on M. After all, the charts $(W, F_s \circ \varphi')$ and $(W, F_t \circ \varphi')$ are not C^r compatible unless $s = t$. This is because when $x \neq 0$,

$$\begin{aligned}((F_s \circ \varphi') \circ (F_t \circ \varphi')^{-1})(x) &= (F_s \circ \varphi' \circ (\varphi')^{-1} \circ F_t^{-1})(x) \\ &= (F_s \circ F_t^{-1})(x) \\ &= F_s(\|x\|_2^{(1-t)/t} x) \\ &= \|(\|x\|_2^{(1-t)/t})x\|_2^{s-1} \cdot \|x\|_2^{(1-t)/t} x \\ &= \|x\|_2^{\frac{(1-t)s}{t}} \|x\|_2^{s-1} x = \|x\|_2^{\frac{s}{t}-1} x = F_{s/t}(x).\end{aligned}$$

Also, you can manually check that the transition map also equals $F_{s/t}(0)$ at $x = 0$. And since $F_{s/t}$ is a diffeomorphism of any class iff $s/t = 1$, we know that the two charts are C^r compatible if and only if $s = t$. ■

In a sense this exercise proves how important it is to keep in mind that we consider a manifold to be smooth with respect to a specific structure. That said, if we're not working with multiple different structures, then it's annoying to explicitely mention the structure over and over. So, we take the approach of calling a chart a smooth chart if it's in our structure.

To finish off today, I want to briefly address the boundary of a manifold.

Like before if M is a topological manifold, we say a point $p \in M$ is an interior point of M if there exists a chart (U, φ) covering p such that $\varphi(U)$ is open in $\mathbb{R}^n$. Meanwhile, if no such chart exists, we say p is a boundary point. Also, we denote the interior of M: $\operatorname{Int} M$, to be to be collection of interior points and the boundary of M: $\partial M \coloneqq M - \operatorname{Int} M$.

It is easy to see that $\operatorname{Int} M$ is an open subset of M and a manifold by itself without a boundary. Based on that we can also easily see that ∂M is a closed subset of M.

Oh, I also forgot to mention: if $U \subseteq M$ is open and M is a C^r manifold, then we can view $U \subseteq M$ as a C^r submanifold of M. Specifically, it's clear that U is a second countable Hausdorff space in the relative topology. Also, given any smooth chart (V, φ) on M, we define a smooth chart $(U \cap V, \varphi|_{U \cap V})$ on U. This gives us a C^r structure on U.

8/24/2025

Let $r, k \in \mathbb{Z}_{>0} \cup \{\infty\}$ and always assume $k \leq r$.

Suppose M is a C^r manifold and $f : M \to \mathbb{R}^m$ is a function. We say f is a C^k function if for every $p \in M$ there exists a smooth chart (U, φ) such that $p \in U$ and $f \circ \varphi^{-1}$ is C^k on the open set $\varphi(U)$ of either H^n or $\mathbb{R}^n$. In this case we denote $f \in C^k(M, \mathbb{R}^m)$ (although when $m = 1$ we usually shorthand this as $f \in C^k(M)$).

Exercise 2.3: Let M be a C^r manifold and suppose $f : M \to \mathbb{R}^k$ is a C^k function where $k \leq r$. Show that $f \circ \psi^{-1} : \varphi(U) \to \mathbb{R}^m$ is C^k for every smooth chart (U, ψ) on M.

Proof:
Given any smooth chart (U, ψ), we can show that $f \circ \psi^{-1}$ is locally C^k as follows. Take any $x = \psi(p) \in \psi(U)$. Now we know there exists another smooth chart (V, φ) satisfying that $p \in V$ and that $f \circ \varphi^{-1}$ is C^k. Also, $\varphi \circ \psi^{-1}$ is a C^r map on $\psi(U \cap V)$. Thus, $f \circ \psi^{-1} = (f \circ \varphi^{-1}) \circ (\varphi \circ \psi^{-1})$ is a C^k map on the open neighborhood $\psi(U \cap V)$ of x in H^n or $\mathbb{R}^n$. ■

Side note: A recurring theme will be that you need M to be at to be at least C^k in order for C^k functions on M to be well-behaved. After all, if $r < k$ then the composite function above is no longer guarenteed to be C^k.

Corollary / Exercise 2.1: $C^k(M, \mathbb{R}^m)$ is a real vector space. If $m = 1$, then $C^k(M)$ is a a real commutative algebra.

We want to generalize the previous definition even more to cover maps from manifolds to other manifolds. To do this, it's worth noting that any open set U of either H^n or $\mathbb{R}^n$ can be thought of as a manifold. Furthermore, U is a C^r manifold for any r when equipped with the standard structure, i.e. the one determined by the atlas $\{(U, \operatorname{Id})\}$.

(Unless specified otherwise, always assume an open subset of $\mathbb{R}^n$ or H^n is equipped with the standard structure...)

Exercise 2.2: Let U be an open subset of H^n or $\mathbb{R}^n$. Then a function $f : U \to \mathbb{R}^m$ is C^k in the traditional real analysis definition iff it is C^k with respect to our new definition.

($\Longrightarrow$)
Suppose (V, φ) is any smooth chart on U. Then $\varphi^{-1} = \mathrm{Id} \circ \varphi^{-1}$ is a C^r function. So, $f \circ \varphi^{-1}$ is C^k.

($\Longleftarrow$)
It must be the case that $f \circ \mathrm{Id}^{-1} = f$ is C^k in the traditional real analysis sense since (U, Id) is a smooth chart. ■

Now note that when viewing $\mathbb{R}^m$ as being a C^r manifold, we can "symmetrize" our definition by noting that a function $f : M \to \mathbb{R}^m$ is C^k if and only if for all $p \in M$ there exists a smooth chart (U, φ) on M and another smooth chart (V, ψ) on $\mathbb{R}^m$ with $f(U) \subseteq V$ such that $\psi \circ f \circ \varphi^{-1}$ is C^k from $\varphi(U)$ into $\psi(V)$.

($\Longrightarrow$)
Since f is C^k, let (U, φ) be a smooth chart such that $f \circ \varphi^{-1}$ is C^k. Then let (V, ψ) be any smooth chart on $\mathbb{R}^m$ with $f(U) \subseteq V$. Note that such a chart must exist since we know $(\mathbb{R}^m, \mathrm{Id})$ works. Now $\psi \circ \mathrm{Id}^{-1} = \psi$ is a C^r map from V. So, $\psi \circ f \circ \varphi^{-1}$ is a C^k from $\varphi(U)$ into $\psi(V)$.

($\Longleftarrow$)
This direction is obvious when you just take (V, ψ) to be the chart $(\mathbb{R}^m, \mathrm{Id})$. ■

This points us to following generalization of differentiability on manifolds. Suppose M and N are both C^r manifolds and let $F : M \to N$ be any map. We say F is a C^k map if for every $p \in M$ there exists a smooth chart (U, φ) on M with $p \in U$ and another smooth chart (V, ψ) on N with $F(U) \subseteq V$ satisfying that the composite map $\psi \circ F \circ \varphi^{-1}$ is a C^k map from $\varphi(U)$ into $\psi(V)$.

Proposition 2.4: Every C^k map between two C^r manifolds M and N is continuous.

Proof:
Suppose $F : M \to N$ is C^k. Then given any $p \in M$, we can show that F is continuous on a neighborhood U of p. Specifically, let (U, φ) and (V, ψ) be smooth charts as in the prior definition. Then $\psi \circ F \circ \varphi^{-1}$ is continuous on the set $\varphi(U)$ on account of it being a differentiable function between two subsets of $\mathbb{R}^n$. Also, since both ψ and φ are homeomorphisms, we have that $F = \psi^{-1} \circ (\psi \circ F \circ \varphi^{-1}) \circ \varphi$ is a continuous map from $U \subseteq M$ into N.

Since each $p \in M$ has a neighborhood on which F is continuous, it follows that F is a continuous map from M to N. ■

Shit I just realized that I've never actually proven that continuity is local like that. So here's a quick lemma…

Lemma: Let $f : X \to Y$ be a map and suppose that every $x \in X$ has a neighborhood N_x such that $f|_{N_x}$ is continuous. Then f is continuous.

Proof:
For any $x \in X$, let N_x be a neighborhood satisfying that $f|_{N_x}$ is continuous. Then given any neighborhood V of $f(x)$ in Y, we know $U := f^{-1}(V) \cap N_x$ must be a neighborhood of x. satisfying that $f(U) \subseteq V$. So, f (not restricted to any subset) is continuous at x.

Since f is continuous at all $x \in X$, we know that f is continuous on X. ■

As a side note, when we were defining what it means for a map between manifolds, $F : M \to N$, to be differentiable, perhaps it felt overly restricting for us to force the chart (V, ψ) on N to satisfy that $F(U) \subseteq V$ in our definition. However, it turns out that without that requirement, it is no longer the case that F being differentiable implies that F is continuous.

Problem 2-1: Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$.

Now for every $x \in \mathbb{R}$ there are smooth coordinate charts (U, φ) containing x and (V, ψ) containing $f(x)$ such that $\psi \circ f \circ \varphi^{-1}$ is smooth as a map from $\varphi(U \cap f^{-1}(V))$ to $\psi(V)$. However f is clearly not continuous, nor smooth according to according to definition of smoothness of maps between two manifolds.

If $x \neq 0$, we can just pick $(U, \varphi) = (\mathbb{R} - \{0\}, \mathrm{Id})$ and $(V, \psi) = (\mathbb{R}, \mathrm{Id})$. Then it's clear that $f(x) \in V$ and $\psi \circ f \circ \varphi^{-1} = f$ is smooth as a map from $\varphi(U \cap f^{-1}(V)) = \mathbb{R} - \{0\}$ to $\mathbb{R}$.

Meanwhile, if $x = 0$, then pick $(U, \varphi) = (\mathbb{R}, \mathrm{Id})$ and $(V, \psi) = ((0, \infty), \mathrm{Id})$. Then it is still the case that $f(x) \in V$. Also, $\psi \circ f \circ \varphi^{-1} = f$ is just the constant function 1 on the set $\varphi(U \cap f^{-1}(V)) = [0, \infty)$. Thus since it can be extended to a differentiable function on an open set containing $[0, \infty)$, we can say that $\psi \circ f \circ \varphi^{-1}$ is also a smooth map from $\varphi(U \cap f^{-1}(V))$.

However, f is not differentiable or even continuous in the traditional analysis sense. Thus, f cannot be a smooth map from $\mathbb{R}$ as a manifold into $\mathbb{R}$ by exercise 2.2. Then in turn, we know from our prior efforts in generalizing differentiability that f is not smooth as a map into $\mathbb{R}$ as a manifold either. ■

Proposition 2.5: Suppose M and N are C^r manifolds, and $F : M \to N$ is a map. Then the following are equivalent.

(a) F is a C^k map.

(b) For every $p \in M$ there exist smooth charts (U, φ) containing p and (V, ψ) containing $F(p)$ such that $U \cap F^{-1}(V)$ is open in M and the map $\psi \circ F \circ \varphi^{-1}$ is C^k from $\varphi(U \cap F^{-1}(V))$ into $\psi(V)$.

(c) F is continuous and there exists atlases $\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}$ of M and $\{(V_\beta, \psi_\beta)\}_{\beta \in B}$ of N consisting of smooth charts such that for each α and β:

$$\psi_\beta \circ F \circ \varphi_\alpha^{-1} \text{ is a } C^k \text{ map from } \varphi_\alpha(U_\alpha \cap F^{-1}(V_\beta)) \text{ into } \psi_\beta(V_\beta).$$

$(a \implies c)$
From the last proposition we know that F is continuous. Also, suppose (U, φ) and (V, ψ) are *any* smooth charts on M and N respectively such that $U \cap F^{-1}(V) \neq \emptyset$. Then we claim that $\psi \circ F \circ \varphi^{-1}$ is a C^k map from $\varphi(U \cap F^{-1}(V))$ into $\psi(V)$.

Proof:
Let $x = \varphi(p)$ be in $\varphi(U \cap F^{-1}(V))$. Then since F is C^k, we know there are smooth charts (W_m, θ_m) in M and (W_n, θ_n) in N such that $p \in W_m$, $F(W_m) \subseteq W_n$, and $\theta_n \circ F \circ \theta_m^{-1}$ is a C^k map from $\theta_m(W_m)$ into $\theta_n(W_n)$. Also, we have that $\psi \circ \theta_n^{-1}$ is a C^r map from $\theta_n(V \cap W_n)$ to $\psi(V \cap W_n)$. And similarly, we have that $\theta_m \circ \varphi^{-1}$ is a C^r map from $\varphi(U \cap W_m)$ to $\theta_m(U \cap W_m)$.

Now we get to composing the functions established above (and I'll do this slowly since my head is already spinning from all the symbols written above).

- $(\psi \circ \theta_n^{-1}) \circ (\theta_n \circ F \circ \theta_m^{-1})$ is a C^k map from $\theta_m(W_m \cap F^{-1}(V \cap W_n))$ into $\psi(V)$.
- Because $F^{-1}(W_n) \supseteq W_m$, we have that $W_m \cap F^{-1}(V \cap W_n) = W_m \cap F^{-1}(V)$.
- Thus $(\psi \circ \theta_n^{-1}) \circ (\theta_n \circ F \circ \theta_m^{-1}) \circ (\theta_m \circ \varphi^{-1})$ is a C^k map from $\varphi(U \cap W_m \cap F^{-1}(V))$ into $\psi(V)$.
- Also $\psi \circ F \circ \varphi^{-1} = (\psi \circ \theta_n^{-1}) \circ (\theta_n \circ F \circ \theta_m^{-1}) \circ (\theta_m \circ \varphi^{-1})$ and we know $x \in \varphi(U \cap W_m \cap F^{-1}(V))$.
- Since F is continuous, we know that $F^{-1}(V)$ is open. It follows that $\varphi(U \cap W_m \cap F^{-1}(V))$ is an open neighborhood of x in either H^n or $\mathbb{R}^n$.
- This shows that any $x \in \varphi(U \cap F^{-1}(V))$ has an open neighborhood in either H^n or $\mathbb{R}^n$ for which $\psi \circ F \circ \varphi^{-1}$ is a C^k map when restricted to that neighborhood. It follows that $\psi \circ F \circ \varphi^{-1}$ is C^k on $\varphi(U \cap F^{-1}(V))$.

Side note: I essentially just proved an analog of exercise 2.3 a few pages ago. So that I can cite it later, I'll write it out as follows...

<u>Proposition</u> If M and N are C^r manifolds and $F : M \to N$ is a C^k map, then $\psi \circ F \circ \varphi^{-1}$ is a C^k map on $\varphi(U \cap F^{-1}(V))$ for all smooth charts (U, φ) on M and (V, ψ) on N.

Based on the prior reasoning, it suffices to choose any covering of M and N of smooth charts and we are done showing (c).

$(c \implies b)$
Let $p \in M$ and let $(U_\alpha, \varphi_\alpha)$ be a smooth chart on M covering p. Next let (V_β, ψ_β) be a smooth chart on N covering $F(p)$. Since F is continuous, we know that $U_\alpha \cap F^{-1}(V_\beta)$ is open. Also, we know by assumption that $\psi_\beta \circ F \circ \varphi_\alpha^{-1}$ is C^k from $\varphi_\alpha(U_\alpha \cap F^{-1}(V_\beta))$ into $\psi_\beta(V_\beta)$. This proves (b).

$(b \implies a)$
Given any $p \in M$ let (U, φ) and (V, ψ) be as in the hypothesis of (b). Then since $U \cap F^{-1}(V)$ is open and contains p, we have that:

$(U', \varphi') := (U \cap F^{-1}(V), \varphi|_{U \cap F^{-1}(V)})$ is another chart on M containing p.

Importantly, $\psi \circ F \circ (\varphi')^{-1}$ will still be a C^k map from $\varphi'(U')$ into $\psi(V)$ since $\psi \circ F \circ \varphi^{-1}$ also is that. But, we also have $F(U') \subseteq V$. This proves (a). ■

Proposition 2.6: Let M and N be C^r manifolds, and let $F : M \to N$ be a map.

- If every point $p \in M$ has an open neighborhood U such that the restriction $F|_U$ is C^k, then F is C^k globaly.
- If F is C^k globaly, then it's restriction to every open subset $U \subseteq M$ is C^k.

Proof:
Hopefully it is obvious that the latter bullet point is just a corollary of the prosition I noted on the last page. Meanwhile, the first bullet point is proved just by noting that any smooth chart in U is also a smooth chart in M. ■

Corollary 2.8: (Gluing lemma) Let M and N be C^r manifolds and let $\{U_\alpha\}_{\alpha\in A}$ be an open cover for M. Suppose that for each $\alpha \in A$ we are given a C^k map $F_\alpha : U_\alpha \to N$ and suppose that $F_\alpha|_{U_\alpha \cap U_\beta} = F_\beta|_{U_\alpha \cap U_\beta}$ for all α and β. Then there exists a unique C^k map $F : M \to N$ such that $F|_{U_\alpha} = F_\alpha$ for all α.

Proof:
Define $F = \bigcup_{\alpha\in A} F_\alpha$. This a well defined function on M since the U_α cover all of M and for any $p \in M$ there is only one $q \in N$ such that $(p, q) \in F$. Also, it is clear that F is the unique map satisying that $F|_{U_\alpha} = F_\alpha$ for all $\alpha \in A$. And since all the F_α are C^k, we know that F is locally C^k. So by the last proposition, we know that F is a C^k map globally. ■

Proposition 2.10: Let M, N, and P be C^r manifolds.

(a) Every constant map $c : M \to M$ is C^r.

Proof:
Suppose $c(p) = q$ for all $p \in M$. Then let (V, ψ) be a smooth chart covering q. If (U, φ) is any smooth chart on M, we know that $c(U) = \{q\} \subseteq V$ and that $\psi \circ c \circ \varphi^{-1} = \psi(q)$ is a constant function on $\varphi(U)$. Therefore $\psi \circ c \circ \varphi^{-1}$ is C^r and we've proven that c is a C^r map. ■

(b) The identity map on M is C^r.

Proof:
Let (U, φ) be any chart on M. Then $\mathrm{Id}_M(U) \subseteq U$ and $\varphi \circ \mathrm{Id}_M \circ \varphi^{-1} = \mathrm{Id}_{\mathbb{R}^n}$ is C^r. This proves that Id_M is C^r.■.

(c) If $U \subseteq M$ is an open submanifold, then the inclusion map $U \hookrightarrow M$ is C^r.

Proof:
Just apply proposition 2.6 to the identity map on M. ■

(d) If $F : M \to N$ and $G : N \to P$ are C^k, then so is $G \circ F : M \to P$.

Proof:
Let $p \in M$. Then by definition there are smooth charts (V, θ) on N and (W, ψ) on P such that $F(p) \in V$, $G(V) \subseteq W$, and $\psi \circ G \circ \theta^{-1}$ is a C^k map on $\theta(V)$. Also, since F is continuous, we know that $F^{-1}(V)$ is an open set in M. Therefore, we can find a smooth chart (U, φ) such that $p \in U \subseteq F^{-1}(V)$. In turn, $(G \circ F)(U) \subseteq W$. And since F is C^k, we know that $\theta \circ F \circ \varphi^{-1}$ is a C^k map from $\varphi(U)$. Hence:

$$\psi \circ (G \circ F) \circ \varphi^{-1} = (\psi \circ G \circ \theta^{-1}) \circ (\theta \circ F \circ \varphi^{-1}) \text{ is a } C^k \text{ map from } \varphi(U)$$

This proves that $G \circ F$ is a C^k map. ■

8/25/2025

Today I'm going to jump back to Guillemin's Differential Forms. My reasoning for this is that I want a working fomulation of Stokes theorem before the end of the Summer, and at the rate I'm going through Lee's book, I'm not going to get that formulation from Lee. I also never quite finished the chapter on tensors before. I will continue to return to Lee off and on though.

The pullback operation on $\Lambda^k(V^*)$:
Let V be an n-dimensional spaces over a field F with characteristic 0, and let W be an m-dimensional vector space over F. Also let $A : V \to W$ be a linear map. Recall that for any $T \in \mathcal{L}^k(W)$ we defined $A^\dagger T(v_1, \ldots, v_k) = T(Av_1, \ldots, Av_k)$.

Lemma 1.8.1: If $T \in \mathcal{I}^k(W)$, then $A^\dagger T \in \mathcal{I}^k(V)$.

Proof:
Since T can be expressed as a linear combination of redundant k-tensors and $A^\dagger$ is a linear map from $\mathcal{L}^k(W)$ to $\mathcal{L}^k(V)$, it suffices to assume T is itself a redundant k-tensor. So let $T = \ell_1 \otimes \cdots \otimes \ell_k$ where each $\ell_j \in W^*$ and $\ell_i = \ell_{i+1}$ for some i. Then by proposition 1.3.18 (on page 103) we have that $A^\dagger T = (A^\dagger \ell_1) \otimes \cdots \otimes (A^\dagger \ell_k)$. It follows that $A^\dagger T$ is a redundant k-tensor. ■.

Let π_W and π_V be the projections of $\mathcal{L}^k(W)$ and $\mathcal{L}^k(V)$ onto $\Lambda^k(W^*)$ and $\Lambda^k(V^*)$ respectively.

If $\omega \in \Lambda^k(W^*)$ and $T \in \mathcal{L}^k(W)$ satisfies that $\pi_W(T) = \omega$, then we define:

$$A^\dagger \omega := \pi_V(A^\dagger T).$$

To see that this is well defined, suppose $\omega = \pi(T) = \pi(T')$. Then $T = T' + S$ for some $S \in \mathcal{I}^k(V)$. So, $A^\dagger T = A^\dagger T' + A^\dagger S$. And since $A^\dagger S \in \mathcal{I}^k(V)$ by the last lemma, we have that $\pi(A^\dagger T) = \pi(A^\dagger T')$.

Proposition 1.8.4: The map $A^\dagger : \Lambda^k(W^*) \to \Lambda^k(V^*)$ sending ω to $A^\dagger\omega$ is linear. Moreover:

- if $\omega_i \in \Lambda^{k_i}(W^*)$ for $i = 1, 2$, then $A^\dagger(\omega_1 \wedge \omega_2) = (A^\dagger\omega_1) \wedge (A^\dagger\omega_2)$;
- if U is a vector space and $B : U \to V$ is a linear map, then for $\omega \in \Lambda^k(W^*)$, $B^\dagger(A^\dagger\omega) = (AB)^\dagger\omega$.

Proof:
Firstly, let $\omega, \omega' \in \Lambda^k(W^*)$. Then if $T \in \pi_W^{-1}(\{\omega\})$ and $T' \in \pi_W^{-1}(\{\omega\})$, we have for any $\lambda, \lambda' \in F$ that $\pi_W(\lambda T + \lambda' T') = \lambda\omega + \lambda'\omega'$. And now, showing that $A^\dagger$ as a map from $\Lambda^k(W^*)$ is linear is as simple as noting that $\pi_V \circ A^\dagger$ is linear (where we view $A^\dagger$ as a map from $\mathcal{L}^k(W)$).

Next, let ω_1, ω_2 be as in the proposition statement. Then suppose T_1 and T_2 both satisfy that $\pi_W(T_1) = \omega_1$ and $\pi_W(T_2) = \omega_2$ (ignore my abuse of notation). Then:

$$\begin{aligned} A^\dagger(\omega_1 \wedge \omega_2) &= \pi_V(A^\dagger(T_1 \otimes T_2)) \\ &= \pi_V((A^\dagger T_1) \otimes (A^\dagger T_2)) \\ &= \pi_V((A^\dagger T_1)) \wedge \pi_V((A^\dagger T_2)) = (A^\dagger\omega_1) \wedge (A^\dagger\omega_2). \end{aligned}$$

Finally, let $\omega \in \Lambda^k(W^*)$ and choose $T \in \mathcal{L}^k(W)$ such that $\pi_W(T) = \omega$. Then if you squint you can see that:

$$B^\dagger(A^\dagger\omega) = \pi_V(B^\dagger(A^\dagger T)) = \pi_V((AB)^\dagger T) = (AB)^\dagger\omega.$$

One application of the pullback operation is that it gives us a way of defining determinants completely independently of any chosen basis. Specifically, let V be an n-dimensional vector space over a field F with characteristic 0, and suppose $A : V \to V$ is a linear map.

(If you want to be pedantic, everything that follows should work so long as F has a characteristic that makes it so that $n! \neq 0$ and $-1 \neq 1$...)

Since $\dim \Lambda^n(V^*) = \binom{n}{n} = 1$ and $A^\dagger : \Lambda^n(V^*) \to \Lambda^n(V^*)$ is linear, it must be that the map $A^\dagger$ is just multiplication by a constant. We denote this constant $\det(A)$ and call it the determinant of A. In other words, we define $\det(A) \in F$ to be the constant such that $A^\dagger\omega = \det(A)\omega$ for all $\omega \in \Lambda^n(V^*)$.

Proposition 1.8.7: If A and B are linear mappings of V into V, then $\det(AB) = \det(A)\det(B)$.

Proof:
Given any $\omega \in \Lambda^n(V^*)$:

$$\det(AB)\omega = (AB)^\dagger\omega = B^\dagger(A^\dagger\omega) = \det(B)A^\dagger\omega = \det(B)\det(A)\omega$$

It follows that $\det(A)\det(B) = \det(AB)$. ■

Proposition 1.8.8: Write $\mathrm{Id}_V : V \to V$ for the identity map. Then $\det(\mathrm{Id}_V) = 1$.

Proof:
Note that if $\omega \in \Lambda^k(V^*)$ for any k and $T \in \mathcal{L}^k(V)$ satisfies that $\pi_V(T) = \omega$, then $\mathrm{Id}^\dagger\omega = \pi_V(\mathrm{Id}^\dagger T) = \pi_V(T) = \omega$. This shows that for any k, $\mathrm{Id}^\dagger$ is the identity map on $\Lambda^k(V^*)$. Hence $\det(\mathrm{Id}) = 1$. ■

Corollary: If $A : V \to V$ is a surjective linear map, then $\det(A) \neq 0$.

Proof:
If A is surjective, then we know by the rank-nullity theorem that A has nullity 0. So, A is bijective and has an inverse A^{-1}. Then by our last two propositions, we know that $\det(A)\det(A^{-1}) = \det(\mathrm{Id}_V) = 1$. Thus, it cannot be that $\det(A) = 0$. ■

Proposition 1.8.9: If $A : V \to V$ is not surjective, then $\det(A) = 0$.

Proof:
Let W be the image of A. If A is not surjective, we know that $\dim W < n$ and thus $\Lambda^n(W^*) = \{0\}$. So, let $A = i_W B$ where i_W is the inclusion map $W \hookrightarrow V$ and let B be the map A with it's codomain restricted to W. Then by proposition 1.8.4 we have that $A^\dagger\omega = B^\dagger(i_W^\dagger\omega)$. But now note $i_W^\dagger\omega \in \Lambda^n(W^*) = \{0\}$. This means that $i_W^\dagger\omega = 0$ and we trivially have that $B^\dagger(i_W^\dagger) = 0$. It follows that $\det(A) = 0$. ■

We still need to show that this definition of the determinant agrees with the usual one. To do this we can first prove something slightly more general.

Suppose $A : V \to W$ is a linear map, and let $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ be bases of V and W respectively. Then let $e_1^*, \ldots, e_n^*$ and $f_1^*, \ldots, f_n^*$ be the corresponding dual bases. If $(a_{i,j})$ is the $n \times n$ matrix representing A with respect to our bases (i.e. $Ae_j = \sum_{i=1}^n a_{i,j} f_i$ for all j), then:

$A^\dagger f_j^* = \sum_{i=1}^n a_{j,i} e_i^*$ for all j (see claim 1.2.15 on page 102…).

In turn:

$$\begin{aligned} A^\dagger(f_1^* \wedge \cdots \wedge f_n^*) &= (A^\dagger f_1^*) \wedge \cdots \wedge (A^\dagger f_n^*) \\ &= (\sum_{i=1}^n a_{1,i} e_i^*) \wedge \cdots \wedge (\sum_{i=1}^n a_{n,i} e_i^*) = \sum_{1 \leq k_1, \ldots, k_n \leq n} a_{1,k_1} \cdots a_{n,k_n} (e_{k_1}^* \wedge \cdots \wedge e_{k_n}^*) \end{aligned}$$

Next, if the multi-index $I = (k_1, \ldots, k_n)$ is repeating, then $e_{k_1}^* \wedge \cdots \wedge e_{k_n}^* = 0$. (This is a consequence of the fact that the wedge product with respect to 1-tensors is anti-commutative). It follows that we can cancel out a bunch of terms in the sum and be left with:

$$A^\dagger(f_1^* \wedge \cdots \wedge f_n^*) = \sum_{\sigma \in S_n} a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} (e_{\sigma(1)}^* \wedge \cdots \wedge e_{\sigma(n)}^*)$$

But now note that:

$$\begin{aligned} (e_{\sigma(1)}^* \wedge \cdots \wedge e_{\sigma(n)}^*) &= \pi_V(e_{\sigma(1)}^* \otimes \cdots \otimes e_{\sigma(n)}^*) \\ &= \pi_V((e_1^* \otimes \cdots \otimes e_n^*)^\sigma) = \mathrm{sgn}(\sigma)\pi_V(e_1^* \otimes \cdots \otimes e_n^*) = \mathrm{sgn}(\sigma) e_1^* \wedge \cdots \wedge e_n^*. \end{aligned}$$

So, we conclude that $A^\dagger(f_1^* \wedge \cdots \wedge f_n^*) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} (e_1^* \wedge \cdots \wedge e_n^*)$.

Letting $W = V$ and $f_i = e_i$ for all i, we in turn have shown that:

$$\det(A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)}.$$

8/26/2025

In physics yesterday we started talking about statistical mechanics, and that inspired me to try focusing on real analysis again. So, I will be returning to Folland for a bit. I think I'll start off where I left off at chapter 7.

Let X be an LCH space and $\mathcal{B}_X$ be the collection of Borel sets on X. At where we left off (on page 62), we had showed that the space $M(X)$ of complex Radon measures on $(X, \mathcal{B}_X)$ is a normed complex vector space when equipped with the norm $\mu \mapsto \|\mu\| = |\mu|(X)$. Also, we had shown that the map $\mu \mapsto I_\mu$ where $I_\mu(f) = \int f \mathrm{d}\mu$ is an isometric isomorphism from $M(X)$ to $C_0(X)^*$.

Exercise 7.8: Suppose that μ is a Radon measure on X. If $\phi \in L^1(\mu)$ and $\phi \geq 0$, then $\nu = \phi \mathrm{d}\mu$ is a Radon measure.

Since $\phi \in L^1(\mu)$, we know that $\nu(E) = \int_E \phi \mathrm{d}\mu$ is finite for all $E \in \mathcal{B}_X$. Thus ν is a finite measure. This trivially satisfies the requirement that $\nu(K)$ is finite for all compact K.

Now let $\varepsilon > 0$ and note that by corollary 3.6 (see my math 240a notes from Fall quarter), there exists $\delta > 0$ such that $\mu(A) < \delta$ implies that $|\nu(A)| = \nu(A) < \varepsilon$ for all $A \in \mathcal{B}_X$. This easily let's us show all the desired regularity properties of ν.

If $E \in \mathcal{B}_X$ then let $U \supseteq E$ be an open set such that $\mu(U - E) < \delta$. Then we know that $\nu(U - E) < \varepsilon$. Taking $\varepsilon \to 0$ shows that ν is outer regular on E.

If $U \subseteq X$ is open, then let $K \subseteq U$ be a compact set such that $\mu(U - K) < \delta$. Then we know that $\nu(U - K) < \varepsilon$. Taking $\varepsilon \to 0$ shows that ν is inner regular on U.

Corollary: If μ is a fixed positive Radon measure and $f \in L^1(\mu)$, then $\nu = f \mathrm{d}\mu$ is a complex Radon measure.

If we write $f = f_1 - f_2 + i(f_3 - f_4)$ and set $\nu_j = f_j \mathrm{d}\mu$ for all j, then it's clear from the last exercise that all the ν_j are finite (and thus complex) Radon measures. Also, $\nu = \nu_1 - \nu_2 + i(\nu_3 - \nu_4)$. So ν is also a complex radon measure.

In one sense the following is completely unnecessary. I will never need the full generality of what I'm about to prove (probably). On the other hand, Folland doesn't prove this and instead points his readers to the bibliography. So, for the first time in my life I went and procured a book from the bibliography of a math textbook. Also, this was especially a pain since when I finally got the textbook, it wouldn't convert to a pdf for some reason and my main E-book reader couldn't read it. So I eventually downloaded a DJVu reader that looks like it's from the fucking 2000s in order to finally read the book.

Considering the fact that there aren't just pdfs of this book floating around willy nilly on the internet, I should probably give a citation instead of just vaguely describing it. The book I will be briefly following along with is Hewitt and Stromberg's <u>Real and Abstract Analysis</u>. I've also made a bibiliography section at the end of the pdf where this book will have the honor of being the first citation.

A measure μ on $(X, \mathcal{M})$ is called decomposable if there is a family $\mathcal{F} \subseteq \mathcal{M}$ with the following properties:

(i) $\mu(F) < \infty$ for all $F \in \mathcal{F}$;

(ii) The members of $\mathcal{F}$ are disjoint and their union is X;

(iii) If $\mu(E) < \infty$, then $\mu(E) = \sum_{F \in \mathcal{F}} \mu(E \cap F)$;

(iv) If $E \subseteq X$ and $E \cap F \in \mathcal{M}$ for all $F \in \mathcal{F}$, then $E \in \mathcal{M}$.

Also, we call $\mathcal{F}$ a decomposition of $(X, \mathcal{M}, \mu)$.

Note that if μ is σ-finite, then we clearly have that μ is decomposable on $(X, \mathcal{M})$.

Lemma 19.26: Let $(X, \mathcal{M})$ be a measurable space and let μ and ν be arbitrary measures on $(X, \mathcal{M})$ such that $\mu(X) < \infty$ and $\nu \ll \mu$. Then there exists a set $E \in \mathcal{M}$ such that:

(i.) For all $A \in \mathcal{M}$ with $A \subseteq E$, either $\nu(A) = 0$ or $\nu(A) = \infty$. Also, if $\nu(A) = 0$, then so does $\mu(A) = 0$.

(ii.) ν is σ-finite on E^{C}.

Proof:

Consider the family:

$$\mathcal{B} \coloneqq \{B \in \mathcal{M} : \forall C \in \mathcal{A},\ C \subseteq B \Longrightarrow \nu(C) = 0 \text{ or } \nu(C) = \infty\}.$$

Importantly, we know that $\mathcal{B} \neq \emptyset$ since $\emptyset \in \mathcal{B}$, and also that $\mu(B) \leq \mu(X) < \infty$ for all $B \in \mathcal{B}$. So, it is well defined to set $\alpha \coloneqq \sup_{B \in \mathcal{B}} \mu(B)$. Next note that if $B, B' \in \mathcal{B}$, then $B \cup B' \in \mathcal{B}$.

Suppose $C \subseteq B \cup B'$ is measurable. Then $\nu(C) = \nu(C \cap B) + \nu(C \cap (B' - B))$. And since $C \cap B$ and $C \cap (B' - B)$ are both measurable subsets of sets in $\mathcal{B}$, we know that both have ν-measure zero or infinity. It follows that $\nu(C) \in \{0, \infty\}$.

We can thus construct a nondecreasing sequence of sets $(B_n)_{n \in \mathbb{N}}$ in $\mathcal{B}$ such that $\lim_{n \to \infty} \mu(B_n) = \alpha$. Setting $D = \bigcup_{n \in \mathbb{N}} B_n$, we have that $\mu(D) = \alpha$ and we can also show that $D \in \mathcal{B}$ using near identical reasoning as in the prior pink text.

Next we show that ν is semifinite on D^{C}. Suppose $F \in \mathcal{M}$ with $F \subseteq D$ and $\nu(F) = \infty$. Then for the sake of contradiction assume that $\nu(G)$ equals 0 or ∞ for every measurable subset $G \subseteq F$. It would follow that $F \cup D \in \mathcal{B}$. But then since $\nu(F) > 0$ and $\nu \ll \mu$, we'd know that $\mu(F) > 0$. Hence, $F \cup D$ would be a set in $\mathcal{B}$ with $\mu(F \cup D) > \alpha$ (and that inequality is strict since $\alpha < \infty$). Yet that contradicts how we defined α.

We furthermore show that ν is σ-finite on D^{C}. Let:

$$\mathcal{F} \coloneqq \{F \in \mathcal{M} : F \subseteq D^{\mathsf{C}} \text{ and } \nu \text{ is } \sigma\text{-finite on } F\}.$$

Like before, there exists a nondecreasing sequence $(F_n)_{n \in \mathbb{N}}$ in $\mathcal{F}$ such that $\lim_{n \to \infty} \mu(F_n) = \sup_{F \in \mathcal{F}} \mu(F) \eqqcolon \beta$. And if we set $F = \bigcup_{n \in \mathbb{N}} F_n$, then we clearly have that $F \in \mathcal{F}$ and $\mu(F) = \beta$. Our claim is that $\nu(F^{\mathsf{C}} \cap D^{\mathsf{C}}) = 0$.

Suppose not. Since ν is semifinite on D^{C}, we would have that there exists a measurable set $H \subseteq F^{\mathsf{C}} \cap D^{\mathsf{C}}$ with $0 < \nu(H) < \infty$. But since $\nu \ll \mu$, we'd have that $\mu(H) > 0$. It would then follow that $F \cup H \in \mathcal{F}$ and satisfies that $\mu(F \cup H) > \beta$. But this contradicts how we defined β.

It easily follows that ν is σ-finite on $F \cup (F^{\mathsf{C}} \cap D^{\mathsf{C}}) = D^{\mathsf{C}}$.

To finish off, let $\mathcal{G} := \{B \in \mathcal{M} : B \subseteq D$ and $\nu(B) = 0\}$. Now $\mathcal{G}$ is nonempty since $\emptyset \in \mathcal{G}$. So, it is well defined to set $\gamma := \sup_{B \in \mathcal{G}} \mu(B)$. Also, like before we have that if $B, B' \in \mathcal{G}$ then $B \cup B' \in \mathcal{G}$. So, we can once again take the union of a nondecreasing sequence of sets to get a set $G \subseteq D$ in $\mathcal{G}$ with $\mu(G) = \gamma$. And finally, set $E := D \cap G^{\mathsf{C}}$.

(i.) Since $E \subseteq D$, we know that any measurable $A \subseteq E$ satisfies that $\nu(A) \in \{0, \infty\}$. Also, if $\nu(A) = 0$ then we must have that $\mu(A) = 0$ since otherwise $G \cup A \in \mathcal{G}$ and $\mu(G \cup A) > \gamma$, which is a contradiction.

(ii.) $E^{\mathsf{C}} = D^{\mathsf{C}} \cup G$. And since ν is σ-finite on D^{C} and $\nu(G) = 0$, we have that ν is σ-finite on E^{C}. ■

Theorem 19.27: An Extension of the Lebesgue-Radon-Nikodym Theorem:

Let $(X, \mathcal{M}, \mu)$ be decomposable via the decomposition $\mathcal{F}$, and let ν be an arbitrary signed measure such that $\nu \ll \mu$. Then there exists an extended real $\mathcal{M}$-measurable function $f : X \to \overline{\mathbb{R}}$ such that $\nu(A) = \int_A f \mathrm{d}\mu$ for all $A \in \mathcal{M}$ for which μ is σ-finite. Also, we can take f to be finite on any F on which ν is σ-finite, and if ν is a positive measure, we can take f to be nonnegative. Furthermore, if g is any extended real $\mathcal{M}$-measurable function such that $\nu(A) = \int_A g \mathrm{d}\mu$ when $\mu(A) < \infty$, then we know $f\chi_E = g\chi_E$ μ-a.e. for all $E \in \mathcal{M}$ on which μ is σ-finite.

Proof:

We shall first consider the simpler case where ν is a positive measure.

Now restricting μ and ν to subspaces of $(X, \mathcal{M})$ won't change that $\nu \ll \mu$. Consequently, by restricting μ and ν to any subspace $F \in \mathcal{F}$ and applying our prior lemma, we can conclude that there are sets $D_F, E_F \in \mathcal{M}$ such that $D_F \cap E_F = \emptyset$; $D_F \cup E_F = F$; ν is σ-finite on D_F; and all measurable $A \subseteq E_F$ satisfy that $\nu(A) = 0 \Longrightarrow \mu(A) = 0$ and $\nu(A) \in \{0, \infty\}$.

Note that if ν is σ-finite on all of F, we can just take $D_F = F$ and $E_F = \emptyset$.

And going one step further, if we restrict μ and ν to the subspace D_F of $(X, \mathcal{M})$, then we know by the typical Lebesgue-Radon-Nikodym theorem that there is a finite measurable nonnegative function $f_0^{(F)} : X \to [0, \infty)$ such that $\nu(A) = \int_A f_0^{(F)} \mathrm{d}\mu$ for all $A \in \mathcal{M}$ with $A \subseteq D_F$.

Now define f by pasting together the functions:

$$f|_F(x) := \begin{cases} f_0^{(F)}(x) & \text{if } x \in D_F \\ \infty & \text{if } x \in E_F \end{cases}$$

- To show that f is measurable, let $(a, \infty]$ be an open ray in $\overline{\mathbb{R}}$. Then:

$$f^{-1}((a, \infty]) \cap F = f|_F^{-1}((a, \infty]) = (f_0^{(F)})^{-1}((a, \infty]) \cup E_F.$$

  And since $f_0^{(F)}$ is measurable on the subspace D_F of $(X, \mathcal{M})$, we've thus proven that $f^{-1}((a, \infty]) \cap F \in \mathcal{M}$ for all F. It follows from the definition of a decomposition that $f^{-1}((a, \infty]) \in \mathcal{M}$. And since the open rays form a basis for $\mathcal{B}_{\overline{\mathbb{R}}}$, we have proven that f is measurable.

- To show that $\nu(A) = \int_A f\mathrm{d}\mu$ when μ is σ-finite on A, first suppose that $\mu(A) < \infty$. Then we know from axiom (iii) of the definition of a decomoposition that there exists a countable subset $\mathcal{F}_0 \subseteq \mathcal{F}$ such that $\mu(A) = \sum_{F\in\mathcal{F}_0} \mu(A \cap F)$. One consequence of this is that $\int_A f\mathrm{d}\mu = \sum_{F\in\mathcal{F}_0} \int_{A\cap F} f\mathrm{d}\mu$.

  Another consequence is that because the $F \in \mathcal{F}$ partition X, we know $\mu(A \cap (\bigcup_{F\in\mathcal{F}_0^\mathsf{c}} F)) = 0$. Then in turn, since $\nu \ll \mu$, we also know that $\nu(A \cap (\bigcup_{F\in\mathcal{F}_0^\mathsf{c}} F)) = 0$. So, we must have that $\nu(A) = \sum_{F\in\mathcal{F}_0} \nu(A \cap F)$. And now we claim for each F that $\nu(A \cap F) = \int_{A\cap F} f\mathrm{d}\mu$.

  Clearly $\nu(A \cap F) = \nu(A \cap D_F) + \nu(A \cap E_F) = \int_{A\cap D_F} f\mathrm{d}\mu + \nu(A \cap E_F)$.

  Also, because $A \cap E_f \subseteq E_F$, we know that $\nu(A \cap E_F) \in \{0, \infty\}$ with $\mu(A \cap E_F) = 0$ if and only if $\nu(A \cap E_f) = 0$. So if $\mu(A \cap E_F) = 0$, then $\int_{A\cap E_F} f\mathrm{d}\mu = 0 = \nu(A \cap E_F)$. Meanwhile, if $\mu(A \cap E_F) > 0$, then $\int_{A\cap E_F} f\mathrm{d}\mu = \int_{A\cap E_F} (\infty)\mathrm{d}\mu = \infty = \nu(A \cap E_F)$. Either way, we have that:

  $$\nu(A \cap F) = \int_{A\cap D_F} f\mathrm{d}\mu + \int_{A\cap E_F} f\mathrm{d}\mu = \int_A f\mathrm{d}\mu$$

  So, we've shown that $\nu(A) = \sum_{F\in\mathcal{F}_0} \int_{A\cap F} f\mathrm{d}\mu = \int_A f\mathrm{d}\mu$ when $\mu(A) < \infty$. The case where $\mu(A) = \infty$ and μ is σ-finite on A then easily follows.

- Finally, suppose g is as in the theorem statement. Then for every $F \in \mathcal{F}$ and every measurable set $A \subseteq D_F$, we have that $\nu(A) = \int_A f\mathrm{d}\mu = \int_A g\mathrm{d}\mu$. The only way this is possible is if $f = g$ μ-a.e. on D_F.

  Meanwhile, we know for any $F \in \mathcal{F}$ that $A := E_F \cap g^{-1}([0, \infty)) \in \mathcal{M}$ because g is measurable. Now suppose $\mu(A) > 0$. Then letting $A_n := \{x \in A : g(x) < n\}$ for each n, we have that the A_n form an increasing sequence of sets whose union is A. So, we know that $\nu(A) = \lim_{n\to\infty} \nu(A_n)$. Also, since $\nu(A) > 0$ we know that $\nu(A_n) > 0$ for some n. Then in turn $\nu(A_n) = \infty$ since $A_n \subseteq E_F$. But this contradicts the fact that $\nu(A_n) = \int g\mathrm{d}\mu \leq n\mu(A_n) < \infty$. So, we conclude that $\mu(A) = 0$.

  As a result, we've shown for any $F \in \mathcal{F}$ that $f = g$ μ-a.e. on $D_F \cup E_F = F$.

  The fact that $g = f$ μ-a.e. on any measurable set such that $\mu(E) < \infty$ is then a simple consequence of the fact that there is a countable subset $\mathcal{F}_0 \subseteq \mathcal{F}$ as well a μ-null set $N \subseteq X$ such that $E = N \cup (\bigcup_{F\in\mathcal{F}_0} (F \cap E))$. And if μ is σ-finite on E, then we can show that $g = f$ μ-a.e. on E by considering E as a countable union of sets on which we've already showed $g = f$ μ-a.e.

Now we return to the case where ν is signed. Let ν^+ and ν^- be the positive and negative variations of ν, and also let P and N be measurable subsets of X such that $\nu^+(N) = 0$ and $\nu^-(P) = 0$. Then by our prior reasoning we know there exists measurable functions f^+ and f^- such that $\nu^+(A) = \int_A f^+\mathrm{d}\mu$ and $\nu^-(A) = \int_A f^-\mathrm{d}\mu$ for all A on which μ is σ-finite. Also, since either ν^+ or ν^- is finite, we know either f^+ or f^- never outputs ∞. So by setting $f = f^+ - f^-$, we get a measurable function such that $\nu(A) = \int_A f\mathrm{d}\mu$ for all A on which μ is σ-finite.

Now suppose g is as in the theorem statement, and let g^+ and g^- be it's positive and negative parts. Given any $F \in \mathfrak{F}$ we must have that $\int_A g\mathrm{d}\mu = \nu(A) = \nu^+(A) \geq 0$ for all measurable $A \subseteq F \cap P$. It follows that $g \geq 0$ μ-a.e. on $F \cap P$, and in turn:

$\int_A g^+\mathrm{d}\mu = \nu^+(A)$ when $A \subseteq F \cap P$.

By similar reasoning, we can show that $g \leq 0$ μ-a.e. on $F \cap N$. This is important because it shows that $g^+ = 0$ μ-a.e. on $F \cap N$. So $\int_A g^+\mathrm{d}\mu = 0 = \nu^+(A)$ for all $A \subseteq F \cap N$. And hence, we can conclude that $\int_A g^+\mathrm{d}\mu = \nu^+(A)$ for all $A \subseteq F$. It easily follows that $\int_A g^+\mathrm{d}\mu = \nu^+(A)$ for all A satisfying that $\mu(A) < \infty$. But then by the prior reasoning we did, we know that $g^+ = f^+$ μ-a.e. on any set E which μ is σ-finite on.

Analogous reasoning can be used to show that $g^- = f^-$ μ-a.e. on any set E which μ is σ-finite on. ■

Corollary: Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space and suppose ν is any arbitrary signed measure such that $\nu \ll \mu$. Then there exists a measurable function $f : X \to \overline{\mathbb{R}}$ such that $\nu = f\mathrm{d}\mu$. Also, if $g : X \to \overline{\mathbb{R}}$ is another function satisfying that $\nu = g\mathrm{d}\mu$, then $f = g$ μ-a.e.

We can of course also write a version of the prior theorem and corollary for when ν is a complex measure. Specifically, just apply the prior theorem to the real and imaginary parts of ν separately. That said, the corollary stops being interesting if ν is complex.

Now, it's unfortunate that this extension of the Lebesgue-Radon-Nikodym theorem loses the ability to decompose ν into a continuous part and a mutually singular part. That said, an extremely convenient fact which makes the last theorem feel more worthwhile is that it turns out that every positive Radon measure is decomposable (with an asterisk attached).

To start off, we need an exercise from Folland.

Note that this exercise references a bunch of exercises which I already did in my Math 240a notes. Look there if you want to see stuff like the definition of the saturation of a measure. I even just now went back and clean up a bunch of problems with my answer for exercise 1.16(e)!! (haha why was I so stupid I swear I want to put an evoker to my head...)

Exercise 1.22: Let $(X, \mathcal{M}, \mu)$ be a measure space and let μ^* be the outer measure induced by treating μ as a premeasure on $\mathcal{M}$. Then let $\mathcal{M}^*$ be the σ-algebra of μ^*-measurable sets and $\overline{\mu} := \mu^*|_{\mathcal{M}^*}$.

(a) If μ is σ-finite, then $(X, \mathcal{M}^*, \overline{\mu})$ is the the completion of $(X, \mathcal{M}, \mu)$.

If $E \in \mathcal{M}^*$, then since μ is σ-finite we know by exercise 1.18 that there exists a set A in $\mathcal{M}_{\sigma\delta} \subseteq \mathcal{M}$ such that $E \subseteq A$ and $\overline{\mu}(A - E) = 0$. Hence $E = A \cup N$ where $A \in \mathcal{M}$ and $\overline{\mu}(N) = 0$. And since $\overline{\mu}$ is the restriction of the outer measure induced by μ, we know there is a null set $N' \in \mathcal{M}$ with $N \subseteq N'$. This proves that $\mathcal{M}^*$ is a subset of the completion of $\mathcal{M}$ (which I will hereafter call $\mathcal{M}^\wedge$).

Meanwhile, by Carathéodory's theorem we know that $(X, \mathcal{M}^*, \overline{\mu})$ is a complete measure space. So, suppose $E = F \cup N$ where $F \in \mathcal{M}$ and $N \subseteq N'$ with $\mu(N') = 0$. Then since $N' \in \mathcal{M}^*$ and $\overline{\mu}(N') = \mu(N') = 0$, we know $N \in \mathcal{M}^*$. In turn this means that $E = F \cup N \in \mathcal{M}^*$ and we've showed that $\mathcal{M}^\wedge \subseteq \mathcal{M}^*$.

Combining the last two paragraphs, we have that $\mathcal{M}^* = \mathcal{M}^\wedge$. And since there is only one measure that extends μ to $\mathcal{M}^\wedge$, we automatically have that $\overline{\mu}$ is the completion of μ.

(b) In general, $\overline{\mu}$ is the saturation of the completion of μ.

For the sake of notation, I will write that $\mathcal{M}^\wedge$ and $\mu^\wedge$ are the completions of $\mathcal{M}$ and μ respectively; that $\widetilde{\mathcal{M}}^\wedge$ is the collection of $\mu^\wedge$-locally measurable sets; and that $\widetilde{\mu}^\wedge$ is the saturation of of $\mu^\wedge$.

Our first goal is to show that $\widetilde{\mathcal{M}}^\wedge = \mathcal{M}^*$. Equivalently, this means we need to show that a set $E \subseteq X$ is μ^*-measurable if and only if it is locally $\mu^\wedge$-measurable.

$(\Longrightarrow)$
Suppose $E \subseteq X$ is μ^*-measurable and let $A \subseteq X$ be any set in $\mathcal{M}^\wedge$ with $\mu^\wedge(A) < \infty$. In part (a) we were able to show without using any σ-finiteness that $\mathcal{M}^\wedge \subseteq \mathcal{M}^*$. Thus, A and $E \cap A$ are μ^*-measurable. Also note that by definition of the completion of a measure, we can pick a set $F \in \mathcal{M}$ with $A \subseteq F$ and $\mu^\wedge(A) = \mu(F)$.

Now we claim by a similar argument as in exercise 1.18 that there exists a set $B \in \mathcal{M}$ such that $E \cap A \subseteq B$ and $\mu^*(B - (A \cap E)) = 0$.

Using exercise 1.18(a), for each $j \in \mathbb{N}$ pick $B_j \in \mathcal{M}$ such that $(A \cap E) \subseteq B_j$ and $\mu^*(B_j) \leq \mu^*(E \cap A) + \frac{1}{j}$. Then since $E \cap A$ is μ^*-measurable and $B \cap (E \cap A) = E \cap A$, we have that:

$$\mu^*(E \cap A) + \mu^*(B_j - (E \cap A)) = \mu^*(B_j) \leq \mu^*(E \cap A) + \tfrac{1}{j}$$

Next note that $\mu^*(E \cap A) \leq \mu^*(A) \leq \mu^*(F) = \mu(F) = \mu^\wedge(A) < \infty$. Thus, we can subtract $\mu^*(E \cap A)$ from both sides of our inequality above to get that:

$$\mu^*(B_j - (E \cap A)) < \tfrac{1}{j}$$

And now, if we define $B = \bigcap_{j \in \mathbb{N}} B_j$, it's clear that $B \in \mathcal{M}$ and that $\mu^*(B - (E \cap A)) \leq \mu^*(B_j - (E \cap A)) < \frac{1}{j}$ for all $j \in \mathbb{N}$. So, $\mu^*(B - (E \cap A)) = 0$.

Now since $\mu^*(B - (A \cap E)) = 0$, we know that there exists a set $C \in \mathcal{M}$ with $\mu(C) = 0$ and $B - (A \cap E) \subseteq C$.

Specifically, for each $n \in \mathbb{N}$ there exists a countable covering $\{C_j^{(n)}\}_{j \in \mathbb{N}}$ of $B - (A \cap E)$ such that $\mu(\bigcup_{j \in \mathbb{N}} C_j^{(n)}) \leq \sum_{j=0}^{\infty} \mu(C_j^{(n)}) < 1/n$ In turn, $C := \bigcap_{n \in \mathbb{N}} (\bigcup_{j \in \mathbb{N}} C_j^{(n)})$ is a set in $\mathcal{M}$ with $B - (A \cap E) \subseteq C$ and $\mu(C) = 0$.

Finally, we have that $N := B - (A \cap E) \in \mathcal{M}^\wedge$ because $N \subseteq C$ and $(X, \mathcal{M}^\wedge, \mu^\wedge)$ is complete. And since $B \in \mathcal{M} \subseteq \mathcal{M}^\wedge$, we have that $(A \cap E) = B - N \in \mathcal{M}^\wedge$. This proves that E is locally $\mu^\wedge$-measurable.

$(\Longleftarrow)$
Suppose $E \subseteq X$ is locally $\mu^\wedge$-measurable and then choose any $F \subseteq X$. If $\mu^*(F) = \infty$, then we trivially have that $\mu^*(F \cap E) + \mu^*(F - E) \leq \mu^*(F)$. So, it suffices to assume that $\mu^*(F) < \infty$. But then by exercise 1.18(a) there exists for each $j \in \mathbb{N}$ a set $A_j \in \mathcal{M}$ such that $F \subseteq A_j$ and $\mu(A_j) < \mu^*(F) + 1/j$. And, by taking the intersection of all the A_j we get a set $A \in \mathcal{M}$ such that $A \subseteq F$ and $\mu(A) = \mu^*(F) < \infty$.

Since E is locally $\mu^\wedge$-measurable, it follows that $A \cap E \in \mathcal{M}^\wedge$. Then since $\mathcal{M}^\wedge \subseteq \mathcal{M}^*$, we know that $\mu^*(F) = \mu^*(F \cap (A \cap E)) + \mu^*(F - (A \cap E))$. And this shows that E is μ^*-measurable since $F \cap (A \cap E) = F \cap E$ and $F - (A \cap E) = F - E$ on account of the fact that $F \subseteq A$.

With that, we've now shown that $\widetilde{\mathcal{M}}^\wedge = \mathcal{M}^*$. Also, since there is only one extension of μ to $\mathcal{M}^\wedge$ and both $\overline{\mu}$ and $\widetilde{\mu}^\wedge$ extend μ to $\mathcal{M}^\wedge$, we must have that $\overline{\mu}(E) = \widetilde{\mu}^\wedge(E)$ whenever $E \in \mathcal{M}^\wedge \subseteq \mathcal{M}^*$ Also, by definition of the saturation of a measure, we have that if $E \in \mathcal{M}^* - \mathcal{M}^\wedge$, then $\widetilde{\mu}^\wedge(E) = \infty$. So, all we need to do left is show that if E is locally $\mu^\wedge$-measurable but not in $\mathcal{M}^\wedge$, then $\overline{\mu}(E) = \mu^*(E) = \infty$.

Suppose E is μ^*-measurable and $\overline{\mu}(E) = \mu^*(E) < \infty$. When we were proving the backwards implication before, we showed that there is a set $A \in \mathcal{M}$ with $A \subseteq E$ and $\mu^*(E) = \mu(A) < \infty$. Next, when we were proving the forwards implication, we showed that there is a set $B \in \mathcal{M}$ with $A \cap E = E \subseteq B$ and $\mu^*(B - (A \cap E)) = \mu^*(B - E) = 0$. Then afterwords, we showed that there exists $C \in \mathcal{M}$ with $\mu(C) = 0$ and $B - E \subseteq C$. Finally, we have that $B - E \in \mathcal{M}^\wedge$ and in turn that $E = B - (B - E) \in \mathcal{M}^\wedge$. ■

As a side note, this shows that you can't indefinitely extend a measure space to larger and larger σ-algebras just by applying the theorem on page 24 of my latex math 240a notes over and over again. After all, the saturation of a complete measure is still complete (by exercise 1.16 in my latex math 240a notes). So, if you complete it again you just get back the same measure space (see page 178 for a proof of this...). Also, it's easy to see that saturating a measure does not add any new locally measurable sets since all the added measurables sets have infinite measure. So, taking a saturation twice back to back is the same as taking it once.

Now let X be an LCH space and let μ be a Radon measure on $(X, \mathcal{B}_X)$. Then set:

$$\mu^*(E) \coloneqq \inf\{\mu(U) : U \supseteq E \text{ with } U \text{ open}\} \text{ for all } E \in \mathcal{P}(X).$$

Recall from the proof of the Riesz Representation thorem in my math 240c notes that μ^* is a well-defined outer measure for which every Borel set is μ^*-measurable and $\mu^*|_{\mathcal{B}_X} = \mu$. Furthermore, if $\mathcal{M}$ is the collection of μ^*-measurable sets and $\overline{\mu} = \mu^*|_{\mathcal{M}}$, we have by the definition of μ^* that $\overline{\mu}$ is outer regular on all of $\mathcal{M}$. Also, $\overline{\mu}$ is fully determined by the linear functional $I(f) \coloneqq \int f d\mu$. (since I uniquely determines $\mu(U)$ for each open U).

Importantly, we can also say that $\mu^*(E) = \inf\{\mu(B) : B \supseteq E \text{ with } B \in \mathcal{B}_X\}$ for all $E \in \mathcal{P}(X)$. This is because (and I will be abusing notation to write this since otherwise it would be really cumbersome) if B represents any Borel set and U any open set, then:

$$\mu^*(E) \leq \mu^*(\bigcap_{B\supseteq E} B) \leq \inf_{B\supseteq E} \mu(B) \leq \inf_{U\supseteq E} \mu(U) = \mu^*(E).$$

Next, we claim that:

$$\inf\{\mu(B) : B \supseteq E \text{ with } B \in \mathcal{B}_X\} = \inf\{\sum_{n=1}^{\infty} \mu(B_n) : \text{ all } B_n \text{ are Borel and } E \subseteq \bigcup_{n\in\mathbb{N}} B_n\}$$

The fact that the right side is at most the left side is trivial. Meanwhile, to show the other inequality we can just note that if $(B_n)_{n\in\mathbb{N}}$ is a covering of E by Borel sets, then by taking differences we can get a sequence of disjoint Borel sets $(B_n')_{n\in\mathbb{N}}$ covering E such that $B_n' \subseteq B_n$ for all n; $B := \bigcup_{n\in\mathbb{N}} B_n$ is a Borel set containing E; and:

$$\mu(B) = \sum_{n\in\mathbb{N}} \mu(B_n') \leq \sum_{n\in\mathbb{N}} \mu(B_n)$$

As a result, μ^* is equal to the outer measure induced by treating μ as a premeasure on $(X, \mathcal{B}_X)$. Combining this with the previous exercise I did, we thus know that $(X, \mathcal{M}, \overline{\mu})$ is precisely the saturation of the completion of $(X, \mathcal{B}_X, \mu)$.

Now returning to Hewitt and Stromberg, here is one more lemma before I reset which variable names I have assigned to what.

Lemma 10.31: For any $A \subseteq X$, the following are equivalent:

(i) A is μ^*-measurable;

(ii) $\mu^*(U) \geq \mu^*(U \cap A) + \mu^*(U - A)$ for all open $U \subseteq X$ such that $\mu(U) < \infty$;

(iii) $A \cap U$ is μ^*-measurable for all open $U \subseteq X$ such that $\mu(U) < \infty$;

(iv) $A \cap K$ is μ^*-measurable for all compact $K \subseteq X$.

Proof:
It's trivial that (i) implies (iv).

Next suppose (iv) holds and let U be an open set such that $\mu(U) < \infty$. Then since μ is inner regular on all open sets, we know for each $n \in \mathbb{N}$ that there is a compact set $F_n \subseteq U$ such that $\mu(F_n) > \mu(U) - 1/n$. Letting $F = \bigcup_{n\in\mathbb{N}} F_n$ we have that $F \subseteq U$; F is μ^*-measurable (since all the F_n are); and $\mu(F) \geq \mu(F_n) > \mu(U) - 1/n$ for all n. It follows that $\mu(F) = \mu(U)$ and $\mu(U - F) = 0$.

One consequence of this is that:

$$\begin{aligned} A \cap U = A \cap (F \cup (U - F)) &= (A \cap F) \cup (A \cap (U - F)) \\ &= (\bigcup_{n\in\mathbb{N}} (A \cap F_n)) \cup (A \cap (U - F)). \end{aligned}$$

And since $A \cap (U - F) \subseteq U - F$ with $\mu(U - F) = 0$, we know by the completeness of $(X, \mathcal{M}, \overline{\mu})$ that $A \cap (U - F)$ is in $\mathcal{M}$. Also, since we have that all the $A \cap F_n$ are in $\mathcal{M}$ by (iv), we know that $A \cap U$ is a countable union of sets in $\mathcal{M}$. This proves (iii).

Now suppose (iii) and let $U \subseteq X$ be an open set such that $\mu(U) < \infty$. Then both U and $U \cap A$ are in $\mathcal{M}$ by (iii). And in turn we also have that $U - A = U - (U \cap A) \in \mathcal{M}$. Thus $\overline{\mu}(U) = \overline{\mu}(U \cap A) + \overline{\mu}(U - A)$ and we've shown (ii).

Finally suppose (ii) and let F be any subset of X. If $\mu^*(F) = \infty$, then we trivially have that $\mu^*(F) \geq \mu^*(A \cap F) + \mu^*(A - F)$. So assume $\mu^*(F) < \infty$. Then for any given $\varepsilon > 0$ we know there is an open set U such that $F \subseteq U$ and $\mu(U) < \mu^*(F) + \varepsilon$. In turn:

$$\mu^*(F) + \varepsilon > \mu(U) \geq \mu^*(U \cap A) + \mu^*(U - A) \geq \mu^*(F \cap A) + \mu^*(F - A)$$

Taking $\varepsilon \to 0$ finishes proving (i). ■

Oh, one more lemma I'll need is that $\overline{\mu}$ is inner regular on all of its σ-finite sets. (You can prove this identically to how we proved μ is inner regular on all of its σ-finite sets [see my math 240c notes]).

Theorem 19.30: Let X be an LCH space and let $(X, \mathcal{M}, \mu)$ be the saturation of the completion of a positive Borel Radon measure on X. Then there exists a family $\mathcal{F}_0$ of subsets of X with the following properties:

(i) the sets in $\mathcal{F}_0$ are compact and have finite measure greater than 0;

(ii) the sets in $\mathcal{F}_0$ are pairwise disjoint;

(iii) if $F \in \mathcal{F}_0$, U is open, and $U \cap F \neq \emptyset$, then $\mu(U \cap F) > 0$;

(iv) if $E \in \mathcal{M}$ and $\mu(E) < \infty$, then $\mu(E \cap F) > 0$ for only countably many $F \in \mathcal{F}_0$;

(v) the set $D := X - (\bigcup_{F \in \mathcal{F}} F)$ is measurable and locally null (which means that $\mu(D \cap A) = 0$ for all $A \in \mathcal{M}$ with $\mu(A) < \infty$ [see my homework from math 240b for more info about this...]);

(vi) if Y is a subset of X such that $Y \cap F \in \mathcal{M}$ for all $F \in \mathcal{F}_0$, then $Y \in \mathcal{M}$.

Proof:
Let $\mathcal{X}$ be the collection of all families of subsets in $\mathcal{M}$ which satisfy properties (i), (ii), and (iii). Then firstly note that $\emptyset \in \mathcal{X}$. So, $\mathcal{X}$ is not empty. Also, it easy to see that if $\mathcal{X}_0$ is a subset of $\mathcal{X}$ that is linearly ordered by the subset relation, then $\mathcal{X}_0$ has an upper bound in $\mathcal{X}$. Just set $\mathcal{F} = \bigcup_{\mathcal{F}'' \in \mathcal{X}_0} \mathcal{F}''$. Then it's obvious that $\mathcal{F}$ satisfies properties (i) and (iii) since every $F \in \mathcal{F}$ is in some $\mathcal{F}'' \in \mathcal{X}_0 \subseteq \mathcal{X}$. Also, if F_1 and F_2 are sets in $\mathcal{F}$, then we know there exists a collection $\mathcal{F}'' \in \mathcal{X}_0$ such that both $F_1, F_2 \in \mathcal{F}''$. And thus in turn we know that $F_1 \cap F_2 = \emptyset$, meaning $\mathcal{F}$ satisfies property (ii).

It follows by Zorn's lemma that $\mathcal{X}$ contains a maximal family $\mathcal{F}_0$. Our goal now is to show that $\mathcal{F}_0$ satisfies properties (iv), (v), and (vi).

Suppose $E \in \mathcal{M}$ and $\mu(E) < \infty$. Then for the sake of finding a contradiction, suppose $E \cap F \neq \emptyset$ for uncountably many $F \in \mathcal{F}_0$. Then by the outer regularity of μ, we know there exists an open set U such that $U \supseteq E$ and $\mu(U) < \mu(E) + 1 < \infty$. Then in turn, by property (ii) we have that $\mu(U) \geq \sum_{F \in \mathcal{F}_0} \mu(U \cap F)$ (because U is a superset of any finite union of the $F \cap U$). But now since $U \cap F \supseteq E \cap F \neq \emptyset$ for uncountably many $F \in \mathcal{F}_0$, we know by property (iii) that $\mu(U \cap F) > 0$ for uncountably many $F \in \mathcal{F}_0$. But that implies that $\sum_{F \in \mathcal{F}_0} \mu(U \cap F) = \infty$, which is a contradiction. Hence, we've proven property (iv).

Next let U be any open set with $\mu(U) < \infty$ and let $\mathcal{F}_1$ be the countable subfamily of $\mathcal{F}_0$ containing all the F such that $\mu(U \cap F) > 0$. Since all $F \in \mathcal{F}_1$ are in $\mathcal{M}$, we know that $\bigcup_{F \in \mathcal{F}_1} F \in \mathcal{M}$. Thus $\mu(U) = \mu(U \cap \bigcup_{F \in \mathcal{F}_1} F) + \mu(U - \bigcup_{F \in \mathcal{F}_1} F)$. But by (iii) we have that $\mathcal{F}_1 = \{F \in \mathcal{F}_0 : U \cap F \neq \emptyset\}$. Thus, it's clear that $U \cap \bigcup_{F \in \mathcal{F}_1} F = U \cap \bigcup_{F \in \mathcal{F}_0} F$ and $U - \bigcup_{F \in \mathcal{F}_1} F = U - \bigcup_{F \in \mathcal{F}_0} F$. And by the previous lemma, this proves that $\bigcup_{F \in \mathcal{F}_0} F$ and $D \coloneqq X - \bigcup_{F \in \mathcal{F}_0} F$ are in $\mathcal{M}$.

Now we still need to show that D is locally null. So suppose for the sake of contradiction that there exists $A \in \mathcal{M}$ with $0 < \mu(A \cap D) < \infty$. Then since μ is inner regular on $A \cap D$, there'd be a compact set $K \subseteq A \cap D$ such that $0 < \mu(K) < \mu(A \cap D) < \infty$. Also, if we consider the collection $\mathcal{U}$ of open sets $U \subseteq X$ such that $\mu(U \cap K) = 0$, then $K \cap \bigcup_{U \in \mathcal{U}} U$ is measurable on account of $\bigcup_{U \in \mathcal{U}} U$ being open. We claim $K \cap \bigcup_{U \in \mathcal{U}} U$ is a null set.

Otherwise, by inner regularity there would exist a compact set $C \subseteq K \cap \bigcup_{U \in \mathcal{U}} U$ with $\mu(C) > 0$. And since $\mathcal{U}$ is an open cover of C, we'd have that there is a finite subcover of sets $U \cap H$, all with measure zero, covering C. This is a contradiction.

Setting $H \coloneqq K - \bigcup_{U \in \mathcal{U}} U$, we'd know that H is compact on account of being a closed subset of K. Also, we'd know that $\mu(H) = \mu(K) > 0$. And it's clear that H would be disjoint from all the $F \in \mathcal{F}_0$. And finally, if V were any open set such that $V \cap H \neq \emptyset$ then we'd know that $V \notin \mathcal{U}$. But that would mean that $\mu(V \cap K) > 0$. And so:

$$\mu(V \cap H) = \mu(V \cap K) - \mu(V \cap (K - \textstyle\bigcup_{U \in \mathcal{U}} U)) = \mu(V \cap K) - 0 > 0$$

Hence, we've shown that $\mathcal{F}_0 \cup \{H\}$ is a collection in $\mathcal{X}$ that is strictly larger than $\mathcal{F}_0$. But that contradicts that $\mathcal{F}_0$ is maximal. Thus, we've proven property (v).

Finally, let Y be as in the theorem statement and consider any open set $U \subseteq X$ such that $\mu(U) < \infty$. Then:

$$U \cap Y = (U \cap Y \cap D) \cup (U \cap Y \cap \bigcup_{F \in \mathcal{F}_0} F) = (U \cap Y \cap D) \cup \bigcup_{F \in \mathcal{F}_0} (U \cap (Y \cap F))$$

But now we know from before that there is a countable subfamily of $\mathcal{F}_0$ containing all the F which intersect U. Also, since D is locally null, we know that $\mu(U \cap D) = 0$ and in turn $U \cap Y \cap D$ is measurable due to it being a subset of a null set. It follows that $U \cap Y$ is a countable union of measurable sets. So, $U \cap Y$ is measurable. By our prior lemma, we thus have that $Y \in \mathcal{M}$. This proves (vi). ■

Side note: You may note that since D is locally null and all compact sets have finite measure, we must have that any compact set $K \subseteq X$ contained in D has measure zero. By inner regularity, this in turn implies that any open set $U \subseteq X$ contained in D has measure zero.

Corollary 19.31: Let X be an LCH space and let $(X, \mathcal{M}, \mu)$ be the saturation of the completion of a positive Borel Radon measure on X. Then $(X, \mathcal{M}, \mu)$ is decomposable.

Proof:

Let $\mathcal{F}_0$ and D be as in the last theorem. Then set $\mathcal{F} \coloneqq \mathcal{F}_0 \cup \{\{x\} : x \in D\}$. We claim $\mathcal{F}$ is a decomposition.

(i) Since X is Hausdorff, we know that $\{x\}$ is compact for all $x \in D$. Hence, every set in $\mathfrak{F}$ is compact and thus has finite measure.

(ii) It's clear that all the elements of $\mathfrak{F}$ form a partition of X.

(iii) If $\mu(E) < \infty$ then since D is locally null, we have that $\mu(E \cap D) = 0$. Also, by how we chose $\mathcal{F}_0$ we know there are only countably many $F \in \mathcal{F}_0$ with $F \cap E \neq \emptyset$. So:

$$\mu(E) = \sum_{F \in \mathfrak{F}} \mu(E \cap F).$$

(iv) This is an immediate consequence of the sixth property that we proved about $\mathfrak{F}_0$. ■

Corollary 19.32: Another Extension of the Lebesgue-Radon-Nikodym Theorem:

Let X be an LCH space and let $(X, \mathcal{M}, \mu)$ be the saturation of the completion of a positive Borel Radon measure on X. Also let ν be any measure on $(X, \mathcal{M})$ such that $\nu \ll \mu$. Then all the conclusions of theorem 19.27 (on page 156 of my journal) hold.

Reflection:

It is here that Hewitt and Stromberg mostly stop being useful. So, I will give one last quote from them before expositing some more about this topic on my own (also geeze I just noticed that this quote has a typo. It's supposed to be "(19.32)"; not "(19.33)"...):

> [redacted] One may well ask if the generality obtained in (19.27) and (19.33) is worth the effort. Many mathematicians believe it is not. But we feel it our duty to show the reader the most general theorems that we can reasonably produce and that he might reasonably need. [redacted]

I'm not gonna lie, while going down research rabbitholes is really fun, I feel like Folland really oversold the significance of this result. When I went into this rabbithole, I was under the impression from Folland that I'd be proving that if μ is an arbitrary Radon measure and $\nu \ll \mu$, then ν always has the form $f d\mu$. But that is not what was proven in theorems 19.27 and 19.32 so I don't know what Folland was yappin about. Considering the lack of specific details that Folland gives, I can't help but suspect that Folland didn't entirely understand what Hewitt and Stromberg actually proved. Although, I don't necessarily blame Folland for that.

Hewitt and Stromberg were difficult for me to understand partly because they had a lot of (in my opinion) very questionable conventions. For example, I tried my best to hide it away as much as possible but Hewitt and Stromberg explicitely treated every measure as the restriction of an outer measure to the σ-algebra of μ^*-measurable sets rather than abstracting that away because I guess they always wanted their σ-algebras to be as large as possible. In fact, while I was skimming the book I even saw some other sections where they talked about extending their measures to even larger σ-algebras. Never mind that there are advantages to working on a smaller σ-algebra (see for instance page 55 of my journal: when working on a larger σ-algebra we lose that continuity implies measurability...).

I'll also mention that the choice of math font in their book is awful and I couldn't help but wonder if they knew what kerning is. Anyways I guess my review of their book is that it feels dated and very clunky. Although, maybe I only had a mixed experience because I'm not an intended reader. (cough cough)

Anyways, while getting a milkshake with my apartment-mates, I thought of a few ideas for how to make the preceding theorems actually useful.

My first challenge is that I'm primarily working with Borel Radon measures while Hewitt and Stromberg weren't. This leads me to the following result:

Proposition: Suppose X is an LCH space and let μ and ν be positive Radon measures on $(X, \mathcal{B}_X)$ such that $\nu \ll \mu$. Then let $(X, \mathcal{M}_\mu, \overline{\mu})$ be the saturation of the completion of $(X, \mathcal{B}_x, \mu)$ and let $(X, \mathcal{M}_\nu, \overline{\nu})$ be the saturation of the completion of $(X, \mathcal{B}_x, \nu)$.

- In general $\mathcal{M}_\mu \neq \mathcal{M}_\nu$.

  Proof:
  Let $X = \mathbb{R}$. Then set μ to be the Lebesgue measure and set ν equal to the zero measure (i.e. the measure for which every set is null). Note that we trivially have that $\nu \ll \mu$. Also, both are easily seen to be Radon measures.

  Since μ and ν are σ-finite, we know by the exercise on pages 158 and 159 that $\mathcal{M}_\mu$ and $\mathcal{M}_\nu$ are just the completions of $\mathcal{M}$ with respect to μ and ν respectively. Since $\nu(X) = 0$, we have that $\mathcal{M}_\nu = \mathcal{P}(X)$. Meanwhile, because of the existance of Vitali sets, we know that $\mathcal{M}_\mu \neq \mathcal{P}(X)$. ■.

- We do always have that $\mathcal{M}_\mu \subseteq \mathcal{M}_\nu$ when μ and ν are Radon and $\nu \ll \mu$.

  Proof: (I got this argument from Hewitt and Stromberg 19.33)
  Suppose $A \in \mathcal{M}_\mu$ and let $F \subseteq X$ be any compact set. Since A is locally measurable with respect to the completion of μ, we know that $A \cap F$ is in the completion of $\mathcal{B}_X$ with respect to μ. So let $A \cap F = E \cup N$ where $E \in \mathcal{B}_X$ and $N \subseteq N'$ with $N' \in \mathcal{B}_X$ and $\mu(N') = 0$. Then since $\mu(N') = 0$ implies that $\nu(N') = 0$, we know that $A \cap F$ is in the complection of $\mathcal{B}_X$ with respect to ν. By the lemma on page 161, this proves that $A \in \mathcal{M}_\nu$. ■

- We do always have that $\overline{\nu}|_{\mathcal{M}_\mu} \ll \overline{\mu}$.

  Proof:
  Suppose $A \in \mathcal{M}_\mu$ with $\overline{\mu}(A) = 0$. Since A has finite measure, we know that A is not merely locally measurable but also that A is in the completion of $\mathcal{B}_X$ with respect to μ. Hence, there exists a set $N \in \mathcal{B}_X$ such that $A \subseteq N$ and $\mu(N) = 0$. But now since $\nu \ll \mu$, we have that $\nu(N) = 0$. In turn, $\overline{\nu}(A) \leq \nu(N) = 0$. This proves that $\overline{\mu}(A) = 0 \Longrightarrow \overline{\nu}(A) = 0$. ■

It follows that if X is an LCH space, then given any two positive Radon measures μ and ν on $(X, \mathcal{B}_X)$ with $\nu \ll \mu$ we can extend μ and ν to measures $\overline{\mu}$ and $\overline{\nu}$ on a common σ-algebra $\mathcal{M}$ with precisely the following properties:

- $\overline{\mu}$ and $\overline{\nu}$ are both outer regular on all sets and inner regular on all σ-finite sets;
- $\overline{\mu}$ and $\overline{\nu}$ are both finite on all compact sets;
- $(X, \mathcal{M}, \overline{\mu})$ is the saturation of the completion of $(X, \mathcal{B}_X, \mu)$;
- $(X, \mathcal{M}, \overline{\mu})$ is decomposable;
- $\overline{\nu} \ll \overline{\mu}$.

Now using the theorems from Hewitt and Stromberg, I want to show that $\overline{\nu} = \overline{f}\mathrm{d}\overline{\mu}$ for some $\overline{\mu}$-measurable function $\overline{f}$. As it turns out, when ν is σ-finite (which in turn means $\overline{\nu}$ is σ-finite), this is always possible:

Proposition: Let $(X, \mathcal{M})$ be a measurable LCH space and suppose μ and ν are positive measures on $(X, \mathcal{M})$ satisfying that ν is inner regular on all σ-finite sets; that μ is finite on all compact sets; that $\nu \ll \mu$; and that $(X, \mathcal{M}, \mu)$ is decomposable via the decomposition $\mathcal{F}$. If ν is σ-finite, then there exists an $\mathcal{M}$-measurable function $f : X \to [0, \infty)$ which vanishes outside a set on which μ is σ-finite such that $\nu(A) = \int_A f\mathrm{d}\mu$ for all $A \in \mathcal{M}$. Furthermore, if ν is finite, then $f \in L^1(\mu)$. And if $g : X \to [0, \infty)$ is another function satisfying $\nu(A) = \int_A g\mathrm{d}\mu$ for all $A \in \mathcal{M}$, then $f = g$ μ-a.e.

Proof:

We shall first suppose that ν is finite. Then by applying theorem 19.27 on page 156, we know there exists a measurable function $f : X \to [0, \infty)$ satisfying that $\nu(A) = \int_A f\mathrm{d}\mu$ whenever μ is σ-finite on A. Furthermore, since $\mu(F) < \infty$ for all $F \in \mathcal{F}$, we thus know that $\nu(F) = \int_F f\mathrm{d}\mu$ for all $F \in \mathcal{F}$.

Now recall that $\sum_{F\in\mathcal{F}} \nu(F)$ is by definition equal to the suprememum of all finite sums of the $\nu(F)$. Thus, while we cannot currently guarentee strict equality due to the fact that $\mathcal{F}$ may be uncountable, we can at least say that $\nu(X) \geq \sum_{F\in\mathcal{F}} \nu(F)$. And since $\nu(X) < \infty$, this proves that there is a countable subset $\mathcal{F}_1 \subseteq \mathcal{F}$ satisfying that $\nu(F) = 0$ for all $F \notin \mathcal{F}_1$. And clearly $\nu(X) = \nu(\bigcup\limits_{F\notin\mathcal{F}_1} F) + \sum\limits_{F\in\mathcal{F}_1} \nu(F)$.

Now we claim $\nu(\bigcup_{F\notin\mathcal{F}_1} F) = 0$. After all, if the set weren't null, then we'd have that $0 < \nu(\bigcup_{F\notin\mathcal{F}_1} F) \subseteq \nu(X) < \infty$. It would thus follow by the inner regularity of ν that there is a compact set $K \subseteq \bigcup_{F\notin\mathcal{F}_1} F$ with $\nu(K) > 0$. But then we'd have that $\mu(K) < \infty$, which means $\nu(K) = \int_K f\mathrm{d}\mu$ and $\mu(K) = \sum_{F\in\mathcal{F}} \mu(F \cap K)$.

Now we can't have that $\mu(K) = 0$ since that contradicts that $\nu \ll \mu$. Hence, since μ is decomposable there's a nonempty countable family $\mathcal{F}_2 \subseteq \mathcal{F}$ such that $\mu(F \cap K) > 0$ for all $F \in \mathcal{F}_2$ and $\mu(K) = \sum_{F\in\mathcal{F}_2} \mu(F \cap K)$ (which in turn means $\mu(\bigcup_{F\notin\mathcal{F}_2} F \cap K) = 0$). Also, since K doesn't intercept any $F \in \mathcal{F}_1$, we know that $\mathcal{F}_1 \cap \mathcal{F}_2 = \emptyset$. But now note that:

$$0 < \nu(K) = \int_K f\mathrm{d}\mu = \sum_{F\in\mathcal{F}_2} \int_{F\cap K} f\mathrm{d}\mu = \sum_{F\in\mathcal{F}_2} \nu(F \cap K).$$

This implies that there is some $F \in \mathcal{F}_2$ such that $\nu(F) > \nu(F \cap K) > 0$. But that is a contradiction since we already know that $\nu(F) = 0$ if $F \notin \mathcal{F}_1$. Hence, we have proven that there is a set $E = \bigcup_{F \in \mathcal{F}_1} F$ which μ is σ-finite on such that $\nu(X) = \nu(E)$. And in turn, we've now proven that $\nu = f\chi_E \mathrm{d}\mu$.

And as a side note: $f\chi_E \in L^1(\mu)$ because $\int f\chi_E \mathrm{d}\mu = \nu(X) < \infty$.

Next, we consider the case where ν is σ-finite. Let $(X_n)_{n\in\mathbb{N}}$ be a partition of X consisting of finitely measurable sets. Then by identical reasoning as before, we can find for each n a set $E_n \subseteq X$ on which μ is σ-finite and $\nu(X_n \cap A) = \int_A f\chi_{E_n}\mathrm{d}\mu$ for all $A \in \mathcal{M}$. Also, since $E_n \subseteq X_n$ for all n, it is clear that they are all disjoint. So, $E := \bigcup_{n\in\mathbb{N}} E_n$ satisfies that μ is σ-finite on E and $f\chi_E = \sum_{n\in\mathbb{N}} f\chi_{E_n}$. Also, it is clear that $\nu = f\chi_E\mathrm{d}\mu$ since:

$$\nu(A) = \sum_{n\in\mathbb{N}} \nu(X_n \cap A) = \sum_{n\in\mathbb{N}} \int_A f\chi_{E_n}\mathrm{d}\mu = \int_A f\chi_E\mathrm{d}\mu \text{ for all } A \in \mathcal{M}.$$

Finally, we show uniqueness. Suppose $g : X \to [0,\infty)$ is another function satisfying that $\nu = g\mathrm{d}\mu$ and let $B = \{x \in X : f(x)\chi_E(x) \neq g(x)\}$. Then it's clear that:

$$B = (B \cap E) \cup (B - E).$$

Since E is σ-finite, we already know from theorem 19.27 that there is a μ-null set N_1 with $B \cap E \subseteq N_1$. Also, $B - E = \{x \in X - E : g(x) > 0\}$. So, if $\mu(B - E) > 0$, then we'd have that $\nu(B - E) = \int_{B-E} g\mathrm{d}\mu > 0$. But that contradicts that $\nu(X - E) = 0$. So, we can conclude that $\mu(B - E) = 0$. This proves that $f\chi_E = g$ μ-a.e. ■

To finish making this useful for our purposes we need to find a $\mathcal{B}_X$-measurable function f such that $\overline{f} = f$ $\overline{\mu}$-a.e. That way, for all $A \in \mathcal{B}_X$ we have that:

$$\int_A f\mathrm{d}\mu = \int_A f\mathrm{d}\overline{\mu} = \int_A \overline{f}\mathrm{d}\overline{\mu} = \overline{\nu}(A) = \nu(A)$$

Fortunately, if we let $(X, \mathcal{N}, \overline{\mu}|_{\mathcal{N}})$ be the completion of $(X, \mathcal{B}_X, \mu)$ (meaning $(X, \mathcal{M}, \overline{\mu})$ is the saturation of $(X, \mathcal{N}, \overline{\mu}|_{\mathcal{N}})$), then we know that $\overline{f}$ is $\mathcal{N}$-measurable.

Proof:
We know $\overline{f}$ vanishes outside a set on which $\overline{\mu}$ is σ-finite. It follows that if $a < 0$, then $\overline{f}^{-1}((a,\infty)) = X \in \mathcal{N}$; and if $a \geq 0$, then $\overline{f}^{-1}((a,\infty))$ is σ-finite. Now we claim that all sets for which $\overline{\mu}$ is σ-finite are in $\mathcal{N}$. After all, if $\overline{\mu}$ is σ-finite on E, then we know there is a sequence of sets $(E_n)_{n\in\mathbb{N}}$ whose union is E and which satisfy that $\overline{\mu}(E_n) < \infty$ for all n. Then since all the E_n have finite measure, we know that $E_n \in \mathcal{N}$ for all n. So, E is also in $\mathcal{N}$.

With that, we've shown that $\overline{f}^{-1}((a,\infty))$ is $\mathcal{N}$-measurable for all $a \in \mathbb{R}$. This proves that $\overline{f}$ is $\mathcal{N}$-measurable. ■

Now, it is a simple application of the proposition on page 46 of my latex math 240a notes to show there exists a $\mathcal{B}_X$-measurable function f such that $\overline{f} = f$ a.e. with respect to $(X, \mathcal{N}, \overline{\mu}|_{N_x})$. And since saturating the latter σ-algebra doesn't add any new null sets, we have that $f = \overline{f}$ $\overline{\mu}$-a.e. Thus, we have found a $\mathcal{B}_X$-measurable function f such that $\nu = f\mathrm{d}\mu$.

One more note I want to make is that if g is another $\mathcal{B}_X$-measurable function such that $\nu = g\mathrm{d}\mu$, then $f = g$ μ-a.e.

Proof:

Since $\overline{f}$ vanishes outside a set on which $\overline{\mu}$ is σ-finite and $\overline{f} = f$ $\overline{\mu}$-a.e., we also know that f vanishes outside a set on which $\overline{\mu}$ is σ-finite. If we call that set E, then we have that $f = f\chi_E$ everywhere. Also note that since E is in $\mathcal{N}$ on account of $\overline{\mu}$ being σ-finite on it, we can expand E to a larger set in $\mathcal{B}_X$ that $\overline{\mu}$ is still σ-finite on. Hence, we may without loss of generality just say that $E \in \mathcal{B}_X$. One final note is that if $\overline{\mu}$ is σ-finite on E, then we also know that μ is σ-finite on E. So, we conclude that $f = f\chi_E$ where $E \in \mathcal{B}_X$ and μ is σ-finite on E.

Now let $B := \{x \in X : g(x) \neq f(x)\}$. Then we know that $B = (B \cap E) \cup (B - E)$. Our claim is that both of the latter sets are null with respect to μ.

First suppose that $\mu(B - E) > 0$. Then since $B - E = \{x \in X - E : g(x) > 0\}$, we'd have to have that $\nu(B - E) = \int_{B-E} g\mathrm{d}\mu > 0$. But that contradicts that $\nu(B - E) = \int_{B-E} f\mathrm{d}\mu = \int_{B-E} 0\mathrm{d}\mu = 0$. So, we conclude that $\mu(B - E) = 0$.

Meanwhile, if $B \cap E$ is not a null set, then because μ is σ-finite (and thus also semifinite) on E, we must be able to pick a set $C \in \mathcal{B}_X$ such that $0 < \mu(C) < \infty$ and $g(x) \neq f(x)$ for all $x \in C$. Then in turn we'd have that:

$$\int_C |f(x) - g(x)|\mathrm{d}\mu > 0$$

But note that since ν is σ-finite, we know there there is a disjoint sequence of sets $\{C_n\}_{n\in\mathbb{N}}$ such that $C = \bigcup_{n\in\mathbb{N}} C_n$ and $\nu(C_n) = \int_{C_n} f\mathrm{d}\mu = \int_{C_n} g\mathrm{d}\mu < \infty$ for all n. Also, for each n we can define $C_n^+ := \{x \in C_n : f(x) > g(x)\}$ and $C_n^- := \{x \in C_n : g(x) > f(x)\}$. And thus:

$$\int_C |f(x) - g(x)|\mathrm{d}\mu = \sum_{n\in\mathbb{N}} \int_{C_n^+} (f(x) - g(x))\mathrm{d}\mu + \sum_{n\in\mathbb{N}} \int_{C_n^-} (g(x) - f(x))\mathrm{d}\mu$$

But now because $\int_{C_n^\pm} f(x)\mathrm{d}\mu = \int_{C_n^\pm} g(x)\mathrm{d}\mu = \nu(C_n^\pm)$ where $\nu(C_n^\pm)$ is finite, we must have that $\int_C |f(x) - g(x)|\mathrm{d}\mu = 0$ as:

$$\int_{C_n^+} (f(x) - g(x))\mathrm{d}\mu = 0 = \int_{C_n^-} (g(x) - f(x))\mathrm{d}\mu \text{ for all } n.$$

This is a contradiction. So, we've proven that $\mu(B \cap E) = 0$. ■

So to sum all my previous work up, we have the following result:

<u>A Third Extension of the Lebesgue-Radon-Nikodym Theorem:</u>

If X is an LCH space and μ and ν are positive Radon measures on $(X, \mathcal{B}_X)$ such that ν is σ-finite and $\nu \ll \mu$, then there exists a measurable function $f : X \to [0, \infty)$ which vanishes outside of a set where μ is σ-finite and which satisfies that $\nu = f\mathrm{d}\mu$. Furthermore, if g is another measurable function satisfying that $\nu = g\mathrm{d}\mu$, then $f = g$ μ-a.e.

Also, we can clearly extend this theorem to the case where ν is signed by just applying the theorem to the positive and negative variations separately of ν. And in order to prove uniqueness in this case we can employ a similar strategy as was shown at the top of page 158.

Going a step further, if ν is complex, then we can apply the prior reasoning to the real and imaginary variations of ν. Thus, we get the most general result that I will attempt to prove:

A Fourth Extension of the Lebesgue-Radon-Nikodym Theorem:

Suppose X is an LCH space and μ is a Radon measure on $(X, \mathcal{B}_X)$. If ν is a complex or σ-finite signed Radon measure with $\nu \ll \mu$, then there exists a measurable function $f : X \to \mathbb{R}$ (or $\mathbb{C}$) which vanishes outside of a set where μ is σ-finite and which satisfies that $\nu = f\mathrm{d}\mu$. Furthermore, if g is another measurable function satisfying that $\nu = g\mathrm{d}\mu$, then $f = g$ μ-a.e.

I will note that this is still not as general of a statement as Folland was claiming Hewitt and Stromberg were making. Yet it is enough for a specific claim which Folland makes and I will take notes on tomorrow to be true. Anyways, it's been four days since I started going down this rabbithole and I need to grade.

8/30/2025

Let X be an LCH space. Given any fixed positive Radon measure μ, we can isometrically embed $L^1(\mu)$ into $M(X)$ as follows:

Define a map $L^1(\mu) \hookrightarrow M(X)$ such that $f \mapsto f\mathrm{d}\mu$ for all $f \in L^1(\mu)$. By the exercise on page 154, we know that $f\mathrm{d}\mu$ is in fact Radon. Also, we know that the image of this map is precisely the subset of $M(X)$ consisting of all complex measures that are absolutely continuous with respect to μ (this works even if μ is not Radon by the rabbithole I went down this past week).

Also, we can easily see that $\|f\mathrm{d}\mu\|_{M(X)} = \|f\|_1$.

I want to be slightly careful about saying this since I'm not assuming μ is σ-finite.

Let $\nu = f\mathrm{d}\mu$. Now we know from math 240a that there exists a measurable function g with $|g| = 1$ $|\nu|$-a.e. such that $\nu = g\mathrm{d}|\nu|$. Furthermore, we know from the rabbit hole before that there exists a nonnegative function $\frac{\mathrm{d}|\nu|}{\mathrm{d}\mu}$ such that $\mathrm{d}|\nu| = \frac{\mathrm{d}|\nu|}{\mathrm{d}\mu}\mathrm{d}\mu$. Hence, we have that $\nu = g\frac{\mathrm{d}|\nu|}{\mathrm{d}\mu}\mathrm{d}\mu$, and this proves that $f = g\frac{\mathrm{d}|\nu|}{\mathrm{d}\mu}$ a.e.. Then in turn we also have that $|f| = \frac{\mathrm{d}|\nu|}{\mathrm{d}\mu}$ and $|\nu| = |f|\mathrm{d}\mu$. This proves that even if μ is not σ-finite, we still have that $\nu = f\mathrm{d}\mu \Longrightarrow |\nu| = |f|\mathrm{d}\mu$.

An important application of the above fact is that if m is the Lebesgue measure on $\mathbb{R}^n$, then we can identify $L^1(m)$ as a subspace $M(\mathbb{R}^n)$. This will be used in my journal later.

Given a net $\langle \mu_\alpha \rangle_{\alpha \in A}$ and another measure μ in $M(X)$, we have that $\mu_\alpha \to \mu$ in the weak* topology on $M(X) = C_0(X)^*$ iff we have that $\int f\mathrm{d}\mu_\alpha \to \int f\mathrm{d}\mu$ for all $f \in C_0(X)$. This topology is important enough to be given a second name: the vague topology on $M(X)$.

Before covering the next theorem in Folland, I'm gonna do another side quest concerning the Stone-Weierstrass Theorem.

Exercise 4.67: Let X be a noncompact LCH space. If $\mathcal{A}$ is a closed subalgebra of $C_0(X, \mathbb{R})$ that separates points, then either $\mathcal{A} = C_0(X, \mathbb{R})$ or $\mathcal{A} = \{f \in C_0(X, \mathbb{R}) : f(x_0) = 0\}$ for some $x_0 \in X$.

Proof:
First suppose there does not exist any $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathcal{A}$. Then let Y be the Alexandroff (i.e. one-point) compactification of X and $i : C_0(X, \mathbb{R}) \hookrightarrow C(Y, \mathbb{R})$ be the injective map that continuously extends each f to Y by setting $f(\infty) = 0$. Now $i(\mathcal{A})$ is a subalgebra of $C(Y, \mathbb{R})$ that separates points and for which $x_0 := \infty$ satisfies that $f(x_0) = 0$ for all $f \in i(\mathcal{A})$. So by the Stone Weierstrass theorem we already proved, we know that:

$$\overline{i(\mathcal{A})} = \{f \in C(Y, \mathbb{R}) : f(\infty) = 0\}$$

But now we claim that $i(\mathcal{A})$ is closed. For suppose $g \in \overline{i(\mathcal{A})}$ and $\{f_n\}_{n \in \mathbb{N}}$ is a sequence in $i(\mathcal{A})$ converging uniformly to g. Then it's clear that $f_n|_X \to g|_X$ uniformly. So $g|_X \in \mathcal{A}$ since $\mathcal{A}$ is closed. And in turn $g \in i(\mathcal{A})$.

This shows that $i(\mathcal{A}) = \overline{i(\mathcal{A})} = \{f \in C(Y, \mathbb{R}) : f(\infty) = 0\}$. It then follows that $\mathcal{A} = C_0(X, \mathbb{R})$.

Next suppose there does exist some $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathcal{A}$. This poses a problem to our previous approach because $i(\mathcal{A})$ no longer separates points. So to get around this, we first consider the subspace $X' = X - \{x_0\}$. Importantly, X' is still Hausdorff. Also, since X' is an open subset of X, we know any $x \in X'$ has a compact neighborhood $N \subseteq X'$. Hence, X' is locally compact.

Next, let j be the map restricting the domain of each $f \in C_0(X, \mathbb{R})$ to X'. Then note that if $f(x_0) = 0$, then $f|_{X'} \in C_0(X', \mathbb{R})$. After all, if $\{x \in X : f(x) > \varepsilon\}$ is compact and entirely contained in X', then we also know that $\{x \in X' : f|_X(x) > \varepsilon\}$ is compact in the subspace topology of X'. As a result of this, we know that j is an injective map from $\{f \in C_0(X, \mathbb{R}) : f(x_0) = 0\}$ into $C_0(X', \mathbb{R})$.

Now, it is easily seen that $j(\mathcal{A})$ is an algebra that separates points and vanishes nowhere. And, it is also seen similarly to earlier that $j(\mathcal{A})$ is closed. But then by the first case we proved in this exercise, we know that $j(\mathcal{A}) = C_0(X', \mathbb{R})$. Hence, it follows that $\mathcal{A} = \{f \in C_0(X, \mathbb{R}) : f(x_0) = 0\}$. ■

Now in the next theorem, Folland brought up this exercise because he needs it to prove that $C_c^1(\mathbb{R})$ is dense in $C_0(\mathbb{R})$. However, in my math 240c notes I already proved a stronger statement that $C_c^\infty(\mathbb{R})$ is dense in $C_0(\mathbb{R})$. So was proving this unnecessary? No comment.

Proposition 7.19: Suppose $\mu, \mu_1, \mu_2, \ldots \in M(\mathbb{R})$ and let $F_n(x) = \mu_n((-\infty, x])$ and $F(x) = \mu((-\infty, x])$.

(a) If $\sup_{n \in \mathbb{N}} \|\mu_n\| < \infty$ and $F_n(x) \to F(x)$ for every x at which F is continuous, then $\mu_n \to \mu$ vaguely.

Proof:
By Folland theorem 3.29 (check my paper math 240a notes), we know F and each of the F_n are all in NBV. In turn, this means that F and all the F_n are continuous except at countably many points. So, our assumed conditions actually guarentee that $F_n \to F$ a.e. with respect to the Lebesgue measure.

Now consider any $f \in C_c^1(\mathbb{R})$. By applying theorem 3.36 from math 240a, we can say that $\int f \mathrm{d}\mu = -\int f' F \mathrm{d}x$ and that $\int f \mathrm{d}\mu_n = -\int f' F_n \mathrm{d}x$ for all n.

Side note: I just realized why theorem 3.36 is a strict generalization of integration by parts as taught in undergrad analysis. If f is a C^1 function on $[a, b]$, then we know f is absolutely continuous on $[a, b]$ via the mean value theorem. And it follows then that the measure $f'\mathrm{d}t$ satisfies that $f(x) - f(a) = \int_{(a,x]} f'(t)\mathrm{d}t$ for all $x \in [a, b]$.

Next suppose f is C^1 everywhere, $f' \in L^1(\mathbb{R})$, and $f(-\infty) = 0$. Then by taking $a \to -\infty$ and $b \to +\infty$ in the last paragraph, we get that $f'\mathrm{d}t$ is a measure satisfying that $f(x) = \int_{(-\infty,x]} f'\mathrm{d}t$.

Thus, so long as f is C^1, $f(-\infty) = 0$, and $f' \in L^1$, we know that $f'\mathrm{d}t$ is the unique borel measure such that $f(x) = \int_{(-\infty,x]} f'\mathrm{d}t$. And since f is continuous, we can always apply theorem 3.36 to f and any other NBV function.

Now if only I had thought about that while I was actually taking math 240. This is why I failed the qual. AAUGH.

But note that $\|F_n\|_u \leq \|\mu_n\|$ for all n. So if we set $C = \|f'\|_u \sup_{n\in\mathbb{N}} \|\mu_n\|$, then we can apply dominated convergence theorem using an upper bound of $C\chi_{\mathrm{supp}(f')}$ to show that $-\int f' F_n \mathrm{d}x \to -\int f' F \mathrm{d}x$ as $n \to \infty$. Hence, $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu$ as $n \to \infty$.

Consequently, we have shown that $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu$ on a dense subset of $C_0(\mathbb{R})$. And so, by proposition 5.17 (see math 240b notes), we know $\int f \mathrm{d}\mu_n \to \int \mathrm{d}\mu$ for all $f \in C_0(\mathbb{R})$. This proves that $\mu_n \to \mu$ vaguely. ■

(b) If $\mu_n \to \mu$ vaguely, then $\sup_{n\in\mathbb{N}} \|\mu_n\| < \infty$. If in addition the μ_n are positive, then $F_n(x) \to F(x)$ at every x at which F is continuous.

Proof:
By the Riesz Representation theore, we already know that the linear functional $f \mapsto \int f \mathrm{d}\mu_n$ from $C_0(\mathbb{R})$ to $\mathbb{C}$ has the operator norm $\|\mu_n\|$. Also, since $\mu_n \to \mu$ vaguely, we know that $\sup_{n\in\mathbb{N}} |\int f \mathrm{d}\mu_n| < \infty$ for all $f \in C_0(\mathbb{R})$. Thus by the uniform boundedness principle, we know that $\sup_{n\in\mathbb{N}} \|\mu_n\| < \infty$.

Next suppose all the μ_n are positive. Then since $\mu_n \to \mu$ vaguely, we claim that μ is positive.

Note: The following the reasoning works on any general LCH space X.

Write $\mu = \mu^{(1)} - \mu^{(2)} + i(\mu^{(3)} - \mu^{(4)})$ where the four latter measures are positive Radon measures. Since $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu$ for all $f \in C_0(X)$ and all the μ_n are positive, we know that if f is nonnegative and real, then $\int f \mathrm{d}\mu \geq 0$. This is enough to show that $\mu^{(i)} = 0$ unless $i = 1$.

Suppose for the sake of contradiction that $\mu^{(2)} = \alpha > 0$ on some set $A \in \mathcal{B}_X$. Then by restricting our set A using a Hahn decomposition, we can also say without loss of generality that $\mu^{(1)}(A) = 0$. But now pick any $\varepsilon \in (0, \frac{\alpha}{2})$. By the outer regularity of $\mu^{(1)}$ we know there exists an open set $U \supseteq A$ such that $\mu^{(1)}(U) < \varepsilon$. At the same time, by the inner regularity of $\mu^{(2)}$ we know there exists a compact set $K \subseteq A$ such that $\mu^{(2)}(K) > \alpha - \varepsilon$. And by Urysohn's lemma, we know there is a function $\phi \in C_c(X, [0,1])$ such that $\phi(K) = \{1\}$ and $\operatorname{supp}(\phi) \subseteq U$. It now follows that:

$$\begin{aligned}\operatorname{Re}(\textstyle\int \phi \mathrm{d}\mu) &= \textstyle\int \phi \mathrm{d}\mu^{(1)} - \int \phi \mathrm{d}\mu^{(2)} \\ &\leq \textstyle\int \chi_U \mathrm{d}\mu^{(1)} - \int \chi_K \mathrm{d}\mu^{(2)} < \varepsilon - (\alpha - \varepsilon) < 0.\end{aligned}$$

But that contradicts our earlier statement that $\int \phi \mathrm{d}\mu$ is real and nonnegative if ϕ is real and nonnegative. So, we conclude no such A exists.

Similar reasoning shows that $\mu^{(3)}$ and $\mu^{(4)}$ are zero.

Now given any a at which F is continuous, choose any $\varepsilon > 0$ and N such that $-N < a - 2\varepsilon$. Then letting $f \in C_c(\mathbb{R})$ be the function that is 1 on $[-N, a]$, 0 on $(-\infty, -N - \varepsilon) \cup (a + \varepsilon, \infty)$, we have that:

$$\begin{aligned}F_n(a) - F_n(-N) &= \mu_n((-N, a]) \\ &\leq \textstyle\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu \leq F(a + \varepsilon) - F(-N - \varepsilon).\end{aligned}$$

And by taking $N \to \infty$ we thus get that $\limsup_{n \to \infty} F_n(a) \leq F(a + \varepsilon)$.

Meanwhile, letting $g \in C_c(\mathbb{R})$ be the function that is 1 on $[-N + \varepsilon, a - \varepsilon]$, 0 on $(-\infty, -N] \cup [a, \infty]$, and linear in between, we have that:

$$\begin{aligned}F_n(a) - F_n(-N) &= \mu_n((-N, a)) \\ &\geq \textstyle\int g \mathrm{d}\mu_n \to \int g \mathrm{d}\mu \geq F(a - \varepsilon) - F(-N + \varepsilon).\end{aligned}$$

And by taking $N \to \infty$ we thus get that $\liminf_{n \in \infty} F_n(a) \geq F(a - \varepsilon)$.

Now by taking $\varepsilon \to 0$ and using the fact that F is continuous at a, we prove that $F_n(a) \to F(a)$ as $n \to \infty$. ■

9/1/2025

Today I want to clean up my knowledge about product measures by doing some exercises from Folland that I never got around to while I was taking math 240.

Exercise 2.11: If $f : \mathbb{R} \times \mathbb{R}^k \to \mathbb{C}$ satisfies that $f(x, \cdot)$ is measurable on $\mathbb{R}^k$ for all $x \in \mathbb{R}$ and that $f(\cdot, y)$ is continuous on $\mathbb{R}$ for all $y \in \mathbb{R}^k$, then f is Borel measurable on $\mathbb{R} \times \mathbb{R}^k$.

Proof:

Let n be any positive integer and for all $i \in \mathbb{Z}$, define $a_i = {}^i/_n$. Then define a function f_n on $\mathbb{R} \times \mathbb{R}^k$ by setting for all $y \in \mathbb{R}^k$ and $x \in [a_i, a_{i+1}]$:

$$f_n(x,y) = \frac{f(a_{i+1}, y)(x - a_i) - f(a_i, y)(x - a_{i+1})}{a_{i+1} - a_i}$$

In order to show that each f_n is well defined, note that if $x = a_i$ for some $i \in \mathbb{Z}$, then:

$$\frac{f(a_i,y)(x-a_{i-1})-f(a_{i-1},y)(x-a_i)}{a_i-a_{i-1}} = f(a_i, y) = \frac{f(a_{i+1},y)(x-a_i)-f(a_i,y)(x-a_{i+1})}{a_{i+1}-a_i}$$

You may note that f_n is essentially linearly interpolating because the differents sets $\{a_i\} \times \mathbb{R}^k$ of the domain. And as $n \to \infty$ we are sampling more often.

We also claim that f_n is measurable. After all, if $E \in \mathcal{M}$, then:

$$f_n^{-1}(E) = \bigcup_{i \in \mathbb{Z}} (f_n^{-1}(E) \cap ([a_i, a_{i+1}] \times \mathbb{R}^k))$$

But now we know that f_n is measurable on the domain $[a_i, a_{i+1}] \times \mathbb{R}^k$ since it is equal to a sum of products of measurable functions. $g(x, y) := f(a_{i+1}, y)$ is measurable as a function from $\mathbb{R} \times \mathbb{R}^k$ because the projection map $(x, y) \mapsto (a_i, y)$ is a continuous map from $\mathbb{R} \times \mathbb{R}^k$ to $\mathbb{R}^k$ and we already assumed that $f(a_i, \cdot)$ is measurable on $\mathbb{R}^k$. Similarly, we can show that $h(x, y) := f(a_i, y)$ is measurable. And since $(x - a_i)$ and $(x - a_{i+1})$ are continuous, we also know those parts of the expression for f_n are measurable. And since $[a_i, a_{i+1}] \times \mathbb{R}^k$ is in $\mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}^k}$ for each i, we can now conclude f_n is measurable on its entire domain.

Now in order to show that f is measurable, we claim that $f_n \to f$ pointwise. To prove this, consider any $x \in \mathbb{R}$ and $y \in \mathbb{R}^k$ and let $\varepsilon > 0$. Since $f(\cdot, y)$ is continuous, we know that there exists $\delta > 0$ such that $|f(x', y) - f(x, y)| < \varepsilon$ for all x' satisfying that $|x' - x| < \delta$. Now suppose n is large enough so that ${}^1/_n < {}^\delta/_3$. Then we know there exists $i \in \mathbb{Z}$ such that $x - \delta < a_i \leq x \leq a_{i+1} < x + \delta$. And in turn:

$$\begin{aligned}
|f_n(x,y) - f(x,y)| &= \left| \tfrac{f(a_{i+1},y)(x-a_i)-f(a_i,y)(x-a_{i+1})}{a_{i+1}-a_i} - f(x,y) \right| \\
&= \left| \tfrac{f(a_{i+1},y)(x-a_i)-f(a_i,y)(x-a_{i+1})}{a_{i+1}-a_i} - f(x,y)\tfrac{(x-a_i)-(x-a_{i+1})}{a_{i+1}-a_i} \right| \\
&= \left| \tfrac{(f(a_{i+1},y)-f(x,y))(x-a_i)}{a_{i+1}-a_i} - \tfrac{(f(a_i,y)-f(x,y))(x-a_{i+1})}{a_{i+1}-a_i} \right| \\
&= |f(a_{i+1},y) - f(x,y)|\left|\tfrac{(x-a_i)}{a_{i+1}-a_i}\right| + |f(a_i,y) - f(x,y)|\left|\tfrac{(x-a_{i+1})}{a_{i+1}-a_i}\right| \\
&\leq |f(a_{i+1},y) - f(x,y)| \cdot 1 + |f(a_i,y) - f(x,y)| \cdot 1 \\
&< \varepsilon + \varepsilon = 2\varepsilon. \blacksquare
\end{aligned}$$

As a corollary, if $f : \mathbb{R}^n \to \mathbb{C}$ is continuous in each variable separately, then f is measurable.

Proof:
Suppose we already proved that if $g : \mathbb{R}^{n-1} \to \mathbb{C}$ is continuous in each variable, then g is Borel measurable. Then consider f as a function from $\mathbb{R} \times \mathbb{R}^{n-1}$ to $\mathbb{C}$. We know by our assumption that $f(\cdot, y)$ is continuous for all $y \in \mathbb{R}^{n-1}$. Also, we know by our inductive hypothesis that $f(x, \cdot)$ is measurable for all $x \in \mathbb{R}$. So by the reasoning in the last exercise, we have that f is Borel measurable on $\mathbb{R} \times \mathbb{R}^{n-1}$.

And since $\mathcal{B}_{\mathbb{R}\times\mathbb{R}^{n-1}} = \mathcal{B}_{\mathbb{R}^n}$ when identifying $\mathbb{R} \times \mathbb{R}^{n-1}$ with $\mathbb{R}^n$, we've shown that f is measurable. ■

I realize I still have not yet properly convinced myself of why $\mathcal{B}_{\mathbb{R}\times\mathbb{R}^{n-1}} \cong \mathcal{B}_{\mathbb{R}^n}$. So, I'd like to set that and a few other conerns about products of measures straight tomorrow. But for now I need to study for physics.

One more note I'd like to make is that this exercise is significant because we don't generally have that $f : \mathbb{R}^n \to \mathbb{C}$ being continuous with respect to each variable separately implies that f is continuous. So essentially this exercise gives a strictly weaker sufficient condition for $f : \mathbb{R}^n \to \mathbb{C}$ to be Borel measurable than strict continuity. Also, this new condition is way easier to show.

9/2/2025

Exercise 2.45: Let $n \geq 3$ and suppose $(X_j, \mathcal{M}_j, \mu_j)$ is a measure space for $j = 1, \ldots, n$, and let us identify $\prod_{j=1}^n X_j$ with $(\prod_{j=1}^{n-1} X_j) \times X_n$ and $X_1 \times (\prod_{j=2}^n X_j)$ in the obvious ways.

- We have that $\bigotimes_{j=1}^n \mathcal{M}_j = (\bigotimes_{j=1}^{n-1} \mathcal{M}_j) \otimes \mathcal{M}_n$.

Proof:
For $i = 1, \ldots, n$ let π_i be the projection of $\prod_{j=1}^n X_j$ onto X_i. Also let $\pi_{\widehat{n}}$ be the projection of $\prod_{j=1}^n X_j$ onto $\prod_{j=1}^{n-1} X_j$ and for $k = 1, \ldots, n-1$ let τ_k be the projection of $\prod_{j=1}^{n-1} X_j$ onto X_k.

We know by definition that $\bigotimes_{j=1}^n \mathcal{M}_j$ is generated by the collection of sets:

$$A_1 \coloneqq \{\pi_j^{-1}(E) : j = 1, \ldots, n \text{ and } E \in \mathcal{M}_j\}.$$

Meanwhile, by the proposition at the top of page 13 of my latex math 240a notes, since $\{\tau_k^{-1}(E) : k = 1, \ldots, n-1 \text{ and } E \in \mathcal{M}_k\}$ is a base for $\bigotimes_{j=1}^{n-1} \mathcal{M}_j$, we have that $(\bigotimes_{j=1}^{n-1} \mathcal{M}_j) \otimes \mathcal{M}_3$ is generated by the collection of sets:

$$A_2 \coloneqq \{\pi_{\widehat{n}}^{-1}(\tau_k^{-1}(E)) : k = 1, \ldots, n-1 \text{ and } E \in \mathcal{M}_k\} \cup \{\pi_n^{-1}(E) : E \in \mathcal{M}_3\}.$$

But now if $E \in \mathcal{M}_1$, we know that:

$$\pi_{\widehat{n}}^{-1}(\tau_1^{-1}(E)) = \pi_{\widehat{n}}^{-1}(E \times X_2 \times \cdots \times X_{n-1}) = E \times X_2 \times \cdots \times X_{n-1} \times X_n.$$

Repeating this reasoning for all $2 \leq k \leq n-1$ we can in fact see that $A_1 = A_2$. This shows that both of the sigma algebras in the problem are equal.

- By nearly identical reasoning, we have that $\bigotimes_{j=1}^n \mathcal{M}_j = \mathcal{M}_1 \otimes (\bigotimes_{j=2}^n \mathcal{M}_j)$. And note that this is enough to show that the $\bigotimes$ operation is associative and we get the same result no matter how we use parentheses to group together the $\mathcal{M}_j$ (see the top of page 113 of my math journal).

- If all the μ_j are σ-finite, then $\mu_1 \times \cdots \times \mu_n = (\mu_1 \times \cdots \times \mu_{n-1}) \times \mu_n$.

  Proof:
  If all the μ_j are σ-finite, then we know that $\mu_1 \times \cdots \times \mu_n$ is the unique measure on $\bigotimes_{j=1}^n \mathcal{M}_j$ satisfying that for any rectangle $E_1 \times \cdots \times E_n$ where $E_k \in \mathcal{M}_k$ for all $k = 1, \ldots, n$:
  $$\mu_1 \times \cdots \times \mu_n(E_1 \times \cdots \times E_n) = \mu_1(E_1) \cdots \mu_n(E_n)$$

  Meanwhile, we also have that $\mu_1 \times \cdots \times \mu_{n-1}$ is a measure on $\bigotimes_{j=1}^{n-1} \mathcal{M}_j$ satisfying that if $E_k \in \mathcal{M}_k$ for all $k = 1, \ldots, n-1$, then:
  $$\mu_1 \times \cdots \times \mu_{n-1}(E_1 \times \cdots \times E_n) = \mu_1(E_1) \cdots \mu_{n-1}(E_{n-1})$$

  Furthermore, $(\mu_1 \times \cdots \times \mu_{n-1}) \times \mu_n$ is a measure on $\bigotimes_{j=1}^n$ satisfying that if $A \in \bigotimes_{j=1}^{n-1} \mathcal{M}_j$ and $E_n \in \mathcal{M}_n$, then:
  $$((\mu_1 \times \cdots \times \mu_{n-1}) \times \mu_n)(A \times E_n) = \mu_1 \times \cdots \times \mu_{n-1}\mu(A) \cdot \mu_n(E_n)$$

  So, if we let $E_k \in \mathcal{M}_k$ for all $k = 1, \ldots, n$, then we can clearly see that:
  $$\begin{aligned} &((\mu_1 \times \cdots \times \mu_{n-1}) \times \mu_n)(E_1 \times \cdots \times E_n) \\ &\qquad = \mu_1 \times \cdots \times \mu_{n-1}(E_1 \times \cdots \times E_{n-1}) \cdot \mu_n(E_n) \\ &\qquad = \mu_1(E_1) \cdots \mu_{n-1}(E_{n-1})\mu_n(E_n) \end{aligned}$$

  This shows that both measures in the problem have the same property which uniquely determines $\mu_1 \times \cdots \times \mu_n$. So, both measures must be equal.

- By nearly identical reasoning we have that $\mu_1 \times \cdots \times \mu_n = \mu_1 \times (\mu_2 \times \cdots \times \mu_n)$ if all the μ_j are σ-finite. Then similarly to two bullet points ago, this is enough to show that the $\times$ operation with respect to σ-finite measures is associative and we get the same result no matter how we use parentheses to group together the μ_j.

  I said yesterday that $\mathcal{B}_{\mathbb{R} \times \mathbb{R}^{n-1}} = \mathcal{B}_{\mathbb{R}^n}$. To prove this, first note that by the corollary on page 14 of my latex math 240a notes:
  $$\mathcal{B}_{\mathbb{R} \times \mathbb{R}^{n-1}} = \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}^{n-1}} = \mathcal{B}_{\mathbb{R}} \times (\textstyle\bigotimes_{j=1}^{n-1} \mathcal{B}_{\mathbb{R}}) \text{ and } \mathcal{B}_{\mathbb{R}^n} = \bigotimes_{j=1}^n \mathcal{B}_{\mathbb{R}}.$$

  Also, by the last exercise we have that $\mathcal{B}_{\mathbb{R}} \times (\bigotimes_{j=1}^{n-1} \mathcal{B}_{\mathbb{R}}) = \bigotimes_{j=1}^n \mathcal{B}_{\mathbb{R}}$. So hopefully this addresses my worry from yesterday.

Next order of business: in my paper math notes for math 240a, I copied down a sentence from Folland stating that the completion of $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, m^n)$ and the completion of $(\mathbb{R}^n, \bigotimes_{j=1}^n \mathcal{L}, m^n)$ are equal. I want to show that now by proving something slightly more general.

Claim: For each $j = 1, \ldots, n$ let $(X_j, \mathcal{M}_j, \mu_j)$ be a σ-finite measure space and let $(X_j, \overline{\mathcal{M}_j}, \overline{\mu_j})$ be the completion of $(X_j, \mathcal{M}_j, \mu_j)$.

- Suppose that $(X, \mathcal{N}, \nu) = (\prod_{j=1}^n X_j, \bigotimes_{j=1}^n \mathcal{M}_j, \prod_{j=1}^n \mu_j)$ and that $(X, \mathcal{N}', \nu') = (\prod_{j=1}^n X_j, \bigotimes_{j=1}^n \overline{\mathcal{M}_j}, \prod_{j=1}^n \overline{\mu_j})$. Then $\mathcal{N} \subseteq \mathcal{N}'$ and $\nu'(A) = \nu(A)$ for all $A \in \mathcal{N}$.

  Proof:
  Since $\mathcal{N}$ is a finite product of σ-algebras, we know that $\mathcal{N}$ is generated by the collection of sets:
  $$A_1 = \{E_1 \times \cdots \times E_n : E_k \in \mathcal{M}_k \text{ for all } k = 1, \ldots, n\}$$

  Similarly, we know that $\mathcal{N}'$ is generated by the collection of sets:
  $$A_2 = \{E_1 \times \cdots \times E_n : E_k \in \overline{\mathcal{M}_k} \text{ for all } k = 1, \ldots, n\}$$

  But clearly $A_1 \subseteq A_2$. So $\mathcal{N} \subseteq \mathcal{N}'$. To prove the other claim, consider the restriction $\nu'|_{\mathcal{N}}$. Since all the μ_j are σ-finite, we know that ν is the unique measure on $\mathcal{N}$ such that $\nu(E_1 \times \cdots \times E_n) = \mu_1(E_1) \cdots \mu_n(E_n)$ for all $E_1 \times \cdots \times E_n \in A^1$. However, we also have that:
  $$\mu_1(E_1) \cdots \mu_n(E_n) = \overline{\mu_1}(E_1) \cdots \overline{\mu_n}(E_n) = \nu'(E_1 \times \cdots \times E_n)$$

  So, $\nu = \nu'|_{\mathcal{N}}$. ■

- Let $(X, \overline{\mathcal{N}}, \overline{\nu})$ be the completion of $(X, \mathcal{N}, \nu)$. Then $\mathcal{N}' \subseteq \overline{\mathcal{N}}$ and $\overline{\nu}(A) = \nu'(A)$ for all $A \in \mathcal{N}'$.

  Proof:
  To start off, first note that ν' is the restriction of the outer measure induced by the function $\pi(E_1 \times \cdots \times E_n) = \overline{\mu_1}(E_1) \cdots \overline{\mu_n}(E_n)$ for all rectangles $E_1 \times \cdots \times E_n$ where $E_k \in \overline{\mathcal{M}_k}$ for all $k = 1, \ldots, n$.

  Technically Folland defined the product measure by taking the outer measure of a premeasure π defined on the collection $\mathcal{A}$ of all finite disjoint unions of rectangles. However, by just expanding each element of a sequence $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$ into an expression of disjoint rectangles, it's obvious how we can find a sequence $\{B_m\}_{m \in \mathbb{N}}$ consisting only of rectangles such that $\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{m \in \mathbb{N}} B_m$ and $\sum_{n \in \mathbb{N}} \pi(A_n) = \sum_{m \in \mathbb{N}} \pi(B_m)$.

  The only reason Folland uses the larger collection $\mathcal{A}$ is that $\mathcal{A}$ is an algebra of sets and so by defining a premeasure on $\mathcal{A}$ we can thus abstract away the outer measure definition and just apply the theorems from chapter 1 of his book. That said, the values that any additive measure takes on $\mathcal{A}$ are clearly determined entirely by the values that the measure takes on the collection of rectangles.

  But now note that if $E_k \in \overline{\mathcal{M}_k}$, then we can pick a set $F_k \in \mathcal{M}_k$ with $E_k \subseteq F_k$ and $\mu_k(F_k) = \overline{\mu_k}(F_k) = \overline{\mu_k}(E_k)$.

Doing this for all k, we have that for any rectangle $E = E_1 \times \cdots \times E_n$ with $E_k \in \overline{\mathcal{M}_k}$ for all k, there exists a rectangle $F = F_1 \times \cdots \times F_n$ with $E_k \subseteq F_k$ and $F_k \in \mathcal{M}_k$ for all k and which satisfies that $\nu'(E) = \nu'(F) = \nu(F)$.

As a consequence of all of the above reasoning, if $A \in \mathcal{N}'$ with $\nu'(A) < \infty$, then for each $k \in \mathbb{N}$ we may pick a sequence of rectangles $\{A_m^{(k)}\}_{m\in\mathbb{N}}$ in $\mathcal{N}'$ such that $A \subseteq \bigcup_{m\in\mathbb{N}} A_m^{(k)}$ and $\sum_{m\in\mathbb{N}} \nu'(A_m^{(k)}) < \nu'(A) + 1/k$. Next, for all k and n we may pick a rectangle $B_m^{(k)} \in \mathcal{N}$ such that $A_m^{(k)} \subseteq B_m^{(k)}$ and $\nu'(B_m^{(k)}) = \nu'(A_m^{(k)})$. And now, if we set $B := \bigcap_{k\in\mathbb{N}} \left(\bigcup_{m\in\mathbb{N}} B_m^{(k)}\right)$, we will know that $A \subseteq B$, $B \in \mathcal{N}$, and $\nu(B) = \nu'(B) = \nu'(A)$. Also note that clearly $\nu'(B - A) = 0$.

And this at long last leads us to the following construction for any $A \in \mathcal{N}'$:

> Let $\{E_i\}_{i\in\mathbb{N}} \subseteq \mathcal{N}$ be a partition of X consisting of sets with finite measure. Then for each i we can pick a set $C_i \in \mathcal{N}$ such that $C_i \supseteq E_i - A$ and $\nu'(C_i - (E_i - A)) = 0$. And in turn, if we let $B_i := E_i - C_i$, we will have that $B_i \in \mathcal{N}$, $B_i \subseteq E_i \cap A$, and $\nu'((A \cap E_i) - B_i) = 0$.
>
> Meanwhile, for each i we may also pick $D_i \in \mathcal{N}$ such that $D_i \supseteq A_i \cap E_i$ and $\nu'(D_i - (A_i \cap E_i)) = 0$. Now it's clear that:
> $$\nu(D_i - B_i) = \nu'(D_i - (A \cap E_i)) + \nu'((A \cap E_i) - B_i) = 0$$
>
> Finally, set $B = \bigcup_{i\in\mathbb{N}} B_i$ and $D = \bigcup_{i\in\mathbb{N}} D$. Then B and D are in $\mathcal{N}$, $B \subseteq A \subseteq D$, and $\nu(D - B) = 0$.
>
> Now $A = B \cup (A - B)$ where $B \in \mathcal{N}$ and $A - B \subseteq D - B$ with $D - B$ being a ν-null set in $\mathcal{N}$. From that it's clear that $A \in \overline{\mathcal{N}}$. Also, since we have that $0 \leq \nu'(A - B) \leq \nu'(D - B) = \nu(D - B) = 0$, we know:
> $$\overline{\nu}(A) = \nu(B) = \nu'(B) = \nu'(A). \blacksquare$$

Note from 9/4/2025: Frick I thought of a way cleaner proof for this. I'll write it below.

Suppose π_k is the projection of X onto X_k and that $A \in \overline{\mathcal{M}_k}$. Then $A = E \cup F$ where $E \in \mathcal{M}_k$ and $F \subseteq N$ where $N \in \mathcal{M}_k$ and $\mu_k(N) = 0$. It now follows that $\pi_k^{-1}(A) = \pi_k^{-1}(E) \cup \pi_k^{-1}(F)$ where $\pi_k^{-1}(E) \in \mathcal{N}$ and $\pi_k^{-1}(F) \subseteq \pi_k^{-1}(N)$ with $\nu(\pi_k^{-1}(N)) = 0$. Hence, $\pi_k^{-1}(A) \in \overline{\mathcal{N}}$.

Now since the collection $\mathcal{A} = \{\pi_k^{-1}(A) : k = 1, \ldots, n \text{ and } A \in \overline{\mathcal{M}_k}\}$ is a base for $\mathcal{N}'$ and $\mathcal{A}$ is a subset of the σ-algebra $\overline{\mathcal{N}}$, we must have that $\mathcal{N}' \subseteq \overline{\mathcal{N}}$.

Finally, consider that for any $A \in \mathcal{N}'$, if we write $A = E \cup F$ where $E \in \mathcal{N}$ and $F \subseteq N$ with $N \in \mathcal{N}$ and $\nu(N) = 0$, then we know that $\nu'(A - E) = 0$. Hence, $\overline{\nu}(A) = \nu(E)$ and $\nu'(A) = \nu'(E) + \nu'(A - E) = \nu(E) + 0$. So $\overline{\nu}(A) = \nu'(A)$.

Now before continuing, we need the following lemmas which I'm shocked I didn't prove during homeworks for math 240a. (To be clear, I thought I had proved this stuff at some point and I've used all of these things in proofs before. But now that I'm looking, I can't actually find a proof written down anywhere in my notes for this stuff.)

Lemma 1: Suppose $(X, \mathcal{M}, \mu)$ and $(X, \mathcal{N}, \nu)$ are measure spaces such that $\mathcal{M} \subseteq \mathcal{N}$ and $\nu|_{\mathcal{M}} = \mu$. Then if $(X, \overline{\mathcal{M}}, \overline{\mu})$ and $(X, \overline{\mathcal{N}}, \overline{\nu})$ are the completions of $(X, \mathcal{M}, \mu)$ and $(X, \mathcal{N}, \nu)$ respectively, we have that $\overline{\mathcal{M}} \subseteq \overline{\mathcal{N}}$ and $\overline{\nu}|_{\overline{\mathcal{M}}} = \overline{\mu}$.

Proof:
Suppose $A \in \overline{\mathcal{M}}$ and let E, F, N be sets satisfying that $A = E \cup F$; $F \subseteq N$; $E, N \in \mathcal{M}$; and $\mu(N) = 0$. Then we also have that $E, N \in \mathcal{N}$ and that $\nu(N) = \mu(N) = 0$. So, it's clear $A \in \overline{\mathcal{N}}$ and $\overline{\mu}(A) = \mu(E) = \nu(E) = \overline{\nu}(A)$.
■

Lemma 2: Suppose $(X, \mathcal{M}, \mu)$ is a complete measure space and let $(X, \overline{\mathcal{M}}, \overline{\mu})$ be the completion of $(X, \mathcal{M}, \mu)$. Then $\mathcal{M} = \overline{\mathcal{M}}$ and $\mu = \overline{\mu}$.

Proof:
If $A \in \overline{\mathcal{M}}$, then there exists sets E, F, N satisfying that $A = E \cup F$; $F \subseteq N$; $E, N \in \mathcal{M}$; and $\mu(N) = 0$. But now since μ is complete, we know that $F \in \mathcal{M}$. So, $A = E \cup F$ is also in $\mathcal{M}$, meaning $\overline{\mathcal{M}} \subseteq \mathcal{M}$. And since the other inclusion is obvious, we know that $\overline{\mathcal{M}} = \mathcal{M}$ Also, we thus have that $\overline{\mu} = \overline{\mu}|_{\mathcal{M}} = \mu$. ■

For the sake of nicer notation in the next corollary, let us refer to a given measure space $(X, \mathcal{M}, \mu)$ using a single symbol such as A. Also let $c(A)$ denote the completion of A. And if $B = (Y, \mathcal{N}, \nu)$ is another measure space, let us write $A \subseteq B$ to mean that $X = Y$, $\mathcal{M} \subseteq \mathcal{N}$, and $\nu|_{\mathcal{M}} = \mu$.

Theorem: If A and B are measures spaces such that $A \subseteq B \subseteq c(A)$, then $c(A) = c(B)$.

Proof:
We know by lemma 1 that $A \subseteq B \subseteq c(A)$ implies that $c(A) \subseteq c(B) \subseteq c(c(A))$. However, by lemma 2 we know that $c(c(A)) = c(A)$. So, $c(A) \subseteq c(B) \subseteq c(A)$. This is the same as saying that $c(A) = c(B)$. ■

Returning to what we were doing before the prior tangent:

- Let $(X, \overline{\mathcal{N}'}, \overline{\nu'})$ be the completion of $(X, \mathcal{N}', \nu')$. Then $(X, \overline{\mathcal{N}'}, \overline{\nu'}) = (X, \overline{\mathcal{N}}, \overline{\nu})$.

  Proof:
  We'll continue using the notation in the theorem I showed right above. Let $A = (X, \mathcal{N}, \nu)$ and $B = (X, \mathcal{N}', \nu')$. Then note that $c(A) = (X, \overline{\mathcal{N}}, \overline{\nu})$ and $c(B) = (X, \overline{\mathcal{N}'}, \overline{\nu'})$.

In the first bullet point of our claim, we showed that $A \subseteq B$. And in the second bullet point of our claim, we showed that $B \subseteq c(A)$. Thus, by applying our theorem right above, we know that $c(B) = c(A)$. ■

As a side note before I clock out today, suppose A, B, C are σ-finite measure spaces and let us write AB to denote the product measure space of A and B. Then we can use the theorems I covered today to say things like:

$$c(c(AB)C) = c(c(c(AB))c(C)) = c(c(AB)c(C)) = c((AB)C) = c(ABC)$$

You can have more fun playing around with that! The rules of the game are:

- $c(A_1 \cdots A_n) = c(c(A_1) \cdots c(A_n))$,
- multiplication is associative (i.e. $A_1A_2A_3 = (A_1A_2)A_3 = A_1(A_2A_3)$).

9/4/2025

One more thing I want to do before finally moving onto new content is finally prove the reformulation of Fubini's theorem but for complete measures. The theorem statement is in my paper notes but I'll also write it below. Also as a side note: if $f(x, y)$ is a function from $X \times Y$ to $\mathbb{C}$ (or $\overline{\mathbb{R}}$), then Folland denotes $f_x := f(x, \cdot)$ and $f^y := f(\cdot, y)$. Additionally, if $E \subseteq X \times Y$, then Folland denotes:

$E_x := \{y \in Y : (x, y) \in E\}$ and $E^y := \{x \in X : (x, y) \in E\}$.

<u>Theorem 2.39: (The Fubini-Tonelli Theorem for Complete Measures)</u>

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be complete σ-finite measure spaces, and let $(X \times Y, \mathcal{L}, \lambda)$ be the completion of $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$. If f is $\mathcal{L}$-measurable and either (a) $f \geq 0$ or (b) $f \in L^1(\lambda)$, then f_x is $\mathcal{N}$-measurable for a.e. x and f^y is $\mathcal{M}$-measurable for a.e. y. Also, in the case that (b) is true, f_x and f^y are integrable for a.e. x and a.e. y.

Moreover, $x \mapsto \int f_x \mathrm{d}\nu$ and $y \mapsto \int f^y \mathrm{d}\mu$ are measurable. And in the case that (b) is true they are also integrable.

And finally, in the case of either (a) or (b), we have that:

$$\int f \mathrm{d}\lambda = \iint f(x, y) \mathrm{d}\mu(x) \mathrm{d}\nu(y) = \iint f(x, y) \mathrm{d}\nu(y) \mathrm{d}\nu(x)$$

Exercise 2.49: Prove the above theorem.

<u>Lemma 1:</u> If $E \in \mathcal{M} \times \mathcal{N}$ and $\mu \times \nu(E) = 0$, then $\nu(E_x) = \mu(E^y) = 0$ for a.e. x and a.e. y.

Proof:
Consider the function $f(x, y) = \chi_E(x, y)$. By the Fubini-Tonelli theorem I already proved in my math 240a notes, we know that:

$$0 = \mu \times \nu(E) = \int f \mathrm{d}(\mu \times \nu) = \int (\int f_x \mathrm{d}\nu(y)) \mathrm{d}\nu(x) = \int (\int f^y \mathrm{d}\mu(x)) \mathrm{d}\nu(y)$$

So, $\int f_x \mathrm{d}\nu(y) = \nu(E_x) = 0$ for a.e. x and $\int f^y \mathrm{d}\mu(x) = \mu(E^y) = 0$ for a.e. y.

Lemma 2: If f is $\mathcal{L}$-measurable and $f = 0$ λ-a.e., then f_x and f^y are integrable for a.e. x and a.e. y, and $\int f_x \mathrm{d}\nu = \int f^y \mathrm{d}\mu = 0$ for a.e. x and a.e. y.

Proof:

Let $A = \{(x,y) \in X \times Y : f(x,y) \neq 0\}$. Then since $(X \times Y, \mathcal{L}, \lambda)$ is the completion of $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$, we know there exists a set $E \in \mathcal{M} \otimes \mathcal{N}$ such that $\mu \times \nu(E) = 0$ and $A \subseteq E$.

By our last lemma, $\nu(E_x) = 0$ for a.e. x and $\mu(E^y) = 0$ for a.e. y. Also, $f_x = \chi_{E_x}$ on $(E_x)^{\mathsf{C}}$ and $f^y = \chi_{E^y}$ on $(E^y)^{\mathsf{C}}$. Hence for a.e. y and a.e. x respectively, we have that $f_x = \chi_{E_x}$ ν-a.e. and $f^y = \chi_{E^y}$ μ-a.e. Because μ and ν are complete, we can in turn conclude by exercise 2.10 in my latex math 240a notes that f_x is ν-measurable and f^y is μ-measurable for a.e. x and a.e. y respectively.

Also, since $|f_x| = \chi_{E_x}$ on $(E_x)^{\mathsf{C}}$ and $|f^y| = \chi_{E^y}$ on $(E^y)^{\mathsf{C}}$, and both $|f_x|$ and $|f^y|$ are measurable, we know for a.e. x that $\int |f_x| \mathrm{d}\nu(y) = \int \chi_{E_x} \mathrm{d}\nu(y) = \nu(E_x) = 0$. And similarly for a.e. y we know that $\int |f^y| \mathrm{d}\mu(x) = \int \chi_{E^y} \mathrm{d}\mu(x) = \mu(E^y) = 0$. Hence, we know that f_x and f^y are integrable for a.e. x and a.e. y, and that $\int f_x \mathrm{d}\nu(y) = 0$ and $\int f^y \mathrm{d}\mu(x) = 0$ for a.e. x and a.e. y.

Now we get to the main part of the proof of our theorem. Let f be any $\mathcal{L}$-measurable function. Then from the proposition on page 46 of my latex math 240a notes, we know that there exists a function g that is $(\mathcal{M} \otimes \mathcal{N})$-measurable such that $f = g$ λ-a.e. So, consider the identity $f = (f - g) + g$.

We have that $f - g = 0$ λ-a.e. So by lemma 2, we know for a.e. x that $(f-g)_x$ is ν-measurable with $\int (f-g)_x \mathrm{d}\nu = 0$. And similarly, we know for a.e. y that $(f-g)^y$ is μ-measurable with $\int (f-g)^y \mathrm{d}\mu = 0$.

Next, if $f \geq 0$ then we can without loss of generality take $g \geq 0$. And then by the Fubini-Tonelli theorem for noncompleted product measures, we know that g_x is ν-measurable for a.e. x; g^y is μ-measurable for a.e. y; $\int g_x \mathrm{d}\nu$ and $\int g^y \mathrm{d}\mu$ are measurable and nonnegative; and $\int g \mathrm{d}(\mu \times \nu) = \iint g_x \mathrm{d}\nu \mathrm{d}\mu = \iint g^y \mathrm{d}\mu \mathrm{d}\nu$.

It then follows that $f_x = (f-g)_x + g_x$ is ν-measurable for a.e. x; that $f^y = (f-g)^y + g^y$ is μ-measurable for a.e. y; and that $\int f_x \mathrm{d}\nu = \int (f-g)_x \mathrm{d}\nu + \int g_x \mathrm{d}\nu$ and $\int f^y \mathrm{d}\mu = \int (f-g)^y \mathrm{d}\mu + \int g \mathrm{d}\mu$ are measurable.

Additionally, note that $\int f \mathrm{d}\lambda = \int g \mathrm{d}\lambda = \int g \mathrm{d}(\mu \times \nu)$ since $f = g$ λ-a.e. And on the other hand we have that:

- $$\begin{aligned} \iint f_x \mathrm{d}\nu \mathrm{d}\mu = \iint (f-g)_x + g_x \mathrm{d}\nu \mathrm{d}\mu &= \int (\int (f-g)_x \mathrm{d}\nu) + (\int g_x \mathrm{d}\nu) \mathrm{d}\mu \\ &= \int 0 + (\int g_x \mathrm{d}\nu) \mathrm{d}\mu \\ &= \iint g_x \mathrm{d}\nu \mathrm{d}\mu = \int g \mathrm{d}(\mu \times \nu) \end{aligned}$$
- $$\begin{aligned} \iint f_y \mathrm{d}\mu \mathrm{d}\nu = \iint (f-g)^y + g^y \mathrm{d}\mu \mathrm{d}\nu &= \int (\int (f-g)^y \mathrm{d}\mu) + (\int g^y \mathrm{d}\mu) \mathrm{d}\nu \\ &= \int 0 + (\int g^y \mathrm{d}\mu) \mathrm{d}\nu \\ &= \iint g^y \mathrm{d}\mu \mathrm{d}\nu = \int g \mathrm{d}(\mu \times \nu) \end{aligned}$$

This proves the theorem for when $f \geq 0$.

The case where $f \in L^1(\lambda)$ is really similar and I'm bored. So I'm going to end the proof here. ■

Something I want to add before moving on is that we can easily extend the Fubini-Tonelli theorem to products of more than two measures by applying the previous things we proved about product measures in the past three days.

For example, if $(X_j, \mathcal{M}_j, \mu_j)$ is a σ-finite measure space for each $j = 1, 2, 3$, then by considering $\mu_1 \times \mu_2 \times \mu_3 = (\mu_1 \times \mu_2) \times \mu_3$ we can say that:

$$\int f \mathrm{d}(\mu_1 \times \mu_2 \times \mu_3) = \iint f \mathrm{d}(\mu_1 \times \mu_2) \mathrm{d}\mu_3 = \iint f \mathrm{d}\mu_3 \mathrm{d}(\mu_1 \times \mu_2).$$

Also, going further we have that:

- $\int f \mathrm{d}(\mu_1 \times \mu_2) = \iint f \mathrm{d}\mu_1 \mathrm{d}\mu_2 = \iint f \mathrm{d}\mu_2 \mathrm{d}\mu_1$
- $\int (\int f \mathrm{d}\mu_3) \mathrm{d}(\mu_1 \times \mu_2) = \iint (\int f \mathrm{d}\mu_3) \mathrm{d}\mu_1 \mathrm{d}\mu_2 = \iint (\int f \mathrm{d}\mu_3) \mathrm{d}\mu_2 \mathrm{d}\mu_1$

Hence we have already shown that:

$$\begin{aligned}\int f \mathrm{d}(\mu_1 \times \mu_2 \times \mu_3) &= \iiint f \mathrm{d}\mu_1 \mathrm{d}\mu_2 \mathrm{d}\mu_3 = \iiint f \mathrm{d}\mu_2 \mathrm{d}\mu_1 \mathrm{d}\mu_3 \\ &= \iiint f \mathrm{d}\mu_3 \mathrm{d}\mu_1 \mathrm{d}\mu_2 = \iiint f \mathrm{d}\mu_3 \mathrm{d}\mu_2 \mathrm{d}\mu_1\end{aligned}$$

Meanwhile, if we identify $\mu_1 \times \mu_2 \times \mu_3$ with $\mu_1 \times (\mu_2 \times \mu_3)$, we can show that:

$$\int f \mathrm{d}(\mu_1 \times \mu_2 \times \mu_3) = \iiint f \mathrm{d}\mu_1 \mathrm{d}\mu_3 \mathrm{d}\mu_2 = \iiint f \mathrm{d}\mu_2 \mathrm{d}\mu_3 \mathrm{d}\mu_1.$$

At long last I shall start making progress in Folland again.

Products of Radon Measures:

Let X and Y be LCH spaces and let π_X and π_Y denote the projections of $X \times Y$ onto X and Y respectively. I will note the following exercise I did for my math 240b homework:

Exercise 4.59: The product of finitely many locally compact spaces is locally compact.

Lemma 1: If $E_\alpha \subseteq X_\alpha$ for all $\alpha \in A$, then the product of the relative topologies of E_α on $\prod_{\alpha \in A} E_\alpha$ (denoted $\mathcal{T}$) is equal to the relative topology of $\prod_{\alpha \in A} E_\alpha$ induced by the product topology on $\prod_{\alpha \in A} X_\alpha$ (denoted $\mathcal{T}''$).

($\Longrightarrow$)

Suppose $V \in \mathcal{T}$ and $x \in V$. Then there are sets V_α all open relative to E_α with all but finitely many V_α equal to E_α and $x \in \prod_{\alpha \in A} V_\alpha \subseteq V$. Now if $V_\alpha = E_\alpha$, set $V'_\alpha = X_\alpha$. Otherwise, let V'_α just be any open set in X_α such that $V_\alpha = V'_\alpha \cap E_\alpha$. That way, $U' := \prod_{\alpha \in A} V'_\alpha$ is open in $\prod_{\alpha \in A} X_\alpha$. And finally, $U := U' \cap \prod_{\alpha \in A} E_\alpha \in \mathcal{T}'$ with $x \in U \subseteq V$. This proves that V is a union of sets in $\mathcal{T}'$.

($\Longleftarrow$)

Suppose $U \in \mathcal{T}'$ and $x \in U$. Then there exists U' in $\prod_{\alpha \in A} X_\alpha$ which is open and satisfies that $U = U' \cap \prod_{\alpha \in A} E_\alpha$. Next, for all $\alpha \in A$ there are open sets V'_α with all but finitely many V'_α equal to X_α and $x \in \prod_{\alpha \in A} V'_\alpha \subseteq U'$. Finally, for all α set $V_\alpha = V'_\alpha \cap E_\alpha$. Then all V_α are open in E_α relative to X_α and all but finitely many V_α are equal to E_α. Also $x \in V := \prod_{\alpha \in A} V_\alpha$. So $V \in \mathcal{T}$ and $x \in V \subseteq U$. This proves that U is a union of sets in $\mathcal{T}$.

As a corollary to the above lemma and because of Tychonoff's theorem, if $N_\alpha \subseteq X_\alpha$ is compact for all $\alpha \in A$, then $\prod_{\alpha \in A} N_\alpha$ is compact in $\prod_{\alpha \in A} X_\alpha$.

Now we finally use the fact A is finite. Note that because A is finite, if $x \in \prod_{\alpha \in A} X_\alpha$ and we are given the neighborhoods N_α of $\pi_\alpha(x)$ for all $\alpha \in A$, then $\prod_{\alpha \in A} N_\alpha$ is a neighborhood of x. Therefore, to get a compact neighborhood of $x \in \prod_{\alpha \in A} X_\alpha$, just take a product of compact neighborhoods of $\pi_\alpha(x)$ for each $\alpha \in A$.

Also, we know from proposition 4.10 in my math 240b notes that arbitrary products of Hausdorff spaces are Hausdorff. Hence, putting the above exercise and that proposition together, we know that a finite topological product of LCH spaces is an LCH space.

Theorem 7.20:

(a) $\mathcal{B}_X \otimes \mathcal{B}_Y \subseteq \mathcal{B}_{X \times Y}$.

Proof:
This is just an obvious generalization of the proposition at the bottom of page 13 of my latex math 240a notes.

(b) If X and Y are second countable, than $\mathcal{B}_X \otimes \mathcal{B}_Y = \mathcal{B}_{X \times Y}$.

Proof:
It suffices to show that $\mathcal{B}_{X \times Y} \subseteq \mathcal{B}_X \times \mathcal{B}_Y$. To do this, let $\mathcal{B}_x$ be a countable basis for X and $\mathcal{B}_y$ be a countable basis for Y. Then $\mathcal{B} = \{B_x \times B_y : B_x \in \mathcal{B}_x$ and $\mathcal{B}_y\}$ is a countable basis for the product topology on $X \times Y$.

It's clear that $\mathcal{B}$ is countable and contains only open sets. Meanwhile, in order to prove that $\mathcal{B}$ is a basis, suppose $(x, y) \in U \subseteq X \times Y$ where U is open. Then we know there is a set $V_1 \times V_2 \subseteq U$ such that V_1 is open in X; V_2 is open in Y; and $x \in V_1$ and $y \in V_2$. Next, there are sets $B_1 \in \mathcal{B}_x$ and $B_2 \in \mathcal{B}_y$ such that $x \in B_1 \subseteq V_1$ and $y \in B_2 \subseteq V_2$. Now it's clear that $B_1 \times B_2 \in \mathcal{B}$ and $(x, y) \in B_1 \times B_2 \subseteq V_1 \times V_2 \subseteq U$.

Now since every open set in $X \times Y$ is a countable union of sets in $\mathcal{B}$, we know that $\mathcal{T} \subseteq \mathcal{M}(\mathcal{B})$ where $\mathcal{T}$ is the topology on $X \times Y$. It follows that $\mathcal{B}_{X \times Y} \subseteq \mathcal{M}(\mathcal{B})$. But at the same time note that $\mathcal{B}$ is a subset of the entire collection of products of open sets, and we know by the proposition at the top of page 13 of my latex math 240a notes that the latter collection generates $\mathcal{B}_X \otimes \mathcal{B}_Y$. So, $\mathcal{B}_{X \times Y} \subseteq \mathcal{M}(\mathcal{B}) \subseteq \mathcal{B}_X \times \mathcal{B}_Y$. ■

(c) If X and Y are second countable and μ and ν are Radon measures on X and Y, then $\mu \times \nu$ is a Radon measure on $X \times Y$.

Proof:
Since we showed in part (b) that $\mu \times \nu$ is a Borel measure on $X \times Y$ and that $X \times Y$ is second countable, we know by Folland theorem 7.8 (see my math 240c notes) that it suffices to show that $\mu \times \nu$ is finite on any compact subset K of $X \times Y$ in order to show that $\mu \times \nu$ is regular and thus Radon. But fortunately since π_X and π_Y are continuous, we know that $\pi_X(K)$ and $\pi_Y(K)$ are compact subsets in X and Y respectively. And since both μ and ν are Radon, we know that $\mu(\pi_X(K)) < \infty$ and $\nu(\pi_Y(K)) < \infty$. Also, $K \subseteq \pi_1(K) \times \pi_2(K)$. So:

$$(\mu \times \nu)(K) \leq (\mu \times \nu)(\pi_1(K) \times \pi_2(K)) = \mu(\pi_1(K))\nu(\pi_2(K)) < \infty. \blacksquare$$

Given functions $g : X \to \mathbb{C}$ and $h : X \to \mathbb{C}$, we define $g \otimes h(x, y) := g(x)h(y)$.

Proposition 7.21: Let $\mathcal{P}$ be the vector space spanned by the functions $g \otimes h$ with $g \in C_c(X)$ and $h \in C_c(Y)$. Then $\mathcal{P}$ is dense in $C_c(X \times Y)$ in the uniform norm. More precisely, given $f \in C_c(X \times Y)$, $\varepsilon > 0$, and precompact open sets $U \subseteq X$ and $V \subseteq Y$ containing $\pi_X(\mathrm{supp}(f))$ and $\pi_Y(\mathrm{supp}(f))$, there exists $F \in \mathcal{P}$ such that $\|F - f\|_u < \varepsilon$ and $\mathrm{supp}(F) \subseteq U \times V$.

Proof:
$\overline{U} \times \overline{V}$ is a compact Hausdorff space. Also, we clearly have that the linear span $\mathcal{A}$ of $\{g \otimes h : g \in C(\overline{U}) \text{ and } h \in C(\overline{V})\}$ is a subalgebra of $C(\overline{U} \times \overline{V})$ which contains all the constant functions and is closed under complex conjugation. We can also see that $\mathcal{A}$ separates points as follows:

Suppose (x_1, y_1) and (x_2, y_2) be distinct points in $\overline{U} \times \overline{V}$. Then we know that either $x_1 \neq x_2$ or $y_1 \neq y_2$. I'll focus on the case that $x_1 \neq x_2$ since the other case is basically identical. We know that $\overline{U}$ is normal (since all compact Hausdorff spaces are normal). So, by Urysohn's lemma there exists a function $g \in C(\overline{U})$ such that $g(x_1) = 1$ and $g(x_2) = 0$. And now if we just set $h(y) = 1 \in C(\overline{V})$, we have that $g \otimes h \in C(\overline{U} \times \overline{V})$ with $g \otimes h(x_1, y_1) = 1$ and $g \otimes h(x_2, y_2) = 0$.

Thus by the Stone-Weierstrass theorem, we know that $\mathcal{A}$ is dense in $C(\overline{U} \times \overline{V})$. In particular, this means that there is an element $G \in \mathcal{A}$ with $\sup_{(x,y)\in\overline{U}\times\overline{V}} |G - f| < \varepsilon$.

Meanwhile, by Urysohn's lemma we know there exists functions $\phi \in C_c(U, [0, 1])$ and $\psi \in C_c(V, [0, 1])$ such that $\phi = 1$ on $\pi_X(\mathrm{supp}(f))$ and $\psi = 1$ on $\pi_Y(\mathrm{supp}(f))$. Thus if we define $F = (\phi \otimes \psi)G$ on $\overline{U} \times \overline{V}$ and $F = 0$ elsewhere, we have that $F \in \mathcal{P}$, $\|F - f\|_u < \varepsilon$, and $\mathrm{supp}(F) \subseteq U \times V$. ■

Proposition 7.22: Every $f \in C_c(X \times Y)$ is $(\mathcal{B}_X \otimes \mathcal{B}_Y)$-measurable. Moreover, if μ and ν are Radon measures on X and Y, then $C_c(X \times Y) \subseteq L^1(\mu \times \nu)$ and:

$$\int f d(\mu \times \nu) = \iint f d\mu d\nu = \iint f d\nu d\mu \text{ for all } f \in C_c(X \times Y).$$

Proof:
If $g \in C_c(X)$ and $h \in C_c(Y)$, then $g \otimes h = (g \circ \pi_X)(h \circ \pi_Y)$ is $(\mathcal{B}_X \otimes \mathcal{B}_Y)$-measurable. Hence since products, sums, and pointwise limits of measurable function are measurable, we know by the last proposition that every $f \in C_c(X, Y)$ is $(\mathcal{B}_X, \mathcal{B}_Y)$-measurable.

Also, every $f \in C_c(X, Y)$ is bounded and supported in a set of finite $(\mu \times \nu)$-measure.

Specifically consider the set $K := \pi_X(\mathrm{supp}(f)) \times \pi_Y(\mathrm{supp}(f)) \in X \times Y$. Then $\mathrm{supp}(f) \subseteq K$, K is compact, and $\mu \times \nu(K) = \mu(\pi_X(\mathrm{supp}(f)))\nu(\pi_Y(\mathrm{supp}(f)))$. Also, the latter expression is finite since $\pi_X(\mathrm{supp}(f))$ and $\pi_Y(\mathrm{supp}(f))$ are compact and μ and ν are Radon.

This proves that every $f \in C_c(X, Y)$ is also in $L^1(\mu \times \nu)$. Finally, even if μ and ν are not σ-finite, we can still show that Fubini's theorem holds. Specifically, consider letting U and V be precompact open sets in X and Y respectively such that $\pi_X^{-1}(\mathrm{supp}(f)) \subseteq U$ and $\pi_Y^{-1}(\mathrm{supp}(f)) \subseteq V$. Then set $\mu'(E) := \mu(E \cap U)$ and $\nu'(E) := \nu(E \cap V)$.

Now it's clear that:

- $\int f\mathrm{d}\mu' = \int f\chi_U\mathrm{d}\mu = \int f\mathrm{d}\mu$,
- $\int(\int f\mathrm{d}\mu')\mathrm{d}\nu' = \int(\int f\mathrm{d}\mu)\mathrm{d}\nu' = \int(\int f\mathrm{d}\mu)\chi_V\mathrm{d}\nu = \int(\int f\mathrm{d}\mu)\mathrm{d}\nu$,

- $\int f\mathrm{d}\nu' = \int f\chi_V\mathrm{d}\nu = \int f\mathrm{d}\nu$,
- $\int(\int f\mathrm{d}\nu')\mathrm{d}\mu' = \int(\int f\mathrm{d}\nu)\mathrm{d}\mu' = \int(\int f\mathrm{d}\nu)\chi_U\mathrm{d}\mu = \int(\int f\mathrm{d}\nu)\mathrm{d}\mu$.

Also, we can see as follows that $\int f\mathrm{d}(\mu\times\nu) = \int f\mathrm{d}(\mu'\times\nu')$...

Note that if $g\in L^1(\mu)$ and $h\in L^1(\nu)$, then we automatically have that $g\in L^1(\mu')$, $h\in L^1(\nu')$, $\int g\mathrm{d}\mu' = \int g\chi_U\mathrm{d}\mu$, and $\int h\mathrm{d}\nu' = \int h\chi_V\mathrm{d}\nu$. Additionally, by applying exercise 2.51 from my math 240a homework, we know that:

$$\int(g\chi_U)\otimes(h\chi_V)\mathrm{d}(\mu\times\nu) = (\int g\chi_U\mathrm{d}\mu)(\int h\chi_V\mathrm{d}\nu)$$
$$= (\int g\mathrm{d}\mu')(\int h\mathrm{d}\nu') = \int(g\otimes h)\mathrm{d}(\mu'\times\nu').$$

It then follows that if $\mathcal{P}$ is as in the last proposition, $F\in\mathcal{P}$, and $\mathrm{supp}(F)\subseteq U\times V$, then we know that $\int F\mathrm{d}(\mu'\times\nu') = \int F\mathrm{d}(\mu\times\nu)$. However, we also showed in the last proposition that we can find a sequence $(F_n)_{n\in\mathbb{N}}$ of such functions such that $F_n\to f$ uniformly. Since $\mu\times\nu(U\times V) = \mu'\times\nu'(U\times V) < \infty$, we can thus conclude via the dominated convergence theorem (using an upper bound of $f+\chi_{U\times V}$) that:

$$\int f\mathrm{d}(\mu\times\nu) = \lim_{n\to\infty}\int F_n\mathrm{d}(\mu\times\nu) = \lim_{n\to\infty}\int F_n\mathrm{d}(\mu'\times\nu') = \int f\mathrm{d}(\mu'\times\nu')$$

A consequence of this is that since $f\in L^1(\mu\times\nu)$, we now know that $f\in L^1(\mu'\times\nu')$. In turn, since both μ' and ν' are finite measures, we can apply the Fubini-Tonelli theorem to see that $\int f\mathrm{d}(\mu'\times\nu') = \iint f\mathrm{d}\mu'\mathrm{d}\nu' = \iint f\mathrm{d}\nu'\mathrm{d}\mu'$. And now we're done since we already showed earlier that $\int f\mathrm{d}(\mu'\times\nu') = \int f\mathrm{d}(\mu\times\nu)$, $\iint f\mathrm{d}\mu'\mathrm{d}\nu' = \iint f\mathrm{d}\mu\mathrm{d}\nu$, and $\iint f\mathrm{d}\nu'\mathrm{d}\mu' = \iint f\mathrm{d}\nu\mathrm{d}\mu$. ■

The significance of the last proposition is that we now know that if μ and ν are positive Radon measures on X and Y respectively, then $I(f) = \int f\mathrm{d}(\mu\times\nu)$ is a well-defined positive linear functional from $C_c(X\times Y)$ to $\mathbb{C}$. Hence, there exists a unique well-defined positive Radon measure (which we denote $\mu\mathbin{\widehat{\times}}\nu$) on $\mathcal{B}_{X\times Y}$ such that $\int f\mathrm{d}(\mu\mathbin{\widehat{\times}}\nu) = \int f(\mu\times\nu)$ for all $f\in C_c(X\times Y)$. We call $\mu\mathbin{\widehat{\times}}\nu$ the Radon product of μ and ν.

I'm going to stop taking notes on this section of Folland for the time being cause I don't think Radon products will be necessary for anything I want to study in Folland for a while. Next time, I will be studying some Fourier analysis from Folland. Also, I'm going to finally add in-pdf hyperlinks to this document.

9/7/2025

Let $M(\mathbb{R}^n)$ be the space of complex Borel Radon measures on $\mathbb{R}^n$. Then given any $\mu,\nu\in M(\mathbb{R}^n)$, we define $\mu\times\nu\in M(\mathbb{R}^n\times\mathbb{R}^n)$ by:

$$\mathrm{d}(\mu\times\nu)(x,y) \coloneqq \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}(x)\frac{\mathrm{d}\nu}{\mathrm{d}|\nu|}(y)\mathrm{d}(|\mu|\times|\nu|)(x,y)$$

It's clear that $\mu \times \nu$ is a well-defined Radon measure on $M(\mathbb{R}^n \times \mathbb{R}^n)$ by *proposition 7.16*, *theorem 7.20(c)*, and the *corollary to exercise 7.8*. Also, this definition doesn't conflict with our previous definition for product measures because if μ and ν are positive measures, then $|\mu| = \mu$ and $|\nu| = \nu$. So $\frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}(x)\frac{\mathrm{d}\nu}{\mathrm{d}|\nu|}(y) = 1$ and $\mathrm{d}(|\mu| \times |\nu|) = \mathrm{d}(\mu \times \nu)$.

Note: we need this new definition of product measures because we want to be able to take products of complex measures rather than just positive measures.

Also, this definition can be used to define products for complex measures in general (as opposed to just on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$).

Exercise 3.12: For $j = 1, 2$, let μ_j, ν_j be σ-finite positive measures on $(X_j, \mathcal{M}_j)$ such that $\nu_j \ll \mu_j$. Then $\nu_1 \times \nu_1 \ll \mu_1 \times \mu_2$ and:

$$\frac{\mathrm{d}(\nu_1 \times \nu_2)}{\mathrm{d}(\mu_1 \times \mu_2)}(x_1, x_2) = \frac{\mathrm{d}\nu_1}{\mathrm{d}\mu_1}(x_1)\frac{\mathrm{d}\nu_2}{\mathrm{d}\mu_2}(x_2)$$

Proof:
Note that for all $E \in \mathcal{M}_1 \otimes \mathcal{M}_2$, since $\frac{\mathrm{d}\nu_j}{\mathrm{d}\mu_j} \geq 0$ for $j = 1, 2$ and $\chi_E \geq 0$, we have by Tonelli's theorem that:

$$\begin{aligned}\nu_1 \times \nu_2(E) = \int \chi_E \mathrm{d}(\nu_1 \times \nu_2) &= \iint \chi_E(x_1, x_2)\mathrm{d}\nu_1(x_1)\mathrm{d}\nu_2(x_2) \\ &= \iint \chi_E(x_1, x_2)\tfrac{\mathrm{d}\nu_1}{\mathrm{d}\mu_1}(x_1)\mathrm{d}\mu_1(x_1)\tfrac{\mathrm{d}\nu_2}{\mathrm{d}\mu_2}(x_2)\mathrm{d}\mu_2(x_2) \\ &= \int_E \tfrac{\mathrm{d}\nu_1}{\mathrm{d}\mu_1}(x_1)\tfrac{\mathrm{d}\nu_2}{\mathrm{d}\mu_2}(x_2)\mathrm{d}(\mu_1 \times \mu_2)(x_1, x_2).\end{aligned}$$

This proves that $\nu_1 \times \nu_2 \ll \mu_1 \times \mu_2$ and that $\frac{\mathrm{d}(\nu_1 \times \nu_2)}{\mathrm{d}(\mu_1 \times \mu_2)}(x_1, x_2) = \frac{\mathrm{d}\nu_1}{\mathrm{d}\mu_1}(x_1)\frac{\mathrm{d}\nu_2}{\mathrm{d}\mu_2}(x_2)$ $(\mu_1 \times \mu_2)$-a.e. ■

Extension (not in Folland): For $j = 1, 2$ let μ_j be a σ-finite measure on $(X_j, \mathcal{M}_j)$ and ν_j be a complex measure on $(X_j, \mathcal{M}_j)$ such that $\nu_j \ll \mu_j$. Then $\nu_1 \times \nu_2 \ll \mu_1 \times \mu_2$ and:

$$\frac{\mathrm{d}(\nu_1 \times \nu_2)}{\mathrm{d}(\mu_1 \times \mu_2)}(x_1, x_2) = \frac{\mathrm{d}\nu_1}{\mathrm{d}\mu_1}(x_1)\frac{\mathrm{d}\nu_2}{\mathrm{d}\mu_2}(x_2)$$

Proof:
Starting off, we know that $\left|\frac{\mathrm{d}\nu_1}{\mathrm{d}|\nu_1|}(x_1)\frac{\mathrm{d}\nu_2}{\mathrm{d}|\nu_2|}(x_2)\right| = 1$ for $(|\nu_1| \times |\nu_2|)$-a.e. (x_1, x_2).

Thus since $|\nu_1|$ and $|\nu_2|$ are finite measures, we know $\frac{\mathrm{d}\nu_1}{\mathrm{d}|\nu_1|}(x_1)\frac{\mathrm{d}\nu_2}{\mathrm{d}|\nu_2|}(x_2) \in L^1(|\nu_1| \times |\nu_2|)$. Hence, we can apply Fubini's theorem to say that:

$$\begin{aligned}\nu_1 \times \nu_2(E) &= \int_E \tfrac{\mathrm{d}\nu_1}{\mathrm{d}|\nu_1|}(x_1)\tfrac{\mathrm{d}\nu_2}{\mathrm{d}|\nu_2|}(x_2)\mathrm{d}(|\nu_1| \times |\nu_2|)(x_1, x_2) \\ &= \iint \chi_E \tfrac{\mathrm{d}\nu_1}{\mathrm{d}|\nu_1|}(x_1)\tfrac{\mathrm{d}\nu_2}{\mathrm{d}|\nu_2|}(x_2)\mathrm{d}|\nu_1|(x_1)\mathrm{d}|\nu_2|(x_2)\end{aligned}$$

Next, since $\nu_j \ll \mu_j$ for $j = 1, 2$ we know that $|\nu_j| \ll \mu_j$ for all $j = 1, 2$. Thus for $j = 1, 2$ there exists $\frac{\mathrm{d}\nu_j}{\mathrm{d}|\nu_j|}$. And so we may write:

$$\iint \chi_E \frac{d\nu_1}{d|\nu_1|}(x_1)\frac{d\nu_2}{d|\nu_2|}(x_2)d|\nu_1|(x_1)d|\nu_2|(x_2)$$
$$= \iint \chi_E \frac{d\nu_1}{d|\nu_1|}(x_1)\frac{d|\nu_1|}{d\mu_1}(x_1)\cdot\frac{d\nu_2}{d|\nu_2|}(x_2)\frac{d|\nu_2|}{d\mu_2}(x_2)d\mu_1(x_1)d\mu_2(x_2)$$
$$= \iint \chi_E \frac{d\nu_1}{d\mu_1}(x_1)\cdot\frac{d\nu_2}{d\mu_2}(x_2)d\mu_1(x_1)d\mu_2(x_2)$$

Finally, $\frac{d\nu_j}{d\mu_j} \in L^1(\mu_j)$ for $j = 1, 2$ since $|\nu_j|$ is a finite measure. Thus by Tonelli's theorem we have that:

$$\int |\frac{d\nu_1}{d\mu_1}(x_1)\frac{d\nu_2}{d\mu_2}(x_2)|d(\mu_1\times\mu_2)(x_1,x_2)$$
$$= \iint |\frac{d\nu_1}{d\mu_1}(x_1)\frac{d\nu_2}{d\mu_2}(x_2)|d\mu_1(x_1)d\mu_2(x_2)$$
$$= \left[\int |\frac{d\nu_1}{d\mu_1}(x_1)|d\mu_1(x_1)\right]\cdot\left[\int |\frac{d\nu_2}{d\mu_2}(x_2)|d\mu_2(x_2)\right] < \infty$$

Thus, $\frac{d\nu_1}{d\mu_1}(x_1)\cdot\frac{d\nu_2}{d\mu_2}(x_2) \in L^1(\mu_1\times\mu_2)$ and we can conclude by Fubini's theorem that:

$$\iint \chi_E \frac{d\nu_1}{d\mu_1}(x_1)\frac{d\nu_2}{d\mu_2}(x_2)d\mu_1(x_1)d\mu_2(x_2) = \int_E \frac{d\nu_1}{d\mu_1}(x_1)\frac{d\nu_2}{d\mu_2}(x_2)d(\mu_1\times\mu_2)(x_1,x_2).$$

So, we've proven that $\nu_1\times\nu_2(E) = \int_E \frac{d\nu_1}{d\mu_1}(x_1)\frac{d\nu_2}{d\mu_2}(x_2)d(\mu_1\times\mu_2)(x_1,x_2)$. And from there all the conclusions we were trying to prove before are obvious. ■

One more note I'd like to make is that if for $j = 1, 2$ we have that ν_j is a complex measure $(X_j, \mathcal{M}_j)$, then $|\nu_1\times\nu_2| = |\nu_1|\times|\nu_2|$. This is easily seen from the fact that:

$$\nu_1\times\nu_2 = \frac{d\nu_1}{d|\nu_1|}(x)\frac{d\nu_2}{d|\nu_2|}(y)d(|\nu_1|\times|\nu_2|)(x,y)$$
$$\Longrightarrow |\nu_1\times\nu_2| = \left|\frac{d\nu_1}{d|\nu_1|}(x)\frac{d\nu_2}{d|\nu_2|}(y)\right|d(|\nu_1|\times|\nu_2|)(x,y) = 1\cdot d(|\nu_1|\times|\nu_2|)$$

Also, since $\left|\frac{d\nu_j}{d|\nu_j|}\right| = 1$ a.e. we know $f \in L^1(|\nu_1|\times|\nu_2|)$ iff $f\cdot\frac{d\nu_1}{d|\nu_1|}\cdot\frac{d\nu_2}{d|\nu_2|} \in L^1(|\nu_1|\times|\nu_2|)$. And when the latter is true then we can easily see via the Fubini-Tonelli theorem that:

$$\int f d\nu_1\times\nu_2 = \iint f d\nu_1 d\nu_2 = \iint f d\nu_2 d\nu_1.$$

And since $|\nu_1\times\nu_2| = |\nu_1|\times|\nu_2|$ is finite, it's clear that any bounded f is in $L^1(\nu_1\times\nu_2)$.

If $\mu, \nu \in M(\mathbb{R}^n)$, we define $\mu * \nu \in M(\mathbb{R}^n)$ by $\mu * \nu(E) := \mu\times\nu(\alpha^{-1}(E))$ where α is addition. (Literally. $\alpha : \mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}^n$ is the map $\alpha(x,y) = x + y$.) We call $\mu * \nu$ the convolution of μ and ν. Also note that $\alpha^{-1}(E) = \{(x,y) \in \mathbb{R}^n\times\mathbb{R}^n : x + y \in E\}$. Thus $\chi_{\alpha^{-1}(E)}(x,y) = \chi_E(x+y)$ and so:

$$\mu * \nu(E) = \mu\times\nu(\alpha^{-1}(E)) = \int \chi_{\alpha^{-1}(E)}(x,y)d(\mu\times\nu)(x,y)$$
$$= \int \chi_E(x+y)d(\mu\times\nu)(x,y)$$

Question: How do we know that $\mu * \nu$ is a well-defined measure in $M(\mathbb{R}^n)$?

To start off, if $E \in \mathcal{B}_{\mathbb{R}^n}$ then we know $\alpha^{-1}(E) \in \mathcal{B}_{\mathbb{R}^n\times\mathbb{R}^n}$. because α is continuous. Next, we clearly have that $\chi_{\alpha^{-1}(E)} \in L^1(|\mu\times\nu|) = L^1(|\mu|\times|\nu|)$ since $|\mu|$ and $|\nu|$ are both finite. This proves that the integral expression $\int \chi_{\alpha^{-1}(E)}d(\mu\times\nu)(x,y)$ is well-defined and finite.

Also, as a side note it means that:

$$\mu * \nu(E) = \iint \chi_E(x+y)\mathrm{d}\mu(x)\mathrm{d}\nu(y) = \iint \chi_E(x+y)\mathrm{d}\nu(y)\mathrm{d}\mu(x).$$

Now showing that $\mu * \nu$ is a countably additive measure is as simple as noting that $\alpha^{-1}(\emptyset) = \emptyset$ and also that if $\{E_n\}_{n\in\mathbb{N}}$ is a sequence of disjoint sets in $\mathcal{B}_{\mathbb{R}^n}$, then $\{\alpha^{-1}(E_n)\}_{n\in\mathbb{N}}$ is a sequence of disjoint sets in $\mathcal{B}_{\mathbb{R}^n} \otimes \mathcal{B}_{\mathbb{R}^n}$. Meanwhile, by theorem 7.8 in my math 240c notes, it suffices to show that $|\mu * \nu|(K) < \infty$ for all compact sets K in order to say that $\mu * \nu$ is regular (and thus Radon). Yet this is trivially true for any complex measure.

9/11/2025

Proposition 8.48:

(a) Convolution of measures is commutative and associative.

Proof:
Commutativity is obvious. Meanwhile to show associativity note that:

- $$\begin{aligned}((\lambda * \mu) * \nu)(E) &= \iint \chi_E(x'+z)\mathrm{d}(\lambda * \mu)(x')\mathrm{d}\nu(z) \\ &= \int (\lambda * \mu)(E + (-z))\mathrm{d}\nu(z) \\ &= \int (\iint \chi_{E+(-z)}(x+y)\mathrm{d}\lambda(x)\mathrm{d}\mu(y)\mathrm{d})\nu(z) \\ &= \iiint \chi_E(x+y+z)\mathrm{d}\lambda(x)\mathrm{d}\mu(y)\mathrm{d}\nu(z)\end{aligned}$$

- Similarly, you can show that:
$$(\lambda * (\mu * \nu))(E) = \iiint \chi_E(x+y+z)\mathrm{d}\lambda(x)\mathrm{d}\mu(y)\mathrm{d}\nu(z).$$

Thus since both measures equal a common expression, we are done. ■

(b) For any bounded Borel measurable function h:

$$\int h\mathrm{d}(\mu * \nu) = \iint h(x+y)\mathrm{d}\mu(x)\mathrm{d}\nu(y).$$

Proof:
This is obvious for characteristic functions, and by applying standard linearity arguments we can show this is true when h is a simple function.

For the general case, suppose $h : X \to \mathbb{C}$ is Borel measurable. Then, using a theorem on page 44 of my latex math 240a notes, we know there exists a sequence $\{\phi_n\}_{n\in\mathbb{N}}$ of simple functions such that $\{\phi_n\}_{n\in\mathbb{N}}$ is monotone increasing and $\phi_n \to h$ pointwise. If h is bounded, then we can conclude by the dominated convergence theorem using $|h|$ and $\int |h(x,y)|\mathrm{d}\mu(x)$ as upper bounds that:

$$\begin{aligned}\int h\mathrm{d}(\mu * \nu) &= \lim_{n\to\infty} \int \phi_n \mathrm{d}(\mu * \nu) \\ &= \lim_{n\to\infty} \iint \phi_n(x+y)\mathrm{d}\mu(x)\mathrm{d}\nu(y) = \iint h(x+y)\mathrm{d}\mu(x)\mathrm{d}\nu(y) \text{ ■}\end{aligned}$$

(c) $\|\mu * \nu\| \leq \|\mu\|\|\nu\|$.

Proof:
Let $h = \frac{\mathrm{d}(\mu*\nu)}{\mathrm{d}|\mu*\nu|}$. Then we have that:

$$\|\mu * \nu\| = |\mu * \nu|(X) = \int \mathrm{d}|\mu * \nu| = \int \overline{h}h\mathrm{d}|\mu * \nu| = \int \overline{h}\mathrm{d}(\mu * \nu).$$

Next, $|\overline{h}| = 1$ a.e. Therefore by applying part (b) we have that:

$$\begin{aligned}\int \overline{h}\mathrm{d}(\mu * \nu) &= |\textstyle\iint \overline{h}(x+y)\mathrm{d}\mu(x)\mathrm{d}\nu(y)| \\ &\leq \textstyle\iint |\overline{h}(x+y)|\mathrm{d}|\mu|(x)\mathrm{d}|\nu|(y) = \iint 1\mathrm{d}|\mu|(x)\mathrm{d}|\nu|(y) \\ &= |\mu|(\mathbb{R}^n)|\nu|(\mathbb{R}^n) = \|\mu\|\|\nu\| \blacksquare\end{aligned}$$

(d) If $\mu = f\mathrm{d}m$ and $\nu = g\mathrm{d}m$, then $\mu * \nu = (f*g)\mathrm{d}m$; that's to say, if we view $L^1(m)$ as a subspace of $M(\mathbb{R}^n)$ via the mapping $f \mapsto f\mathrm{d}m$ (where m is the Lebesgue measure), then our old and new definition of convolutions coincide.

Proof:
Let $E \in \mathcal{B}_{\mathbb{R}^n}$. Now $f(x)g(y) \in L^1(m \times m)$ by exercise 2.51 (in my math 240a latex notes) plus the fact that $|\mu|$ and $|\nu|$ are finite. So by Fubini's theorem plus the translation invariance of the Lebesgue measure, we have that:

$$\begin{aligned}\mu * \nu(E) &= \textstyle\iint \chi_E(x+y)\mathrm{d}x\mathrm{d}y \\ &= \textstyle\iint \chi_E(x+y)f(x)g(y)\mathrm{d}x\mathrm{d}y \\ &= \textstyle\iint \chi_E(x)f(x-y)g(y)\mathrm{d}x\mathrm{d}y \\ &= \textstyle\int_E \left(\int f(x-y)g(y)\mathrm{d}y\right)\mathrm{d}x = \int_E f * g\mathrm{d}x \blacksquare\end{aligned}$$

From now on if I write L^p, let it be assumed that I'm referring to $L^p(\mathbb{R}^n, m)$ where m is the Lebesgue measure. Then, we shall define convolutions of measures with functions in L^p for any p.

Proposition 8.49: If $f \in L^p(\mathbb{R}^n)$ (where $1 \leq p \leq \infty$) and $\mu \in M(\mathbb{R}^n)$, then the integral $f * \mu(x) := \int f(x-y)\mathrm{d}\mu(y)$ exists for a.e. x, $f * \mu \in L^p$, and $\|f * \mu\|_p \leq \|f\|_p\|\mu\|$.

Proof:
To start off, we know by Minkowski's inequality that:

$$\begin{aligned}\left(\int \left|\int |f(x-y)|\mathrm{d}|\mu|(y)\right|^p \mathrm{d}x\right)^{1/p} &\leq \int \left(\int |f(x-y)|^p\mathrm{d}x\right)^{1/p}\mathrm{d}|\mu|(y) \\ &= \int \left(\int |f(x)|^p\mathrm{d}x\right)^{1/p}\mathrm{d}|\mu|(y) \\ &= \int \|f\|_p\mathrm{d}|\mu| = \|f\|_p \int \mathrm{d}|\mu| = \|f\|_p\|\mu\|.\end{aligned}$$

Then since $\|f\|_p\|\mu\| < \infty$, this tells us that $\int |f(x-y)|\mathrm{d}|\mu|(y)$ exists and is finite for a.e. x. In turn, this means that $\int f(x-y)\mathrm{d}\mu(y)$ exists and is finite for a.e. x. And since $|\int f(x+y)\mathrm{d}\mu(y)| \leq \int |f(x-y)\mathrm{d}|\mu|(y) = |\int |f(x-y)\mathrm{d}|\mu|(y)|$, we're done. $\blacksquare$

Note that if $p=1$, then our two definitions of $f*\mu$ coincide (where again we are considering L^1 as a subspace of $M(\mathbb{R}^n)$). After all,

$$\int_E f*\mu(x)\mathrm{d}x = \iint \chi_E(x)f(x-y)\mathrm{d}\mu(y)\mathrm{d}x = \iint \chi_E(x+y)(f(x)\mathrm{d}x)\mathrm{d}\mu(y).$$

Thus, consider equipping $M(\mathbb{R}^n)$ with the convolution operator in addition to its other vector space operations. Then it's easy to see that $\mu*(\nu+\lambda) = (\mu*\nu)+(\mu*\lambda)$ for all $\mu,\nu,\lambda\in M(\mathbb{R}^n)$. Also, consider the point mass measure μ defined by $\iota(E)=1$ if $0\in E$ and $\iota(E)=0$ otherwise. Clearly $\iota\in M(X)$ since ι is a measure on $\mathbb{R}^n$ that is finite on all compact sets. Also, if μ is any other measure in $M(\mathbb{R}^n)$ then we have that $\iota*\mu=\mu$. After all:

$$\iota*\mu(E) = \iint \chi_E(x+y)\mathrm{d}\iota(x)\mathrm{d}\mu(y) = \int \chi_E(y)\mathrm{d}\mu(y) = \mu(E)$$

Thus it follows that $M(\mathbb{R}^n)$ when equipped with pointwise addition and convolution is a commutative ring, and by proposition 8.49 we know that L^1 is an ideal in $M(\mathbb{R}^n)$.

Also, it's clear that $c(\mu*\nu) = (c\mu)*\nu = \mu*(c\nu)$. So, $M(\mathbb{R}^n)$ is also an algebra. And it's clear either by proposition 8.48(d) or the fact that L^1 is an ideal that we also know that L^1 is a subalgebra of $M(\mathbb{R}^n)$.

We now extend the Fourier Transform from L^1 to $M(\mathbb{R}^n)$ in the obvious way. Specifically, if $\mu\in M(\mathbb{R}^n)$, then define the function $\widehat{\mu}:\mathbb{R}^n\to\mathbb{C}$ by $\widehat{\mu}(\xi) = \int e^{-2\pi i\xi\cdot x}\mathrm{d}\mu(x)$.

Right now it is relevant to express a generalization of a result that Folland gave in chapter 2 of his book.

<u>Theorem:</u> Consider any measure space $(X,\mathcal{M},\mu)$ (where μ is positive), let $U\subseteq\mathbb{R}^n$ be an open set, and suppose that $f: X\times U\to\mathbb{C}$ is a function such that $f(\cdot,y): X\to\mathbb{C}$ is integrable for all $y\in U$. Also define $F(y) = \int_X f(x,t)\mathrm{d}\mu(x)$.

(a) Suppose there exists $g\in L^1(\mu)$ such that $|f(x,y)|\leq g(x)$ for all $x\in X$ and $y\in U$. If $\lim_{y\to y'} f(x,y) = f(x,y')$ for all $x\in X$, then $\lim_{y\to y'} F(y) = F(y')$. In particular, if $f(x,\cdot)$ is continuous for each x, then F is continuous.

(b) Denote any $y\in U$ as $y=(y_1,\ldots,y_n)$. Now suppose that $\partial f/\partial y_j$ exists for all $y\in U$ and that there exists $g\in L^1(\mu)$ such that $|(\partial f/\partial y_j)(x,y)|\leq g(x)$ for all $x\in X$ and $y\in U$. Then $\partial F/\partial y_j$ exists and $\frac{\partial F}{\partial y_j}(y) = \int(\partial f/\partial y_j)(x,y)\mathrm{d}\mu(x)$.

The proof for part (a) is identical to that of theorem 2.27 in Folland (see my paper math 240a notes). Just apply dominated convergence theorem to the functions $f_m(x) := f(x,y^{(m)})$ where $(y^{(m)})_{m\in\mathbb{N}}$ is any sequence in U converging to y'.

Side note: technically the proof for part (a) doesn't require U to be open.

To prove part (b), let $e_1,\ldots,e_n$ be the standard basis vectors for $\mathbb{R}^n$ and consider for any $y\in U$ that there exists some $r>0$ such that $B_r(y)\subseteq U$. Thus for any sequence $(t_m)_{m\in\mathbb{N}}$ in $(-r,r)$ converging to 0 we have that:

$$\frac{\partial f}{\partial y_j}(x,y) = \lim_{m\to\infty} h_m(x) \text{ where } h_m(x) = \frac{f(x,y+t_m e_j)-f(x,y)}{t_m}.$$

(Note that this guarentees that $\frac{\partial f}{\partial y_j}$ is measurable for all y.)

Now importantly $\tilde{f}(t) := f(x, y + te_j)$ is differentiable with $\frac{\mathrm{d}\tilde{f}}{\mathrm{d}t}(t) = \frac{\partial f}{\partial y_j}(x, y + te_j)$.

So, we know by the mean value theorem that $|\tilde{f}(t) - \tilde{f}(0)| \leq |t| \cdot \sup_{t \in (-r,r)} |\tilde{f}'(t)|$. And in turn, we have that:

$$|h_m(x)| \leq \sup_{t \in (-r,r)} \frac{\partial f}{\partial y_j}(x, y + te_j) \leq g(x).$$

Hence, we may apply dominated convergence theorem to say that:

$$\begin{aligned}\lim_{m \to \infty} \tfrac{F(y+t_m) - F(y)}{t} &= \lim_{m \to \infty} \frac{\int f(x, y + t_m e_j) \mathrm{d}\mu(x) - \int f(x,y)\mathrm{d}\mu(x)}{t} \\ &= \lim_{m \to \infty} \int h_m(x) \mathrm{d}\mu(x) = \int \lim_{m \to \infty} h_m(x)\mathrm{d}\mu(x) = \int \tfrac{\partial f}{\partial y_j}(x,y)\mathrm{d}\mu(x)\end{aligned}$$

This proves that $\frac{\partial F}{\partial y_j}(y) = \int \frac{\partial f}{\partial y_j}(x,y)\mathrm{d}\mu(x)$. ■

Also note that if μ is a complex measure, then by expressing $\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$ where each μ_j is a positive measure and applying what we already proved to each $\int f(x,y)\mathrm{d}\mu_j(x)$, we can show that the previous theorem holds for any function $F(y) = \int f(x,y)\mathrm{d}\mu(x)$ provided that there exists some $g : X \to \mathbb{C}$ in $L^1(|\mu|)$ such that $|f(x,y)| \leq g(x)$ for all $x \in X$ and $y \in U \subseteq \mathbb{R}^n$.

Since $e^{-2\pi i \xi \cdot x}$ is continuous with respect to x and $|e^{-2\pi i\xi\cdot x}| = 1 \in L^1(|\mu|)$, we know by the prior reasoning that $\widehat{\mu}$ is continuous. Also, it is clear that $|\widehat{\mu}(\xi)| \leq \|\mu\|$. And, we can see for any $\mu, \nu \in M(\mathbb{R}^n)$ that $(\mu * \nu)^\wedge = \widehat{\mu}\widehat{\nu}$. After all:

$$\begin{aligned}(\mu * \nu)^\wedge(\xi) &= \int e^{-2\pi i \xi \cdot x'} \mathrm{d}(\mu * \nu)(x') \\ &= \iint e^{-2\pi i \xi \cdot (x+y)} \mathrm{d}\mu(x)\mathrm{d}\nu(y) \\ &= \left(\int e^{-2\pi i \xi \cdot x}\mathrm{d}\mu(x)\right)\left(\int e^{-2\pi i \xi \cdot y}\mathrm{d}\nu(y)\right) = \widehat{\mu}(\xi)\widehat{\nu}(\xi)\end{aligned}$$

Proposition 8.50: Suppose that $\{\mu_k\}_{k \in \mathbb{N}}$ is a sequence in $M(\mathbb{R}^n)$ and $\mu \in M(\mathbb{R}^n)$. If $\|\mu_k\| \leq C < \infty$ for all k and $\widehat{\mu}_k \to \widehat{\mu}$ pointwise, then $\mu_k \to \mu$ vaguely.

Proof:
If $f \in \mathcal{S}$ (where $\mathcal{S}$ is the collection of Schwartz functions), then $f^\vee \in \mathcal{S}$. So by the Fourier inversion theorem, we have that:

$$\begin{aligned}\int f\mathrm{d}\mu_k &= \int(\int f^\vee(x) e^{-2\pi i \xi \cdot x}\mathrm{d}x)\mathrm{d}\mu_k(\xi) \\ &= \int f^\vee(x)(\int e^{-2\pi i \xi \cdot x}\mathrm{d}\mu_k(\xi))\mathrm{d}x = \int f^\vee(x)\widehat{\mu}_k(x)\mathrm{d}x\end{aligned}$$

By similar resoning we know that $\int f\mathrm{d}\mu = \int f^\vee(x)\widehat{\mu}(x)\mathrm{d}x$. But now since $\|\widehat{\mu_k}\|_u \leq C$ for all k and $f^\vee\widehat{\mu}_k \to f^\vee\widehat{\mu}$ pointwise, we can conclude by the dominated convergence theorem that:

$$\begin{aligned}\lim_{k \to \infty} \int f\mathrm{d}\mu_k &= \lim_{k \to \infty} \int f^\vee(x)\widehat{\mu_k}(x)\mathrm{d}x \\ &= \int \lim_{k \to \infty} f^\vee(x)\widehat{\mu_k}(x)\mathrm{d}x = \int f^\vee(x)\widehat{\mu}(x)\mathrm{d}x = \int f\mathrm{d}\mu\end{aligned}$$

To finish off, note that since $\mathcal{S}$ is dense in $C_0(\mathbb{R}^n)$, we know by Folland proposition 5.17 (see my math 240b notes) that $\int f\mathrm{d}\mu_k \to \int f\mathrm{d}\mu$ as $k \to \infty$ for all $f \in C_0(\mathbb{R}^n)$. ■

Here's a relevant exercise from Folland:

Exercise 7.26: Suppose X is an LCH space. If $\{\mu_k\}_{k\in\mathbb{N}} \subseteq M(X)$, $\mu_k \to \mu$ vaguely, and $\|\mu_k\| \to \|\mu\|$, then $\int f\mathrm{d}\mu_k \to \int f\mathrm{d}\mu$ for every $f \in BC(X)$.

In the case that $\mu = 0$, then this is trivial from the fact that $\|\mu_k\| = |\mu_k|(X) \to 0$ as $k \to \infty$. Meanwhile, if $\mu \neq 0$, then consider any $\varepsilon > 0$. We can show that there exists a compact set $F \subseteq X$ such that $|\mu_k|(F^{\mathsf{C}}) < 4\varepsilon$ for all k and that $|\mu|(F^{\mathsf{C}}) < 4\varepsilon$.

By Lusin's theorem, we know there exists a function $h \in C_c(X, [0,1])$ such that $\|h\|_u \leq 1$ and $h = \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}$ everwhere outside a set E of $|\mu|$-measure less than ε. Thus we have that $|\int h\mathrm{d}\mu_k| \leq \int |h|\mathrm{d}|\mu_k| \leq \|\mu_k\|$, that $|\int h\mathrm{d}\mu| \leq \|\mu\|$, and that:

$$|\int h\mathrm{d}\mu| = |\int_{E^{\mathsf{C}}} \mathrm{d}|\mu| + \int_E h\mathrm{d}\mu| \geq |\mu|(E^{\mathsf{C}}) - |\mu|(E) > \|\mu\| - 2\varepsilon$$

Also, since $\mu_k \to \mu$ vaguely, we know that $\int h\mathrm{d}\mu_k \to \int h\mathrm{d}\mu$ as $k \to \infty$. Hence, there is some $N_1 \in \mathbb{N}$ such that for all $k > N_1$, $|\int h\mathrm{d}\mu_k - \int h\mathrm{d}\mu| < \varepsilon$. But in turn we have that $|\int h\mathrm{d}\mu_k| > \|\mu\| - 3\varepsilon$ for all $k > N_1$. Meanwhile, since $\|\mu_k\| \to \|\mu\|$, we can find some $N_2 \in \mathbb{N}$ greater than N_1 such that $|\int h\mathrm{d}\mu_k| > \|\mu_k\| - 4\varepsilon$ for all $k > N_2$. Thus, by setting $F' = \mathrm{supp}(h)$ it is clear that $|\mu_k|(X - F') < 4\varepsilon$ for all $k > N_2$ and that $|\mu|(X - F') < \varepsilon < 4\varepsilon$.

Now to finish off, for all $k \in \mathbb{N}$ with $k \leq N_2$, we can find by the inner regularity of $|\mu_k|$ on X a compact set F_k such that $|\mu_k|(F_k^{\mathsf{C}}) < 4\varepsilon$. And by setting $F = F' \cup (\bigcup_{k=1}^{N_2} F_k)$, we are done.

Now by Urysohn's lemma, we know there exists a function $g \in C_c(X, [0,1])$ such that $g = 1$ on F. It follows for any $f \in BC(X)$ that:

$$|\int (f - gf)\mathrm{d}\mu_k| = |\int_{F^{\mathsf{C}}} (1-g)f\mathrm{d}\mu_k| \leq \int_{F^{\mathsf{C}}} 1\mathrm{d}|\mu_k| < 4\varepsilon$$

Similarly we have that $|\int (f - gf)\mathrm{d}\mu| < 4\varepsilon$. And importantly since $fg \in C_c(X)$ for any $f \in BC(X)$ and $\mu_k \to \mu$ vaguely, we know that $\lim_{k\to\infty} |\int gf\mathrm{d}\mu_k - \int gf\mathrm{d}\mu| = 0$. So, we may conclude that:

$$\begin{aligned}&\lim_{k\to\infty} |\textstyle\int f\mathrm{d}\mu_k - \int f\mathrm{d}\mu| \\ &\qquad \leq \lim_{k\to\infty} (|\textstyle\int (f-gf)\mathrm{d}\mu_k| + |\int gf\mathrm{d}\mu_k - \int gf\mathrm{d}\mu| + |\int (f-gf)\mathrm{d}\mu|) < 8\varepsilon\end{aligned}$$

And taking $\varepsilon \to 0$ we have shown what we wanted.

Moreover, the hypothesis that $\|\mu_k\| \to \|\mu\|$ can't be omitted.

Consider $X = \mathbb{R}$ and let $\mu = 0$ and μ_k be the point mass centered at k (i.e. $\mu_k(E) = 1$ if $k \in E$ and 0 otherwise). Now clearly $\|\mu_k\| \not\to \|\mu\|$ as $k \to \infty$ since $\|\mu_k\| = 1$ for all k but $\|\mu\| = 0$. Also it's clear that $\mu_k \to \mu$ vaguely since for all $f \in C_0(\mathbb{R})$, we have that $\int f\mathrm{d}\mu_k = f(k) \to 0 = \int f\mathrm{d}\mu$ as $k \to \infty$ since f vanishes at infinity. That said, it is not the case that $\int f\mathrm{d}\mu_k \to \int f\mathrm{d}\mu$ for all $f \in BC(\mathbb{R})$, and a simple example illustrating that is $f = 1$. ■

A corollary of the above exercise is that if $\{\mu_k\}_{k\in\mathbb{N}}$ is a sequence in $M(\mathbb{R}^n)$, $\mu \in M(\mathbb{R}^n)$, $\mu_k \to \mu$ vaguely, and $\|\mu_k\| \to \|\mu\|$, then $\widehat{\mu}_k \to \widehat{\mu}$ pointwise. This is because $e^{-2pii\xi\cdot x}$ is a bounded continuous function for all $\xi \in \mathbb{R}^n$.

Exercise 8.40: $L^1(\mathbb{R}^n)$ is vaguely dense in $M(\mathbb{R}^n)$.

Let $\varphi \in L^1(\mathbb{R}^n)$ such that $\varphi \geq 0$ and $\int \varphi(x)\mathrm{d}x = 1$. Then put $\varphi_t(x) := \frac{1}{t^n}\varphi(x/t)$ for all $t > 0$. Now our claim is that for any $\mu \in M(\mathbb{R}^n)$, we have that $(\varphi_t * \mu)\mathrm{d}x \to \mu$ vaguely as $t \to 0$.

To start off, we know that $\|(\varphi_t * \mu)\mathrm{d}x\| \leq \|\varphi_t\|_1\|\mu\| = \|\mu\| < \infty$ for all $t > 0$.

Also note that $((\varphi_t * \mu)\mathrm{d}x)^\wedge(\xi) = \widehat{\varphi}_t(\xi)\widehat{\mu}(\xi)$. But consider that $\widehat{\varphi}$ is continuous and $\widehat{\varphi}_t(\xi) = \int \frac{1}{t^n}\varphi(\frac{x}{t})e^{-2\pi i\xi\cdot x}\mathrm{d}x = \int \varphi(u)e^{-2\pi i\xi\cdot tu}\mathrm{d}u = \widehat{\varphi}(t\xi)$. Thus it's clear that $\widehat{\varphi}_t(\xi) \to \widehat{\varphi}(0)$ pointwise as $t \to 0$. And since $\int \varphi(x)\mathrm{d}x = 1$, we know that $\widehat{\varphi}(0) = 1$. Hence, $((\varphi_t * \mu)\mathrm{d}x)^\wedge \to \widehat{\mu}$ pointwise.

Applying *proposition 8.50*, we are done. ■

9/14/2025

I'd been visiting my parents in Ohio for the entire past week and didn't manage to get that much done. But now that I'm heading back to California, I ought to start doing work again. Today I'll be starting on Folland chapter 10 (where he covers probability). But before covering any theorems, I'd just like to shamelessly rip off this table that Folland put together:

Analysts' Term	*Probabilists' Term*
Measure space $(X, \mathcal{M}, \mu)$ $(\mu(X) = 1)$	Sample space $(\Omega, \mathcal{B}, P)$
$(\sigma$-)algebra	$(\sigma$-)field
Measurable set	Event
Measurable real-valued function f	Random variable X
Integral of f, $\int f\,d\mu$	Expectation or mean of X, $E(X)$
L^p [as adjective]	Having finite pth moment
Convergence in measure	Convergence in probability
Almost every(where), a.e.	Almost sure(ly), a.s.
Borel probability measure on $\mathbb{R}$	Distribution
Fourier transform of a measure	Characteristic function of a distribution
Characteristic function	Indicator function

Also, given some proposition Q, probabilists typically write $\{Q(\omega)\}$ as opposed to $\{\omega \in \Omega : Q(\omega) = \text{true}\}$. And given a probability measure P they write $P(Q(\omega))$ as opposed to $P(\{Q(\omega)\})$. As long as I'm studying probability theory, I'm going to use probability terminology and notation.

Let X be a random variable. We define it's variance σ^2 and standard deviation σ as:

$$\sigma^2(X) := \inf_{a\in\mathbb{R}} E[(x-a)^2] \text{ and } \sigma(X) := \sqrt{\sigma^2(X)}$$

If $X \notin L^2$, then we always have $E((X-a)^2) = \infty$ for all a and thus $\sigma^2(X) = \infty$. Otherwise, since $L^2 \subseteq L^1$ (since P is finite), we have that $X \in L^1 \cap L^2$. So:

$$E((X-a)^2) = E(X^2 - 2aX + a^2) = E(X^2) - 2aE(X) + a^2.$$

Note that this is a quadratic which attains it's minimum at $a = E(X)$.

Sanity check, by completing the square we have:

$$a^2 - 2aE(X) = a^2 - 2aE(X) + (E(X))^2 - (E(X))^2 = (a - E(X))^2 - (E(X))^2.$$

Hence, for all $X \in L^2$ we have that:

$$\sigma^2(X) = E(X^2) - 2(E(X))^2 + (E(X))^2 = E(X^2) - (E(X))^2.$$

Let $(\Omega, \mathcal{B}, P)$ be a sample space and let $(\Omega', \mathcal{B}')$ be another measurable space. Given a $(\mathcal{B}, \mathcal{B}')$-measurable map ϕ, we may define the image measure $P_\phi(E) \coloneqq P(\phi^{-1}(E))$ on $(\Omega', \mathcal{B}')$. This is in fact a measure because preimages commute with unions, intersections, and complements. Also importantly $P_\phi(\Omega') = P(\Omega) = 1$. So P_ϕ is another probability measure.

Proposition 10.1: If $f : \Omega' \to \mathbb{R}$ is a random variable, then $\int f \mathrm{d}P_\phi = \int (f \circ \phi)\mathrm{d}P$ whenever either side exists.

This obviously holds for simple functions and you can extend the result from there using standard methods.

If X is a random variable on Ω, then P_X is a probability measure on $\mathbb{R}$ called the distribution of X. Also, we call $F(t) = P_X((-\infty, t]) = P(X \leq t)$ the distribution function of X. If $\{X_\alpha\}_{\alpha \in A}$ is a family of random variables, we say that $\{X_\alpha\}_{\alpha \in A}$ are identically distributed if $P_{X_\alpha} = P_{X_\beta}$ for all $\alpha, \beta \in A$.

Basically all properties of a random variable we care about can be gathered just from knowing the variable's distribution. For example, by proposition 10.1 we know that:

$$E(X) = \int t\mathrm{d}P_X(t) \text{ and } E[(X-a)^2] = \int (t-a)^2 \mathrm{d}P_X(t).$$

And from there we can easily see if $t \in L^1(P_X) \cap L^2(P_X)$, then:

$$\sigma^2(X) = \int (t - E(X))^2 \mathrm{d}P_X(t).$$

If $X_1, \ldots, X_n$ is a finite sequence of random variables on Ω, then we can consider $(X_1, \ldots, X_n)$ as a map from ω to $\mathbb{R}^n$. Then the induced measure $P_{(X_1, \ldots, X_n)}$ on $\mathbb{R}^n$ is called the joint distribution of $X_1, \ldots, X_n$.

Hopefully it's a given but if you doubt $(X_1, \ldots, X_n)$ is measurable then see the proposition on page 41 of my math 240a latex notes.

Note by proposition 10.1 that given any two random variables X, Y on Ω:

$$\begin{aligned}\int (t+s)\mathrm{d}P_{X,Y}(t,s) &= \int [\pi_1 \circ \mathrm{Id}]\mathrm{d}P_{X,Y} + \int [\pi_2 \circ \mathrm{Id}]\mathrm{d}P_{X,Y} \\ &= \int [\pi_1 \circ \mathrm{Id} \circ (X,Y)]\mathrm{d}P + \int [\pi_2 \circ \mathrm{Id} \circ (X,Y)]\mathrm{d}P \\ &= \int X\mathrm{d}P + \int Y\mathrm{d}P = E(X) + E(Y) = E(X+Y).\end{aligned}$$

Because of the convenience of distributions, given a Borel probability measure λ on $\mathbb{R}$ we'll often talk of the mean $\overline{\lambda}$ and variance $\sigma^2(\lambda)$ of λ:

$$\overline{\lambda} \coloneqq \int t\mathrm{d}\lambda(t) \text{ and } \sigma^2(\lambda) \coloneqq \int (t - \overline{\lambda})^2 \mathrm{d}\lambda(t)$$

Let $(\Omega, \mathcal{F}, P)$ be a sample space and $E \subseteq \Omega$ be an event such that $P(E) > 0$. Then $P_E(F) := P(E \cap F)/P(E)$ is a probability measure on E which gives the conditional probability of F given E. $P_E(F)$ is also denoted $P(F|E)$.

It's clear if $P(E) > 0$ that we then have that $P(E \cap F) = P(E)P(F|E)$. An important case to consider is when $P(F|E) = P(F)$ (which corresponds to when knowing $\omega \in E$ doesn't give us any information about if $\omega \in F$). In that case we say that E and F are independent. More generally, we call two events E and F independent whenever $P(E \cap F) = P(E)P(F)$. Even more generally, we call an arbitrary collection $\{E_\alpha\}_{\alpha \in A}$ of events in Ω independent iff for all $n \in \mathbb{N}$ and $\alpha_1, \ldots, \alpha_n \in A$, we have that $P(\bigcap_{j=1}^n E_{\alpha_j}) = \prod_{j=1}^n P(E_{\alpha_j})$.

From this definition hopefully it is obvious to you that if $C \subseteq A$, then $\{E_\alpha\}_{\alpha \in C}$ is a collection of independent random variables.

An arbitrary collection of random variables $\{X_\alpha\}_{\alpha \in A}$ are called independent if $\{X_\alpha^{-1}(B_\alpha)\}_{\alpha \in A}$ is independent for all choices of $B_\alpha \in \mathcal{B}_\mathbb{R}$.

Observe that:

$$\begin{aligned} P(X_1^{-1}(B_1) \cap \ldots \cap X_n^{-1}(B_n)) &= P((X_1, \ldots, X_n)^{-1}(B_1 \times \ldots \times B_n)) \\ &= P_{X_1,\ldots,X_n}(B_1 \times \ldots \times X_n) \end{aligned}$$

Meanwhile $\prod_{j=1}^n P(X_j^{-1}(B_j)) = \prod_{j=1}^n P_{X_j}(B_j) = \left(\prod_{j=1}^n P_{X_j}\right)(B_1 \times \ldots \times B_n)$.

Thus, we have that $P_{X_1,\ldots,X_n} = \prod_{j=1}^n P_{X_j}$ if $X_1, \ldots, X_n$ are independent. Also, we can easily show the converse by just setting $B_j = \mathbb{R}$ for any j we want to ignore when proving independence.

Proposition 10.2: Let $\{X_{n,j} : 1 \leq j \leq J(n), 1 \leq n \leq N\}$ be independent random variables, and let $f_n : \mathbb{R}^{J(n)} \to \mathbb{R}$ be Borel measurable for $1 \leq n \leq N$. Then the random variables $\{Y_n = f_n(X_{n,1}, \ldots, X_{n,J(n)}) : 1 \leq n \leq N\}$ are independent.

Proof:

Let $Z_n = (X_{n,1}, \ldots, X_{n,J(n)})$. If $B_1, \ldots, B_N$ are Borel subsets of $\mathbb{R}$, we have that $Y_n^{-1}(B_n) = Z_n^{-1}(f_n^{-1}(B_n))$. Hence:

$$\begin{aligned} (Y_1, \ldots, Y_N)^{-1}(B_1 \times \ldots \times B_n) &= \textstyle\bigcap_{n=1}^N Y_n^{-1}(B_n) \\ &= \textstyle\bigcap_{n=1}^N Z_n^{-1}(f_n^{-1}(B_n)) \\ &= (Z_1, \ldots, Z_N)^{-1}(f_1^{-1}(B_1) \times \ldots \times f_N^{-1}(B_N)) \end{aligned}$$

But now note that $(Z_1, \ldots, Z_n)$ and $(X_{1,1}, \ldots, X_{N,J(N)})$ are identical functions. And so by the independence of the $X_{n,j}$ and *Exercise 2.45*, note that:

$$\begin{aligned} P_{Z_1,\ldots,Z_n} &= P_{X_{1,1},\ldots,X_{N,J(N)}} \\ &= \prod_{n=1,\, j=1}^{N,\, J(n)} P_{X_{n,j}} = \prod_{n=1}^{N}\left(\prod_{j=1}^{J(n)} P_{X_{n,j}}\right) = \prod_{n=1}^{N}\left(P_{X_{n,1},\ldots,X_{n,J(n)}}\right) = \textstyle\prod_{n=1}^N P_{Z_n} \end{aligned}$$

Therefore, we have that:

$$\begin{aligned} P_{Y_1,\ldots,Y_N}(B_1 \times \ldots \times B_N) &= P((Y_1,\ldots,Y_N)^{-1}(B_1 \times \ldots \times B_N)) \\ &= P((Z_1,\ldots,Z_N)^{-1}(f_1^{-1}(B_1) \times \ldots \times f_N^{-1}(B_N))) \\ &= P_{Z_1,\ldots,Z_n}(f_1^{-1}(B_1) \times \ldots \times f_N^{-1}(B_N)) \\ &= \prod_{n=1}^N P_{Z_n}(f_n^{-1}(B_n)) \\ &= \prod_{n=1}^N P(Z_n^{-1}(f_n^{-1}(B_n))) \\ &= \prod_{n=1}^N P(Y_n^{-1}(B_n)) = \prod_{n=1}^N P_{Y_n}(B_n). \blacksquare \end{aligned}$$

9/15/2025

Quick lemma: Let $(\Omega_1, \mathcal{F}_1, P)$ be a sample space, let $(\Omega_2, \mathcal{F}_2)$, $(\Omega_3, \mathcal{F}_3)$ be measurable spaces, and let $\phi : \Omega_1 \to \Omega_2$, $\psi : \Omega_2 \to \Omega_3$ be measurable maps. Then $P_{\psi \circ \phi} = (P_\phi)_\psi$.

Proof:

$$P_{\psi\circ\phi}(E) = P((\psi \circ \phi)^{-1}(E)) = P(\phi^{-1}(\psi^{-1}(E))) = P_\phi(\psi^{-1}(E)) = (P_\phi)_\psi(E).$$

(I'm mentioning this because Folland implicitely uses this in the next proof...)

By easy induction on our proof for *proposition 8.48(a)*, we can see that if $\lambda_1, \ldots, \lambda_n$ are all in $M(\mathbb{R}^n)$, then:

$$\lambda_1 * \ldots * \lambda_n(E) = \int \cdots \int \chi_E(\textstyle\sum_{i=1}^n t_n) \mathrm{d}\lambda_1(t_1) \cdots \mathrm{d}\lambda_n(t_n).$$

Proposition 10.4: If $X_1, \ldots, X_n$ are independent random variables, then:

$$P_{X_1+\ldots+X_n} = P_{X_1} * \ldots * P_{X_n}.$$

Proof:

Let $A(t_1, \ldots, t_n) = \sum_{j=1}^n t_j$. Then A is a Borel measurable map from $\mathbb{R}^n$ to $\mathbb{R}$. So:

$$\begin{aligned} P_{X_1+\ldots+X_n}(E) &= (P_{X_1,\ldots,X_n})_A(E) \\ &= \left(\textstyle\prod_{j=1}^n P_{X_j}\right)_A(E) = \left(\textstyle\prod_{j=1}^n P_{X_j}\right)(A^{-1}(E)) \\ &= (P_{X_1} * \ldots * P_{X_n})(E). \blacksquare \end{aligned}$$

Propostion 10.5: Suppose that $X_1, \ldots, X_n$ are independent random variables. If $X_j \in L^1$ for all $1 \leq j \leq n$, then $\prod_{j=1}^n X_j \in L^1$ and $E(\prod_{j=1}^n X_j) = \prod_{j=1}^n E(X_j)$.

Proof:

Let $f(t_1, \ldots, t_n) = \prod_{j=1}^n |t_j|$. Then by the independence of the X_j we have that:

$$E(|\textstyle\prod_{j=1}^n X_j|) = \int f \circ (X_1, \ldots, X_n) \mathrm{d}P = \int f \mathrm{d}P_{X_1,\ldots,X_n} = \int f \mathrm{d}(P_{X_1} \times \ldots \times P_{X_n})$$

And since f is nonnegative, we know by Tonelli's theorem that:

$$\int f \mathrm{d}(P_{X_1} \times \ldots \times P_{X_n}) = \textstyle\prod_{j=1}^n \int |t_j| \mathrm{d}P_{X_j} = \prod_{j=1}^n E(|X_j|)$$

Hence if all the $X_j \in L^1$, we've shown that $E(|\prod_{j=1}^n X_j|) = \prod_{j=1}^n E(|X_j|) < \infty$. This proves our first assertion that $\prod_{j=1}^n X_j \in L^1$. And to our prove our second assertion, just do the prior reasoning all over again but remove all the absolute value signs and use Fubini's theorem instead of Tonelli's. $\blacksquare$

Corollary 10.6: Suppose that $X_1, \ldots, X_n$ are independent random variables and in L^2. Then $\sigma^2(X_1 + \ldots + X_n) = \sum_{j=1}^n \sigma^2(X_j)$.

Proof:
Let $Y_j = X_j - E(X_j)$. Then by proposition 10.2, we have that all the Y_j are independent and have an expected value of 0. So by our last proposition, we know for all $j \neq k$ that $E(Y_jY_k) = E(Y_j)E(Y_k) = 0$. And hence:

$$\begin{aligned}\sigma^2(X_1 + \ldots + X_n) &= E((X_1 + \ldots + X_n - E(X_1 + \ldots + X_n))^2) \\ &= E((Y_1 + \ldots + Y_n)^2) \\ &= \sum_{j=1}^n \sum_{k=1}^n E(Y_jY_k) = \sum_{j=1}^n E(Y_j^2) = \sum_{j=1}^n \sigma^2(X_j). \blacksquare\end{aligned}$$

These prior properties show us that independence is actually a very strong assumption to make. In fact, by propositions 10.2 and 10.5, if X and Y are independent random variables with $E(X) = 0$, then given any Borel measurable function f such that $f \circ Y \in L^1$, we have that $E(X \cdot f(Y)) = E(X)E(f(Y)) = 0$. Hence X is orthogonal in L^2 to every function of Y with finite 2nd moment.

That said, fortunately there is an easy way of constructing collections of independent random variables. Specifically, given a collection of sample spaces $\{(\Omega_j, \mathcal{B}_j, P_j)\}_{j=1}^n$, let $\Omega = \Omega_1 \times \ldots \times \Omega_n$, $\mathcal{B} = \mathcal{B}_1 \otimes \ldots \otimes \mathcal{B}_n$, and $P = P_1 \times \ldots \times P_n$. Then any collection $\{X_1, \ldots, X_n\}$ of random variables on Ω such that X_j depends only on the jth. coordinate of Ω for each j is independent.

After all, express every X_j as $f_j \circ \pi_j$ where f_j is some $\mathcal{B}_j$-measurable function and π_j is the projection of Ω onto Ω_j. Now given any rectangle $B_1 \times \ldots \times B_n$ in $\mathcal{B}_{\mathbb{R}^n}$, we have that:

$$\begin{aligned}P_{X_1,\ldots,X_n}(B_1 \times \ldots \times B_n) &= P((X_1, \ldots, X_n)^{-1}(B_1 \times \ldots \times B_n)) \\ &= P(\prod_{j=1}^n f_j^{-1}(B_j)) = \prod_{j=1}^n P_j(f_j^{-1}(B_j)) = \prod_{j=1}^N (P_j)_{f_j}(B_j).\end{aligned}$$

Now it is easily seen that:

$$\begin{aligned}P_{\pi_j}(E) &= P(\Omega_1 \times \ldots \times \Omega_{j-1} \times E \times \Omega_{j+1} \times \ldots \times \Omega) \\ &= P_j(E) \cdot \prod_{i \neq j} P_i(\Omega_i) = P_j(E) \cdot 1 = P_j(E)\end{aligned}$$

Thus $(P_j)_{f_j} = (P_{\pi_j})_{f_j} = P_{f_j \circ \pi_j} = P_{X_j}$ and we've shown that $P_{X_1,\ldots,X_n}$ and $\prod_{j=1}^n P_{X_j}$ agree on all rectangles. It now follows by standard arguments that they equal for all sets in $\mathcal{B}_{\mathbb{R}^n}$.

Exercise 10.9: Suppose that $(X_n)_{n \in \mathbb{N}}$ is a sequence of random variables. If $X_n \to X$ in probability, then $P_{X_n} \to P_X$ vaguely.

Note that we trivially have that $\sup_{n \in \mathbb{N}} \|P_{X_n}\| = 1 < \infty$ since all the P_{X_n} are probability measures. Meanwhile, define $F_n(t) := P_{X_n}((-\infty, t])$ and $F(t) := P_{X_n}((-\infty, t])$. We want to show that if F is continuous at $t_0 \in \mathbb{R}$ then we have that $F_n(t_0) \to F(t_0)$ as $n \to \infty$. This is so that we can then apply *proposition 7.19* and be done.

So given any $\varepsilon > 0$, use the continuity of F at t_0 to pick $\delta > 0$ such that $|F(t) - F(t_0)| < \varepsilon$ whenever $|t - t_0| < \delta$. Since $X_n \to X$ in measure, we may pick $N > 0$ such that $P(|X_n - X| > \delta/2) < \varepsilon$ for all $n \geq N$. Also, let $E_n = \{|X_n - X| \leq \delta/2\}$ for all $n \geq N$. Thus we have for all $n \geq N$ that:

$$\begin{aligned}|F_n(t_0) - F(t_0)| = |P(X_n^{-1}((-\infty, t_0]) \cap E_n) + P(X_n^{-1}((-\infty, t_0]) \cap E_n^{\mathsf{C}})\\ - P(X^{-1}((-\infty, t_0 + \delta/2])) + P(X^{-1}((t_0, t_0 + \delta/2]))|\\ \leq |P(X_n^{-1}((-\infty, t_0]) \cap E_n) - P(X^{-1}((-\infty, t_0 + \delta/2]))|\\ + |P(X_n^{-1}((-\infty, t_0]) \cap E_n^{\mathsf{C}}) + P(X^{-1}((t_0, t_0 + \delta/2]))|\end{aligned}$$

Now because $P(E_n^{\mathsf{C}}) < \varepsilon$ and $|P(X^{-1}((t_0, t_0 + \delta/2]))| = |F(t_0 + \delta/2) - F(t_0)| < \varepsilon$, we know that $|P(X_n^{-1}((-\infty, t_0]) \cap E_n^{\mathsf{C}}) + P(X^{-1}((t_0, t_0 + \delta/2]))| < 2\varepsilon$.

Meanwhile, $X_n^{-1}((-\infty, t_0]) \cap E_n \subseteq X^{-1}((-\infty, t_0 + \delta/2])$. Therefore, we have that:

$$\begin{aligned}&|P(X_n^{-1}((-\infty, t_0]) \cap E_n) - P(X^{-1}((-\infty, t_0 + \delta/2]))|\\ &= P(X^{-1}((-\infty, t_0 + \delta/2]) - (E_n \cap X_n^{-1}((-\infty, t_0])))\\ &< P(E_n^{\mathsf{C}}) + P(E_n \cap (X^{-1}((-\infty, t_0 + \delta/2] - X_n^{-1}((-\infty, t_0])))) < \varepsilon + 0\end{aligned}$$

So $|F_n(t_0) - F(t_0)| < 3\varepsilon$ for all $n \geq N$. This proves that $F_n(t_0) \to F(t_0)$ as $n \to \infty$ and we are done. ■

Exercise 10.4: Let X, Y, and Z be positive independent random variables with a common distribution λ, and let $F(t) = \lambda((-\infty, t])$. The probability that the polynomial $Xt^2 + Yt + Z$ has real roots is $\int_0^\infty \int_0^\infty F(t^2/4s)\mathrm{d}\lambda(t)\mathrm{d}\lambda(s)$.

This is really easy. Note that the quadratic has real roots iff $Y^2 - 4XZ \geq 0$. Rewriting this, we have that the quadratic has real roots when $0 \leq Z \leq Y^2/4X$. Therefore:

$$\begin{aligned}P(Y^2 - 4XZ \geq 0) &= \int \chi_{\{t^2 - 4rs \geq 0\}} \mathrm{d}P_{X,Y,Z}(r,s,t)\\ &= \int \chi_{\{t^2 - 4rs \geq 0\}} \mathrm{d}\lambda^3(r,s,t)\\ &= \int_0^\infty \int_0^\infty \int_0^{t^2/4s} \mathrm{d}\lambda(r)\mathrm{d}\lambda(t)\mathrm{d}\lambda(s)\\ &= \int_0^\infty \int_0^\infty F(t^2/4s)\mathrm{d}\lambda(t)\mathrm{d}\lambda(s). \blacksquare\end{aligned}$$

Exercise 10.5: If X is a random variable with distribution $P_X = f(t)\mathrm{d}t$ where $f(t) = f(-t)$, then the distribution of X^2 is $P_{X^2} = t^{-1/2} f(t^{1/2})\chi_{(0,\infty)}(t)\mathrm{d}t$.

Note that for any $s \geq 0$, $\{X^2 \leq s\} = \{-\sqrt{s} \leq X \leq \sqrt{s}\}$. Hence:

$$\begin{aligned}P_{X^2}((-\infty, s]) = P(X^2 \leq s) = P(-\sqrt{s} \leq X \leq \sqrt{s})\\ = \int_{-\sqrt{s}}^{\sqrt{s}} \mathrm{d}P_X = \int_{-\sqrt{s}}^{\sqrt{s}} f(t)\mathrm{d}t = \int_0^{\sqrt{s}} 2f(t)\mathrm{d}t\\ = \int_0^s 2f(u^{1/2})(\tfrac{1}{2}u^{-1/2}\mathrm{d}u)\\ = \int_0^s u^{-1/2} f(u^{1/2})\mathrm{d}u\end{aligned}$$

Meanwhile if $s < 0$, then $P_{X^2}((-\infty, s]) = 0$. Hence, we have for all $s \in \mathbb{R}$ that:

$$P_{X^2}((-\infty, s]) = \int_{-\infty}^s t^{-1/2} f(t^{1/2})\chi_{(0,\infty)}\mathrm{d}t$$

This is enough by theorem 3.29 (see my math 240a paper notes) that to see that $P_{X^2} = t^{-1/2} f(t^{1/2})\chi_{(0,\infty)}\mathrm{d}t$. ■

I won't be reintroducing distributions or proving things that are already in my math 180a notes. That said, there is a result I'd like to give more clarity to. To start off, Folland denotes:

- δ_t (where $t \in \mathbb{R}$) is the point mass measure centered at E (meaning $\delta_t(E) = 1$ if $t \in E$ and 0 otherwise).
- $\beta_p^{*n} := \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k$ is the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$.
- $\lambda_a := e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} \delta_k$ is the Poisson distribution with parameter $a > 0$.

Exercise 10.8(c): $\beta_{a/n}^{*n} \to \lambda_a$ vaguely as $n \to \infty$.

Like in my answer to exercise 10.9, our strategy for this exercise will be to apply *proposition 7.19*. So, set $F_n(t) := \beta_{a/n}^{*n}((-\infty, t])$ and $F(t) := \lambda_a((-\infty, t])$ for all $t \in \mathbb{R}$. Then the set of discontinuities of F is precisely $\mathbb{N}$ (here I'm having $\mathbb{N}$ include 0). So, we just need to show that $F_n(t) \to F(t)$ whenever $t \notin \mathbb{N}$.

If $t < 0$, then it's clear that $0 = F_n(t) \to F(t) = 0$ as $n \to \infty$. Meanwhile, suppose that $N < t < N+1$ where $N \in \mathbb{N}$. Then $F(t) = e^{-a} \sum_{k=0}^{N} \frac{a^k}{k!}$.

Meanwhile $\frac{n(n-1)\cdots(n-k+1)}{n^k} \to 0$, $(1 - \frac{a}{n})^n \to e^{-a}$, and $(1-\frac{a}{n})^{-k} \to 1^{-k} = 1$ as $n \to \infty$ for all k. Therefore for $N < t < N+1$ we have:

$$F_n(t) = \sum_{k=0}^{N} \binom{n}{k} (a/n)^k (1 - a/n)^{n-k} \to \sum_{k=0}^{N} \frac{a^k e^{-a}}{k!} = F(t) \text{ as } n \to \infty. \blacksquare$$

As a side note, since $\|\beta_{a/n}^{*n}\| = 1 = \|\lambda_a\|$ for all n since all of them are probability measures, we can apply *Exercise 7.26* to say that $\int f \mathrm{d}\beta_{a/n}^{*n} \to \int f \mathrm{d}\lambda_a$ as $n \to \infty$ for all $f \in BC(\mathbb{R})$.

Exercise 10.10: (The Moment Convergence Theorem)

Let $X_1, X_2, \ldots, X$ be random variables such that $P_{X_n} \to P_X$ vaguely and $\sup_{n \in \mathbb{N}} E(|X_n|^r) < \infty$ for some $r > 0$. Then $E(|X_n|^s) \to E(|X|^s)$ for all $s \in (0, r)$, and if $s \in \mathbb{N}$ also, then $E((X_n)^s) \to E(X^s)$.

Proof:
Fix $C \geq \sup_{n \in \mathbb{N}} E(|X_n|^r)$. Then by Chebyshev's inequality (see my math 240b notes), we know that $P(|X_n| > \alpha) \leq \frac{E(|X_n|^r)}{\alpha^r}$ for all $\alpha > 0$ and $n \in \mathbb{N}$. And hence for all $n \in \mathbb{N}$ we have that $P(|X_n| > \alpha) \leq C/\alpha^r$ when $\alpha > 0$.

Next, given any $\alpha > 0$, let $\phi_\alpha \in C_c(\mathbb{R}, [0,1])$ such that $\phi_\alpha(t) = 1$ when $|t| \leq \alpha$ and $\phi_\alpha(t) = 0$ when $|t| > \alpha + 1$. Importantly, $\phi_\alpha(t)|t|^s$ (and $\phi_\alpha(t)t^s$ when s is an integer) is in $C_c(\mathbb{R})$. Thus by the vague convergence of the P_{X_n} we know for all $s \in (0, r)$ that $\int \phi_\alpha(t)|t|^s \mathrm{d}P_{X_n} \to \int \phi_\alpha(t)|t|^s \mathrm{d}P_X$ as $n \to \infty$ (and similarly if s is also an integer we have that $\int \phi_\alpha(t)t^s \mathrm{d}P_{X_n} \to \int \phi_\alpha(t)t^s \mathrm{d}P_X$).

Meanwhile, since $s < r$, we know by Hölder's inequality that:

$$\begin{aligned} 0 \leq \int (1 - \phi_\alpha(t))|t|^s \mathrm{d}P_{X_n} &\leq \left[\int (1-\phi_\alpha(t))^{r/(r-s)} \mathrm{d}P_{X_n}\right]^{(r-s)/r} \cdot \left[\int \left(|t|^s\right)^{r/s} \mathrm{d}P_{X_n}\right]^{s/r} \\ &\leq \left[\int \chi_{\{|t|>a\}} \mathrm{d}P_{X_n}\right]^{(r-s)/r} \cdot \left[E(|X_n|^r)\right]^{s/r} \\ &\leq (P(|X_n| > a))^{(r-s)/r} \cdot C^{s/r} \leq \left(\tfrac{C^{1/r}}{\alpha}\right)^{r-s} C^{s/r} = \tfrac{C}{\alpha^{r-s}}. \end{aligned}$$

Now $\int |t|^s \mathrm{d}P_{X_n} = \int \phi_\alpha(t)|t|^s \mathrm{d}P_{X_n} + \int (1-\phi_\alpha(t))|t|^s \mathrm{d}P_{X_n}$. So, we can say that:

$$\begin{aligned}\int_{-\alpha}^{\alpha} |t|^s \mathrm{d}P_X \leq \int \phi_\alpha(t)|t|^s \mathrm{d}P_X \leq \liminf_{n\to\infty} \int |t|^s \mathrm{d}P_{X_n} \\ \leq \limsup_{n\to\infty} \int |t|^s \mathrm{d}P_{X_n} \leq \int \phi_\alpha(t)|t|^s \mathrm{d}P_X + \tfrac{C}{\alpha^{r-s}} \\ \leq \int_{-\alpha-1}^{\alpha+1} |t|^s \mathrm{d}P_X + \tfrac{C}{\alpha^{r-s}}\end{aligned}$$

And by taking $\alpha \to \infty$, we know by the monotone convergence theorem and the fact that $\frac{1}{\alpha^{r-s}} \to 0$ that $\lim_{n\to\infty} E(|X|^s) = \lim_{n\to\infty} \int |t|^s \mathrm{d}P_{X_n} = \int |t|^s \mathrm{d}P_X = E(|X|^s)$ for all $s \in (0, r)$.

A notable consequence of this is that $X \in L^s(P)$ for all $s \in (0, r)$. After all, by Folland proposition 6.12 (see my math 240b notes):

$$\|X_n\|_s \leq \|X_n\|_r P(X)^{1/s - 1/r} = \|X_n\|_r \leq C^{1/r} \text{ for all } n \in \mathbb{N}.$$

Thus $(E(|X_n|^s))_{n\in\mathbb{N}}$ is a sequence bounded above by $C^{s/r}$, and since $E(|X_n|^s) \to E(|X|^s)$, we have that $E(|X|^s) \leq C^{s/r} < \infty$ for all $s \in (0, r)$. This now let's us address the special case that s is an integer. After all, note that:

$$\begin{aligned}\limsup_{n\to\infty} |\textstyle\int t^s \mathrm{d}P_{X_n} - \int t^s \mathrm{d}P_X| = \limsup_{n\to\infty} |\int t^s \phi_\alpha(t) \mathrm{d}P_{X_n} + \int t^s(1-\phi_\alpha(t))\mathrm{d}P_{X_n} \\ - \textstyle\int t^s \phi_\alpha(t)\mathrm{d}P_X - \int t^s(1-\phi_\alpha(t))\mathrm{d}P_X| \\ \leq 0 + \limsup_{n\to\infty} |\textstyle\int t^s(1-\phi_\alpha(t))\mathrm{d}P_{X_n}| \\ + \limsup_{n\to\infty} |\textstyle\int t^s(1-\phi_\alpha(t))\mathrm{d}P_X|\end{aligned}$$

Now $0 \leq |\int (1-\phi_\alpha(t))t^s \mathrm{d}P_{X_n}| \leq \int (1-\phi_\alpha(t))|t|^s \mathrm{d}P_{X_n} \leq \frac{C}{\alpha^{r-s}}$ for all n.

Meanwhile pick $q \in (s, r)$ and note by Chebyshev's inequality that $P(|X| > \alpha) \leq \frac{C^{q/r}}{\alpha^q}$ for all $\alpha > 0$. And then by similar reasoning to earlier we can show using Hölder's inequality that:

$$\begin{aligned}0 \leq |\textstyle\int (1-\phi_\alpha(t))t^s \mathrm{d}P_X| \leq \int (1-\phi_\alpha t)|t|^s \mathrm{d}P_X \\ \leq (P(|X| > \alpha))^{(q-s)/q} \cdot [E(|X|^q)]^{s/q} \\ \leq (\tfrac{C^{1/r}}{\alpha})^{q-s} \cdot [C^{q/r}]^{s/q} = \tfrac{1}{\alpha^{q-s}} C^{(\frac{q}{r} - \frac{s}{r} + s/r)} = \tfrac{1}{\alpha^{q-s}} C^{q/r}\end{aligned}$$

Thus $\limsup_{n\to\infty} |\int t^s \mathrm{d}P_{X_n} - \int t^s \mathrm{d}P_X| \leq \frac{C}{\alpha^{r-s}} + \frac{1}{\alpha^{q-s}} C^{q/r}$ and the latter goes to zero as $\alpha \to \infty$. ■

9/16/2025

In the next section of Folland, he writes a lot of theorems involving infinite sequences of independent random variables. Now unfortunately actually constructing those sequences requires some theorems I skipped over earlier. So for now just I'm just going to ignore the issue. However, I'll come back and address this issue later on *page 215*.

Theorem 10.9: (The Weak Law of Large Numbers)

Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of independent L^2 random variables with means $\{\mu_j\}_{j\in\mathbb{N}}$ and variances $\{\sigma_j^2\}$. If $n^{-2}\sum_{j=1}^{n}\sigma_j^2 \to 0$ as $n\to\infty$, then $n^{-1}\sum_{j=1}^{n}(X_j-\mu_j)\to 0$ in probability as $n\to\infty$.

Proof:

$n^{-1}\sum_{j=1}^{n}(X_j-\mu_j)$ has mean $n\cdot 0 = 0$ and variance:

$$\sum_{j=1}^{n}(\sigma^2(\tfrac{X_j-\mu_j}{n})) = \sum_{j=1}^{n}(E(\tfrac{(X_j-\mu_j)^2}{n^2}-0)) = \tfrac{1}{n^2}\sum_{j=1}^{n}\sigma^2(X_j).$$

Hence by Chebyshev's inequality, for any $\varepsilon > 0$ we have that:

$$P(|n^{-1}\sum_{j=1}^{n}(X_j-\mu_j)| > \varepsilon) \leq \tfrac{1}{\varepsilon^2}\left(\tfrac{1}{n^2}\sum_{j=1}^{n}\sigma^2(X_j)\right)$$

And since $n^{-2}\sum_{j=1}^{n}\sigma_j^2 \to 0$ as $n\to\infty$, we know $P(|n^{-1}\sum_{j=1}^{n}(X_j-\mu_j)| > \varepsilon)\to 0$ as $n\to\infty$. ■

As a side note: if $\sup_{j\in\mathbb{N}}\sigma_j^2 \leq C < \infty$, then $n^{-2}\sum_{j=1}^{n}\sigma_j^2 \leq n^{-1}C \to 0$ as $n\to\infty$. Hence, this theorem is especially useful for the case where all the X_j are identically distributed.

Before continuing, we need an exercise.

Exercise 10.3(b): Suppose that $\{E_\alpha\}_{\alpha\in A}$ is a collection of independent events in Ω. Then so is $\{F_\alpha\}_{\alpha\in A}$ where each F_α is equal to either E_α or E_α^{C}.

Proof:

I already showed in my math 180a notes that this is true whenever A is finite. So I won't be repeating that proof. In the general case, for any $n\in\mathbb{N}$ and $\alpha_1,\ldots,\alpha_n\in A$ we know that the subcollection $\{E_{\alpha_1},\ldots,E_{\alpha_n}\}$ is also a collection of independent events. And so we can apply the finite case proved in my math 180a notes to say that $F_{\alpha_1},\ldots,F_{\alpha_n}$ are independent events, and in turn that $P(\bigcap_{j=1}^{n}F_{\alpha_j}) = \prod_{j=1}^{n}P(F_{\alpha_j})$. ■

Given a sequence $\{A_n\}_{n\in\mathbb{N}}$ of measurable sets / events, we define:

$\limsup A_n := \bigcap_{n\in\mathbb{N}}(\bigcup_{k=n}^{\infty}A_k)$ and $\liminf A_n := \bigcup_{n\in\mathbb{N}}(\bigcap_{k=n}^{\infty}A_k)$

For some intuition:

- $x\in\liminf A_n \iff \exists N>0\ s.t.\ \forall n\geq N,\ \ x\in A_n$
  (i.e. x is eventually in each A_n),
- $x\in\limsup A_n \iff \forall N>0,\ \ \exists n\geq N\ s.t.\ x\in A_n$
  (i.e. x is frequently in the A_n).

Clearly $\liminf A_n \subseteq \limsup A_n$.

The Borel-Cantelli Lemma: Let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of events.

(a) If $\sum_{n=1}^{\infty}P(A_n) < \infty$, then $P(\limsup A_n) = 0$.

(b) If the A_n are all independent and $\sum_{n=1}^{\infty}P(A_n) = \infty$, then $P(\limsup A_n) = 1$.

Proof:
Part (a) is simple. $P(\limsup A_n) \leq P(\bigcup_{k=n}^{\infty} A_k) \leq \sum_{k=n}^{\infty} A_k$ for all $n \in \mathbb{N}$. And if $\sum_{n=1}^{\infty} P(A_n) < \infty$, then we know that $\sum_{k=n}^{\infty} A_k \to 0$ as $n \to \infty$.

As for part (b), note that $P(\limsup A_n) = 1$ if and only if:

$$P((\limsup A_n)^{\mathsf{C}}) = P((\bigcap_{n\in\mathbb{N}}(\bigcup_{k=n}^{\infty} A_k))^{\mathsf{C}}) = P(\bigcup_{n\in\mathbb{N}}(\bigcup_{k=n}^{\infty} A_k)^{\mathsf{C}}) = P(\bigcup_{n\in\mathbb{N}}(\bigcap_{k=n}^{\infty} A_k^{\mathsf{C}})) = 0$$

So, it suffices to show $P(\bigcap_{k=n}^{\infty} A_k^{\mathsf{C}}) = 0$ for all $n \in \mathbb{N}$. But fortunately all the A_k^{C} are independent according to exercise 10.3(b) on the last page. Thus we have for all $K > n$ that $P(\bigcap_{k=n}^{K} A_k^{\mathsf{C}}) = \prod_{k=n}^{K}(1 - P(A_k))$. And by applying the monotonicity of measures, we have that:

$$P(\textstyle\bigcap_{k=n}^{\infty} A_k^{\mathsf{C}}) = \lim_{K\to\infty} P(\bigcap_{k=n}^{K} A_k^{\mathsf{C}}) = \lim_{K\to\infty} \prod_{k=n}^{K}(1 - P(A_k)).$$

But now note that $1 - t \leq e^{-t}$ for all $t \in \mathbb{R}$.

My apartment mate immediately recognized this trick so I guess it's commonly used. Also, it probably would have been efficient to use when I was doing the exercise on *pages 51-52* last Christmas. Anyways, it's easy to see that this inequality is true by just looking at the derivative of $e^{-t} - (1 - t)$ and seeing that the function attains a global minimum at $t = 0$.

Hence:

$$\begin{aligned}\lim_{K\to\infty} \textstyle\prod_{k=n}^{K}(1 - P(A_k)) &\leq \limsup_{K\to\infty} \textstyle\prod_{k=n}^{K} e^{-P(A_k)} \\ &= \limsup_{K\to\infty} \exp(-\textstyle\sum_{k=n}^{K} P(A_k)).\end{aligned}$$

And since $\sum_{k=n}^{\infty} P(A_k) = +\infty$ for each n, we thus have that:

$$P(\textstyle\bigcap_{k=n}^{\infty} A_k^{\mathsf{C}}) \leq \limsup_{K\to\infty} \exp(-\sum_{k=n}^{K} P(A_k)) = 0. \blacksquare$$

Kolmogorov's Inequality: Let $X_1, \ldots, X_n$ be independent random variables with mean 0 and variances $\sigma_1^2, \ldots, \sigma_n^2$, and let $S_k = X_1 + \ldots + X_k$ for all $k \in \{1, \ldots, n\}$. Then for any $\varepsilon > 0$, we have that:

$$P(\max_{1\leq k\leq n} |S_k| \geq \varepsilon) \leq \varepsilon^{-2} \sum_{k=1}^{n} \sigma_k^2.$$

Proof:
For each k let A_k be the set where $|S_j| < \varepsilon$ for $j < k$ and $|S_k| \geq \varepsilon$. That way all the A_k are disjoint and their union is the set where $\max_{1\leq k\leq n} |S_k| \geq \varepsilon$. Also, note that $P(A_k) = E(\chi_{A_k}) \leq \varepsilon^{-2} E(\chi_{A_k} S_k^2)$ since $S_k^2 \geq \varepsilon^2$ on A_k. Therefore, we know that:

$$P(\max_{1\leq k\leq n} |S_k| \geq \varepsilon) = \textstyle\sum_{k=1}^{n} P(A_k) \leq \varepsilon^{-2} \sum_{k=1}^{n} E(\chi_{A_k} S_k^2).$$

Now, I kinda hate how Folland does the next bit because he just expands out his expression in a way that looks completely out of nowhere. So here's my attempt at explaining what insights I think led to Kolmogorov or Folland to doing the following reasoning.

Note that we already know that $E(S_n^2) = \sum_{k=1}^{n} \sigma_k^2$ by *corollary 10.6* since $E(S_n^2) = \sigma^2(S_n)$ on account of the fact that $E(S_n) = 0$. So, if we could just bound $\sum_{k=1}^{n} E(\chi_{A_k} S_k^2)$ from above by $E(S_n^2)$ then we'd be done. Luckily, note that $S_n - S_k = X_{k+1} + \ldots + X_n$. Thus, by *proposition 10.2* we know that S_k and $S_n - S_k$ are independent.

Going a step further, note that $\chi_{A_k} S_k$ can be rewritten as $f \circ (S_1, \ldots, S_k)$ where $f(t_1, \ldots, t_k) = t_k \chi_{(-\varepsilon,\varepsilon)^{k-1} \times (-\varepsilon,\varepsilon)^{\mathsf{C}}}$ is a Borel measurable function from $\mathbb{R}^k$ to $\mathbb{R}$. In turn there is clearly another Borel measurable function $g : \mathbb{R}^k \to \mathbb{R}^k$ such that $g \circ (X_1, \ldots, X_k) = (S_1, \ldots, S_k)$. Thus $\chi_{A_k} S_k = f \circ g \circ (X_1, \ldots, X_k)$. And by applying *proposition 10.2*, we have that $\chi_{A_k} S_k$ and $S_n - S_k$ are independent.

Importantly, $E(S_n - S_k) = 0$ for all k. Thus:
$E(\chi_{A_k} S_k (S_n - S_k)) = E(\chi_{A_k} S_k) E(S_n - S_k) = E(\chi_{A_k} S_k) \cdot 0 = 0$ for all k.

And hopefully it's now obvious how to proceed.

Observe that:

$$\begin{aligned} E(S_n^2) \geq \sum_{k=1}^{n} E(\chi_{A_k} S_n^2) &= \sum_{k=1}^{n} E(\chi_{A_k}(S_n^2 + S_k^2 - S_k^2 + 2S_k S_n - 2S_k S_n)) \\ &= \sum_{k=1}^{n} E(\chi_{A_k}(S_k^2 + 2S_k S_n + (S_n - S_k)^2)) \\ &= \sum_{k=1}^{n} E(\chi_{A_k}(3S_k^2 + 2S_k S_n - 2S_k S_k + (S_n - S_k)^2)) \\ &\geq \sum_{k=1}^{n} [3E(\chi_{A_k} S_k^2) + 2E(\chi_{A_k} S_k (S_n - S_k)) + 0] \\ &= 3 \sum_{k=1}^{n} E(\chi_{A_k} S_k^2). \blacksquare \end{aligned}$$

As a side note: I realize I proved a tighter bound then what Folland did. It doesn't help that Folland made a mistake which I fixed while doing the above manipulations. But anyways, I actually showed that:

$$P(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon) \leq \tfrac{1}{3\varepsilon^2} \sum_{k=1}^{n} \sigma_k^2.$$

<u>Kolmogorov's Strong Law of Large Numbers:</u> If $\{X_n\}_{n \in \mathbb{N}}$ is a sequence of independent L^2 random variables with means $\{\mu_n\}_{n \in \mathbb{N}}$ and variances $\{\sigma_n^2\}$ such that $\sum_{n \in \mathbb{N}} n^{-2} \sigma_n^2 < \infty$, then $n^{-1} \sum_{j=1}^{n} (X_j - \mu_j) \to 0$ a.s. as $n \to \infty$.

Proof:

Let $S_n = \sum_{k=1}^{n} (X_k - \mu_k)$ for all n. Next, given any $\varepsilon > 0$, for each $k \in \mathbb{N}$ let A_k be the set where $n^{-1}|S_n| \geq \varepsilon$ for some n such that $2^{k-1} \leq n < 2^k$. Then we know for all outcomes in A_k that $|S_n| \geq \varepsilon 2^{k-1}$ for some $n < 2^k$. And thus by Kolmogorov's inequality, we have that:

$$P(A_k) \leq \tfrac{1}{3} (\varepsilon 2^{k-1})^{-2} \sum_{n=1}^{2^k} \sigma_n^2$$

In turn we have (and note the usage of Tonelli's theorem at the end) that:

$$\sum_{k=1}^{\infty} P(A_k) \leq \sum_{k=1}^{\infty} [\tfrac{1}{3}(\varepsilon 2^{k-1})^{-2} \sum_{n=1}^{2^k} \sigma_n^2] = \tfrac{4}{3\varepsilon^2} \sum_{k=1}^{\infty} \sum_{n=1}^{2^k} 2^{-2k} \sigma_n^2 = \tfrac{4}{3\varepsilon^2} \sum_{n=1}^{\infty} (\sum_{k \geq \log_2(n)} 2^{-2k}) \sigma_n^2.$$

Also note that:

$$\sum_{k \geq \log_2(n)} 2^{-2k} = 2^{-2\lceil \log_2(n) \rceil} \sum_{k=0}^{\infty} (2^{-2})^k = 2^{-2\lceil \log_2(n) \rceil} \cdot \tfrac{4}{3} \leq 2^{-2\log_2(n)} \cdot \tfrac{4}{3} = \tfrac{4}{3n^2}.$$

Therefore, $\sum_{k=1}^{\infty} P(A_k) \leq \frac{16}{9\varepsilon^2} \sum_{n=1}^{\infty} n^{-2}\sigma_n^2 < \infty$.

By the Borel-Cantelli lemma, we now know that $P(\limsup A_k) = 0$. But note that $\limsup A_k$ consists precisely of every outcome where $n^{-1}|S_n| \geq \varepsilon$ for infinitely many n. Therefore, $P(\limsup_{n\to\infty} n^{-1}|S_n| \geq \varepsilon) = 0$ for all $\varepsilon > 0$. And consequently, we have that:

$$\begin{aligned} P(n^{-1}S_n \not\to 0) &= P(\bigcup_{m\in\mathbb{N}} \{\limsup_{n\to\infty} n^{-1}|S_n| \geq {}^1\!/\!{}_m\}) \\ &\leq \sum_{m\in\mathbb{N}} P(\limsup_{n\to\infty} n^{-1}|S_n| \geq {}^1\!/\!{}_m) = 0. \blacksquare \end{aligned}$$

As a side note: it's still clear that if $\sup_{n\in\mathbb{N}} \sigma_n^2 \leq C < \infty$, then $\sum_{n=1}^{\infty} n^{-2}\sigma_j^2 \leq C \sum_{n=1}^{\infty} n^{-2} < \infty$. In particular, this means that this theorem applies when all the X_j are identically distributed.

Khinchine's Strong Law of Large Numbers: If $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of independent identically distributed L^1 random variables with mean μ, then $n^{-1}\sum_{j=1}^{n} X_j \to \mu$ a.s. as $n \to \infty$.

Proof:

Let λ be the common distribution of the X_n and note that $\int |t| \mathrm{d}\lambda(t) < \infty$. Next, for each j let $Y_j = X_j$ on the set where $|X_j| \leq j$ and $Y_j = 0$ elsewhere. Then:

$$\sum_{j=1}^{\infty} P(Y_j \neq X_j) = \sum_{j=1}^{\infty} P(|X_j| > j) = \sum_{j=1}^{\infty} \lambda(\{|t| > j\}) = \sum_{j=1}^{\infty}\sum_{k=j}^{\infty} \lambda(\{k < |t| \leq k+1\}).$$

By swapping the order of summation (which we can do via Tonelli's theorem), we have that:

$$\begin{aligned} \sum_{j=1}^{\infty}\sum_{k=j}^{\infty} \lambda(\{k < |t| \leq k+1\}) &= \sum_{k=1}^{\infty}\sum_{j=1}^{k} \lambda(\{k < |t| \leq k+1\}). \\ &= \sum_{k=1}^{\infty} k \cdot \lambda(\{k < |t| \leq k+1\}) \leq \int |t| \mathrm{d}\lambda(t) < \infty. \end{aligned}$$

Thus by the Borel-Cantelli lemma, if $A_j = \{Y_j \neq X_j\}$, then $P(\limsup A_j) = 0$. Or in other words, $P((\limsup A_j)^{\mathsf{C}}) = 1$. This proves that almost surely there exists $J > 0$ such that $X_j = Y_j$ for all $j > J$. And luckily from this we can now conclude that it suffices to show $n^{-1}\sum_{j=1}^{n} Y_j \to \mu$ a.s. in order to show that $n^{-1}\sum_{j=1}^{n} X_j \to \mu$ a.s.

Why?

Suppose that $n^{-1}\sum_{j=1}^{n} Y_j \to \mu$, that there exists J such that $Y_j = X_j$ for all $j > J$, and that all the X_j are finite-valued. Then for any $n > J$ we have that:

$$n^{-1}\sum_{j=1}^{n} X_j = n^{-1}(\sum_{j=1}^{J} X_j) + n^{-1}(\sum_{j=J+1}^{n} Y_j).$$

Now it's clear that $n^{-1}(\sum_{j=1}^{J} X_j) \to 0$ as $n \to \infty$ since $\sum_{j=1}^{J} X_j$ is fixed. Also by similar reasoning we know that $n^{-1}(\sum_{j=1}^{J} Y_j) \to 0$ as $n \to \infty$. And this is enough to say that $\lim_{n\to\infty} n^{-1}(\sum_{j=J+1}^{n} Y_j) = \lim_{n\to\infty} n^{-1}(\sum_{j=1}^{n} Y_j) = \mu$ since:

$$\begin{aligned} \lim_{n\to\infty} n^{-1}(\sum_{j=1}^{n} Y_j) &= \lim_{n\to\infty} n^{-1}\sum_{j=1}^{J} Y_j + \lim_{n\to\infty} n^{-1}(\sum_{j=J+1}^{n} Y_j) \\ &= 0 + \lim_{n\to\infty} n^{-1}(\sum_{j=J+1}^{n} Y_j) \end{aligned}$$

And as a side note since it took me a while to process this, if you were wondering why it wasn't trivial that $Y_j - X_j \to 0$ a.s., start by realizing that if the range of the X_n includes values greater than $j+1$, then $\{Y_j = X_j\} \not\subseteq \{Y_{j+1} = X_{j+1}\}$.

Now $\sigma^2(Y_n) \leq E(Y_n^2) = \int_{\{|t| \leq n\}} t^2 \mathrm{d}\lambda(t)$. Therefore:

$$\sum_{n=1}^{\infty} n^{-2}\sigma^2(Y_n) = \sum_{n=1}^{\infty}\sum_{j=1}^{n} n^{-2} \int_{\{j-1<|t|\leq j\}} t^2 \mathrm{d}\lambda(t) \leq \sum_{n=1}^{\infty}\sum_{j=1}^{n} jn^{-2} \int_{\{j-1<|t|\leq j\}} |t| \mathrm{d}\lambda(t)$$

And by using Tonelli's theorem to swap the order of summation, we can say that:

$$\begin{aligned}\sum_{n=1}^{\infty}\sum_{j=1}^{n} jn^{-2} \int_{\{j-1<|t|\leq j\}} |t| \mathrm{d}\lambda(t) &= \sum_{j=1}^{\infty}\sum_{n=j}^{\infty} jn^{-2} \int_{\{j-1<|t|\leq j\}} |t| \mathrm{d}\lambda(t) \\ &= \sum_{j=1}^{\infty}[j \cdot \left(\sum_{n=j}^{\infty} n^{-2}\right) \cdot \int_{\{j-1<|t|\leq j\}} |t| \mathrm{d}\lambda(t)]\end{aligned}$$

Next note that if $j > 1$, then: $\sum_{n=j}^{\infty} n^{-2} \leq \int_{j-1}^{\infty} x^{-2}\mathrm{d}x = \frac{1}{2(j-1)}$. Furthermore, note that when $j > 1$: $(\frac{1}{j})/(\frac{1}{2(j-1)}) = \frac{2j-2}{j} = 2 - \frac{2}{j} \geq 1$. Thus $\frac{1}{2(j-1)} < \frac{1}{j}$ for all $j > 1$. Meanwhile, $\sum_{n=1}^{\infty} n^{-2}$ famously equals $\frac{\pi^2}{6}$ which is easily checked to be less than $2 = 2 \cdot (1)^{-1}$. Thus, we can conclude that:

$$\begin{aligned}\sum_{j=1}^{\infty}[j \cdot \left(\sum_{n=j}^{\infty} n^{-2}\right) \cdot \int_{\{j-1<|t|\leq j\}} |t| \mathrm{d}\lambda(t)] &\leq \sum_{j=1}^{\infty}[j \cdot \tfrac{2}{j} \cdot \int_{\{j-1<|t|\leq j\}} |t| \mathrm{d}\lambda(t)] \\ &= 2 \int |t| \mathrm{d}\lambda(t) < \infty\end{aligned}$$

Therefore, if $\mu_j := E(Y_j)$ for all j, then we know by Kolmogorov's strong law of large numbers that $n^{-1}\sum_{j=1}^{n}(Y_j - \mu_j) = \left(n^{-1}\sum_{j=1}^{n} Y_j\right) - \left(n^{-1}\sum_{j=1}^{n} \mu_j\right) \to 0$ a.s. as $n \to \infty$.

But now note by the dominated convergence theorem that:

$$\mu_j = \int_{j}^{j} t \mathrm{d}\lambda(t) \to \int_{-\infty}^{\infty} t\mathrm{d}\lambda(t) = \mu \text{ as } j \to \infty.$$

Therefore, by exercise 10.12 below, we know that $n^{-1}\sum_{j=1}^{n} \mu_j \to \mu$ as $n \to \infty$. And this proves that $n^{-1}\sum_{j=1}^{n} Y_j \to \mu$ a.s. as $n \to \infty$. ■

Exercise 10.12: Suppose $\mathcal{X}$ is a normed vector space and $(a_n)_{n\in\mathbb{N}}$ is a sequence in $\mathcal{X}$ such that $a_n \to a$ for some $a \in \mathcal{X}$. Then $n^{-1}\sum_{j=1}^{n} a_j \to a$ as well.

Proof:

Suppose J is any integer and note that for any $n > J$:

$$\begin{aligned}\|(n^{-1}\textstyle\sum_{j=1}^{n} a_j) - a\| &\leq n^{-1}\textstyle\sum_{j=1}^{n}\|a_j - a\| \\ &\leq \tfrac{J}{n}\max_{1\leq j\leq J}\|a_j - a\| + \tfrac{n-J}{n}\sup_{j>J}\|a_j - a\|\end{aligned}$$

Thus by taking $n \to \infty$ we clearly have that:

$$\limsup_{n\to\infty}\|(n^{-1}\textstyle\sum_{j=1}^{n} a_j) - a\| \leq \sup_{j>J}\|a_j - a\|.$$

And since $a_j \to a$ as $j \to \infty$, we know that $\sup_{j>J}\|a_j - a\| \to 0$ as $J \to \infty$. Thus, taking $J \to \infty$ shows that $\limsup_{n\to\infty}\|(n^{-1}\sum_{j=1}^{n} a_j) - a\| = 0$. Or in other words, $n^{-1}\sum_{j=1}^{n} a_j \to a$ as $n \to \infty$. ■

The physical consequence of the laws of large numbers is that as a person plays more games of chance (that are independent of each other), their average outcome will approach the expected average outcome. Or to put into other words, a person's luck will tend to balance out the more independent games of chance they play (although that does not mean that past games of chance have any predictive power over future games of chance).

9/18/2025

I want to start today by doing some exercises to expand on some of the earlier results. Firstly, here is a proof that the hypotheses for the weak law of large numbers are weaker than the hypotheses for Kolmogorov's strong law of large numbers.

Exercise 10.11: Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in $[0,\infty)$ such that $\sum_{n=1}^{\infty} n^{-2}a_n < \infty$. Then:

$$\lim_{n\to\infty} n^{-2}\sum_{j=1}^{n} a_j = 0.$$

Proof:
For any $N \in \mathbb{N}$ and $n > N$, we have that:

$$n^{-2}\sum_{j=1}^{n} a_j = n^{-2}\left(\sum_{j=1}^{N} a_j\right) + \left(\sum_{j=N+1}^{n} n^{-2}a_j\right) \leq n^{-2}\left(\sum_{j=1}^{N} a_j\right) + \left(\sum_{j=N+1}^{n} j^{-2}a_j\right)$$

Then taking $n \to \infty$ we have that:

$$0 \leq \limsup_{n\to\infty} n^{-2}\sum_{j=1}^{n} a_j \leq 0 + \sum_{n=N+1}^{\infty} n^{-2}a_n.$$

And since $\sum_{n=1}^{\infty} n^{-2}a_n < \infty$, we know that $\sum_{n=N+1}^{\infty} n^{-2}a_n \to 0$ as $N \to \infty$. Hence, we've shown that $\limsup_{n\to\infty} n^{-2}\sum_{j=1}^{n} a_j = 0$. And since $\liminf_{n\to\infty} n^{-2}\sum_{j=1}^{n} a_j \geq 0$, this proves that $\lim_{n\to\infty} n^{-2}\sum_{j=1}^{n} a_j = 0$. ■

Also, we can actually weaken our hypotheses for the weak law of large numbers.

Exercise 10.13: The weak law of large numbers remains valid if the hypothesis of independence is replaced by the hypothesis that $E[(X_j-\mu_j)(X_k-\mu_k)] = 0$ for all $j \neq k$.

We still have that $E(n^{-1}\sum_{j=1}^{n}(X_j-\mu_j)) = n^{-1}\sum_{j=1}^{n} E(X_j-\mu_j) = 0$. And in turn:

$$\begin{aligned}\sigma^2(n^{-1}\textstyle\sum_{j=1}^{n}(X_j-\mu_j)) &= E((n^{-1}\textstyle\sum_{j=1}^{n}(X_j-\mu_j)-0)^2)\\ &= \tfrac{1}{n^2}E((\textstyle\sum_{j=1}^{n}(X_j-\mu_j))^2)\\ &= \tfrac{1}{n^2}\left(\textstyle\sum_{j=1}^{n}\sigma_j^2 + \sum_{j\neq k} E[(X_j-\mu_j)(X_k-\mu_k)]\right)\\ &= \tfrac{1}{n^2}\textstyle\sum_{j=1}^{n}\sigma_j^2 + 0.\end{aligned}$$

And now the rest of the proof is identical to our original proof. ■

This hypothesis is strictly weaker. For example, it holds as long as the random variables are pairwise independent, and it's possible for random variables to be pairwise independent but not independent.

On a complete tangent, here's another interesting result.

Exercise 10.14: If $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of independent random variables such that $E(X_n) = 0$ for all n and $\sum_{n=1}^{\infty}\sigma^2(X_n) < \infty$, then $\sum_{n=1}^{\infty} X_n$ converges almost surely.

Proof:
Fix $\varepsilon > 0$ and let $N \in \mathbb{N}$. Also denote $S_n = \sum_{j=1}^n X_j$ for all $n \in \mathbb{N}$. Then note that by Kolmogorov's inequality, we have for all $m > N$ that:

$$P(\max_{N<n\leq m} |S_n - S_N| \geq \varepsilon/2) \leq 4\varepsilon^{-2} \sum_{n=N+1}^{m} \sigma^2(X_n).$$

And by taking $m \to \infty$, we get that:

$$P(\exists n > N \ s.t. \ |S_n - S_N| \geq \varepsilon/2) \leq 4\varepsilon^{-2} \sum_{n=N+1}^{\infty} \sigma^2(X_n).$$

But now since $\sum_{n=1}^\infty \sigma^2(X_n) < \infty$, we know that $\sum_{n=N+1}^\infty \sigma^2(X_k) \to 0$ as $N \to \infty$. This is enough to let us conclude that there almost surely exists some $N \in \mathbb{N}$ such that $|S_n - S_N| < \varepsilon/2$ for all $n > N$.

Why is this?

$$\begin{aligned} P(\forall N \in \mathbb{N},\ \ \exists n > N \ s.t. \ |S_n - S_N| \geq \varepsilon/2) & \\ &= P(\bigcap_{N\in\mathbb{N}} \{\exists n > N \ s.t. \ |S_n - S_N| \geq \varepsilon/2\}) \\ &\leq \inf_{N\in\mathbb{N}} P(\exists n > N \ s.t. \ |S_n - S_N| \geq \varepsilon/2) \\ &\leq \inf_{N\in\mathbb{N}} 4\varepsilon^{-2} \sum_{n=N+1}^{\infty} \sigma^2(X_n) = 0. \end{aligned}$$

Also note that if $m > n > N$, $|S_m - S_N| < \varepsilon/2$, and $|S_n - S_N| < \varepsilon/2$, then:

$$|S_m - S_n| \leq |S_m - S_N| + |s_n - S_N| < \varepsilon.$$

Hence, we have successfully proven that for any $\varepsilon > 0$, there almost surely exists some $N \in \mathbb{N}$ such that for all $m > n > N$ we have that $|S_m - S_n| < \varepsilon$. And in turn by considering a countable sequence $(\varepsilon_k)_{k\in\mathbb{N}}$ in $(0, \infty)$ converging to 0 we can thus easily see that the partial sums of the X_n almost surely satisfy the Cauchy criterion. ■

Corollary (still part of the prior exercise): If the plus and minus signs in $\sum_{n=1}^\infty \pm n^{-1}$ are determined by successive tosses of a fair coin, the resulting series converges almost surely.

In this case, we have for each n that $X_n = +n^{-1}$ 50% of the time and $-n^{-1}$ the other 50% of the time. And since each X_n is a simple function, we can easily evaluate that $E(X_n) = 0$ and $\sigma^2(X_n) = E((X_n - 0)^2) = n^{-2}$. Hence, it's clear that we can apply what we just proved to this sequence of X_n. ■

Now to finish off today, I want to do (and give a little commentary) to an exercise that is heavily relevant to statistics.

Exercise 10.17: A collection or "population" of N objects may be considered a sample space in which each object individually has probability N^{-1}. Also let X be a random variable on this space with mean μ and variance σ^2. A central goal of statistics is to determine what μ and σ^2 are by randomly sampling our population and measuring X for each object sampled.

To mathematically model this process, we imagine generating a sequence $\{X_n\}_{n\in\mathbb{N}}$ of numbers that are values of independent random variables with the same distribution as X. (specifically we can think of X_n as the value of X for the nth. object we sampled). And now we want to infer what μ and σ^2 are using only the first however so many numbers of our sequence.

The nth sample mean is defined as $M_n := n^{-1}\sum_{j=1}^n X_j$. Note that $E(M_n) = \mu$ and $\sigma^2(M_n) = n^{-1}\sigma^2$ for all n and that $M_n \to \mu$ a.s. as $n \to \infty$.

The final statement is easily seen by applying Khinchine's strong law of large numbers. Meanwhile, the first result can be easily evaluated using the linearity of expectation (i.e. the fact that $E(n^{-1}\sum_{j=1}^n X_j) = n^{-1}\sum_{j=1}^n E(X_j)$). And finally, note by *corollary 10.6* plus the easily checked fact that $\sigma^2(aX) = a^2\sigma^2(X)$ that:

$$\sigma^2(M_n) = \sigma^2(n^{-1}\sum_{j=1}^n X_j) = n^{-2}\sigma^2(\sum_{j=1}^n X_j) = n^{-2}\sum_{j=1}^n \sigma^2 = n^{-1}\sigma^2$$

Next, the nth sample variance is defined as $S_n^2 := (n-1)^{-1}\sum_{j=1}^n (X_j - M_n)^2$.

- Why do we use $n-1$ instead of n?

Consider letting $E_n^2 := n^{-1}\sum_{j=1}^n (X_j - M_n)^2$ for all n. Then:

$$\begin{aligned}
E(E_n^2) &= n^{-1}\sum_{j=1}^n E((X_j - M_n)^2) \\
&= n^{-1}\sum_{j=1}^n E(((X_j-\mu) - (M_n - \mu))^2) \\
&= n^{-1}\sum_{j=1}^n \left[E((X_j-\mu)^2) - 2E((X_j-\mu)(M_n-\mu)) + E((M_n-\mu)^2)\right] \\
&= n^{-1}\left[(\sum_{j=1}^n \sigma^2) - 2(\sum_{j=1}^n E((M_n-\mu)(X_j-\mu)) + (\sum_{j=1}^n n^{-1}\sigma^2))\right] \\
&= n^{-1}\left[(n+1)\sigma^2 - 2\sum_{j=1}^n E((M_n-\mu)(X_j-\mu))\right]
\end{aligned}$$

Also note for all j that:

$$\begin{aligned}
E((M_n-\mu)(X_j-\mu)) &= E((n^{-1}\sum_{i=1}^n X_i - \mu)\cdot(X_j-\mu)) \\
&= n^{-1}\sum_{i=1}^n E((X_i-\mu)(X_j-\mu)) \\
&= n^{-1}\left(E((X_j-\mu)^2) + \sum_{i\neq j} E(X_i-\mu)E(X_j-\mu)\right) \\
&= n^{-1}\left(\sigma^2 + \sum_{i\neq j} 0\right)
\end{aligned}$$

Hence, we've shown that:

$$E(E_n^2) = n^{-1}((n+1)\sigma^2 - 2(n\cdot n^{-1}\sigma^2)) = n^{-1}(n-1)\sigma^2.$$

And this proves that E_n^2 will typically under-estimate the actual variance of the population.

As a side note, the physical intuition for this is that when sampling a population you aren't likely to measure the outliers of a population and it is those that tend to dominate what would be the theoretical expression for the population variance.

Fortunately though, notice that upon replacing n^{-1} with $(n-1)^{-1}$, then it does work out that $E(S_n^2) = (n-1)^{-1}(n-1)\sigma^2 = \sigma^2$.

- Also note that $S_n^2 \to \sigma^2$ a.s. as $n \to \infty$.

To start off, this statement would be way easier to prove if we were working with E_n^2 instead of S_n^2. So let's first prove that if $E_n^2 \to \sigma^2$ a.s. as $n \to \infty$, then so does $S_n^2 \to \sigma^2$ a.s. as $n \to \infty$.

It's clear that $\frac{1}{n} \leq \frac{1}{n-1}$. So it will always be the case that:

$$\liminf_{n\to\infty} S_n^2 \geq \lim_{n\to\infty} E_n^2.$$

On the other hand, note that for all $a > 1$, we have that $\frac{a}{n} \geq \frac{1}{n-1}$ so long as $n \geq \frac{a}{a-1}$.

Why? For $x > 1$ and $a > 1$:

$$\frac{a}{x} - \frac{1}{x-1} = \frac{ax-a-x}{x(x-1)} \geq 0 \iff ax - a - x \geq 0 \iff (a-1)x \geq a$$
$$\iff x \geq \frac{a}{a-1}.$$

This proves that $\limsup_{n\to\infty} S_n^2 \leq a \lim_{n\to\infty} E_n^2$ for all $a > 1$. And now taking $a \to 1$ we get the desired result that $\lim_{n\to\infty} S_n^2 = \lim_{n\to\infty} E_n^2$.

So now we just need to show that $E_n^2 \to \sigma^2$ almost surely. Fortunately, note that:

$$\begin{aligned}
E_n^2 &= n^{-1}\textstyle\sum_{j=1}^n (X_j - M_j)^2 \\
&= n^{-1}\textstyle\sum_{j=1}^n ((X_j - \mu) - (M_j - \mu))^2 \\
&= n^{-1}\textstyle\sum_{j=1}^n \left[(X_j - \mu)^2 - 2(X_j - \mu)(M_n - \mu) + (M_n - \mu)^2\right] \\
&= \left(n^{-1}\textstyle\sum_{j=1}^n (X_j - \mu)^2\right) - \left(2(M_n - \mu)\cdot n^{-1}\textstyle\sum_{j=1}^n (X_j - \mu)\right) + (M_n - \mu)^2
\end{aligned}$$

Now we already proved that $M_n \to \mu$ a.s. Therefore it's clear that, $(M_n - \mu)^2 \to 0$ and $2(M_n - \mu) \to 0$ a.s. as $n \to \infty$. Meanwhile, by a straightforward application of Khinchine's strong law of large numbers we know that almost surely:

$$n^{-1}\textstyle\sum_{j=1}^n (X_j - \mu) \to 0 \text{ and } n^{-1}\textstyle\sum_{j=1}^n (X_j - \mu)^2 \to \sigma^2 \text{ as } n \to \infty$$

So, we almost surely have that $E_n^2 \to \sigma^2 - (0 \cdot 0) + 0 = \sigma^2$ as $n \to \infty$. ■

9/19/2025

Given any $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, we define the measure $\nu_\mu^{\sigma^2} := \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/(2\sigma^2)} \mathrm{d}t$.

You can check my math 180a notes to see that $\nu_\mu^{\sigma^2}(\mathbb{R}) = 1$. Also, while I realize that my math 180a notes don't prove that $\int |t| \mathrm{d}\nu_\mu^{\sigma^2}(t) < \infty$, come on it's not that hard to prove and I don't feel like entirely redoing all my notes from that class. So just note that $\int t \mathrm{d}\nu_\mu^{\sigma^2}(t) = \mu$ and $\int (t-\mu)^2 \mathrm{d}\nu_\mu^{\sigma^2}(t) = \sigma^2$.

$\nu_\mu^{\sigma^2}$ is called the normal / Gaussian distribution with mean μ and variance σ^2. Also ν_0^1 is called the standard normal distribution. Interestingly, this probability distribution was observed to be common in nature before the following technical reasoning was ever come up with for why.

Before continuing on, I need to lay some groundwork for some notation used in Folland's next proof. I don't know of any particular resources for this so I'm just going to wing this next part and also look at wikipedia for definitions (see bibliography for a link).

Given a function $g : \mathbb{R} \to \mathbb{R}$ and $t_0 \in \overline{\mathbb{R}}$, we say that $f(t) = o(g(t))$ as $t \to t_0$ if there exists $\alpha : \mathbb{R} \to \mathbb{R}$ (or $\mathbb{C}$) such that $f(t) = \alpha(t)g(t)$ and $\alpha(t) \to 0$ as $t \to t_0$. This is called little o notation.

Note that I'm letting α take on complex values because I want to let f also take on complex values. Although, for the sake of simplicity, I'm requiring g to be real-valued.

Note that if $g(t) \neq 0$ when $t \neq t_0$ on some neighborhood of t_0, then $f(t) = o(g(t))$ as $t \to t_0$ if and only if $\frac{f(t)}{g(t)} \to 0$ as $t \to t_0$.

Lemma 1: If $f(t) = o(g(t))$ as $t \to t_0$ and $c \neq 0$, then $f(t) = o(cg(t))$ as $t \to t_0$.

Proof:
Suppose $f(t) = \alpha(t)g(t)$ where $\alpha(t) \to 0$ as $t \to t_0$. Then just let $\alpha'(t) := c^{-1}\alpha(t)$ and we have that $f(t) = \alpha'(t) \cdot cg(t)$ where $\alpha'(t) \to 0$ as $t \to t_0$.

Lemma 2: If $f(t) = o(g(t))$ as $t \to t_0$ and $c \in \mathbb{C}$, then $cf(t) = o(g(t))$ as $t \to t_0$.

Proof:
Once again write $f(t) = \alpha(t)g(t)$. Then $cf(t) = c\alpha(t)g(t)$ and $c\alpha(t) \to 0$ as $t \to t_0$.

Lemma 3: $f(t) = o(g(t))$ as $t \to t_0$ iff $f(t) = o(|g(t)|)$ as $t \to t_0$.

$(\Longrightarrow)$
Suppose $f(t) = o(g(t))$ and write $f(t) = \alpha(t)g(t)$. Then set $\alpha'(t) := \alpha(t)g(t)|g(t)|^{-1}$ when $g(t) \neq 0$ and $\alpha'(t) := 0$ when $g(t) = 0$. Then it's clear that $f(t) = \alpha'(t)|g(t)|$ and that $|\alpha'(t)| \leq |\alpha(t)| \cdot 1 \to 0$ as $t \to t_0$. Hence $f(t) = o(|g(t)|)$.

$(\Longleftarrow)$
Suppose $f(t) = o(|g(t)|)$ and write $f(t) = \alpha(t)|g(t)|$. Then when $g(t) \neq 0$ set $\alpha'(t) := \alpha(t)|g(t)|(g(t))^{-1}$ and when $g(t) = 0$ set $\alpha'(t) := 0$. And, it's clear by identical reasoning as before that $f(t) = o(|g(t)|)$.

Lemma 4: If g_1 and g_2 are nonnegative and $f_1(t) = o(g_1(t))$ and $f_2(t) = o(g_2(t))$ as $t \to t_0$, then $(f_1 + f_2)(t) = o((g_1 + g_2)(t))$ as $t \to t_0$.

Proof:
Write $f_1(t) = \alpha_1(t)g_1(t)$ and $f_2(t) = \alpha_2(t)g_2(t)$, and then set:

$\alpha'(t) := \frac{\alpha_1(t)g_1(t)+\alpha_2(t)g_2(t)}{g_1(t)+g_2(t)}$ when $g_1(t) \neq 0$ and $g_2(t) \neq 0$, and $\alpha'(t) := 0$ otherwise.

Then $(f_1 + f_2)(t) = \alpha'(t)(g_1 + g_2)(t)$ and:

$$|\alpha'(t)| = |\alpha_1(t)|\frac{g_1(t)}{g_1(t)+g_2(t)} + |\alpha_2(t)|\frac{g_2(t)}{g_1(t)+g_2(t)} \leq |\alpha_1(t)|\frac{g_1(t)}{g_1(t)} + |\alpha_2(t)|\frac{g_2(t)}{g_2(t)}$$
$$= |\alpha_1(t)| + |\alpha_2(t)| \to 0 \text{ as } t \to t_0.$$

Corollary 5: If g_1 and g_2 always have the same sign and $f_1(t) = o(g_1(t))$ and $f_2(t) = o(g_2(t))$ as $t \to t_0$, then $(f_1 + f_2)(t) = o((g_1 + g_2)(t))$ as $t \to t_0$.

Proof:
Just apply lemmas 3 and 4 and use the fact the fact $|g_1| + |g_2| = |g_1 + g_2|$.

Corollary 6: If $f_1(t) = o(g(t))$ and $f_2(t) = o(g(t))$ as $t \to t_0$, then $(f_1 + f_2)(t) = o(g(t))$ as $t \to t_0$.

Proof:
Apply corollary 5 as well as lemma 1.

Lemma 7: Suppose $f_1(t) = o(g_1(t))$ and $f_2(t) = o(g_2(t))$ as $t \to t_0$. Then $(f_1 \cdot f_2)(t) = o((g_1 \cdot g_2)(t))$ as $t \to t_0$.

Proof:
Write $f_1(t) = \alpha_1(t)g_1(t)$ and $f_2(t) = \alpha_2(t)g_2(t)$. Then $\alpha'(t) := \alpha_1(t)\alpha_2(t)$ satisfies that $\alpha'(t) \to 0$ as $t \to t_0$ and $(f_1 \cdot f_2)(t) = \alpha'(t)(g_1 \cdot g_2)(t)$.

<u>Lemma 8:</u> Suppose $f(t) = o(g(t))$ as $t \to t_0$ and $h : \mathbb{R} \to \mathbb{R}$ is another function. Then $(f \cdot h)(t) = o((g \cdot h)(t))$ as $t \to t_0$.

Proof:
If $f(t) = \alpha(t)g(t)$, then $(f \circ h)(t) = \alpha(t)(g \circ h)(t)$.

<u>Lemma 9:</u> Suppose $f(t) = o(g(t))$ as $t \to t_0$ and $g(t) = o(h(t))$ as $t \to t_0$. Then $f(t) = o(h(t))$ as $t \to t_0$.

Proof:
Write $f(t) = \alpha_1(t)g(t)$ and $g(t) = \alpha_2(t)h(t)$. Then $f(t) = (\alpha_1(t)\alpha_2(t))h(t)$ and it's clear that $\alpha_1(t)\alpha_2(t) \to 0$ as $t \to t_0$.

<u>Lemma 10:</u> Suppose $f(t) = o(g(t) + \sum_{j=1}^{n} h_j(t))$ as $t \to t_0$ where each $h_j(t) = o(g(t))$ as $t \to t_0$ for each j. Then $f(t) = o(g(t))$ as $t \to t_0$.

Proof:
Write $h_j(t) = \alpha_j(t)g(t)$ for each j and $f(t) = \alpha(t) \cdot (g(t) + \sum_{j=1}^{n} h_j(t))$. Then $\alpha'(t) := \alpha(t)\left(1 + \sum_{j=1}^{n} \alpha_j(t)\right)$ satisfies that $f(t) = \alpha'(t)g(t)$ and:

$$\alpha'(t) \to 0(1 + \textstyle\sum_{j=1}^{n} 0) = 0 \text{ as } t \to t_0.$$

<u>Lemma 11:</u> Let $t_0, t_1 \in \overline{\mathbb{R}}$ and suppose $h : \mathbb{R} \to \mathbb{R}$ is some function such that $h(t) \to t_0$ as $t \to t_1$. Also suppose either:

- that $h(t) \neq t_0$ on $U - \{t_1\}$ for some open neighborhood of t_1,
- or that f and g are continuous at t_0 (so that α is too).

Then if $f(t) = o(g(t))$ as $t \to t_0$, we know that $(f \circ h)(t) = o((g \circ h)(t))$ as $t \to t_1$.

Proof:
Write $f(t) = \alpha(t)g(t)$. Then $(f \circ h)(t) = \alpha(h(t))g(h(t))$. And no matter which assumption proposed above that we make, we have that $\alpha(h(t)) \to 0$ as $t \to t_1$.

<u>Lemma 12:</u> Suppose $f(t) = o(g(t))$ as $t \to t_0$ and $|g| \leq |h|$ on some neighborhood of t_0. Then $f(t) = o(h(t))$.

Proof:
Write $f(t) = \alpha(t)g(t)$ and define $\alpha'(t) := \alpha(t)\frac{g(t)}{h(t)}$ when $h(t) \neq 0$ and $\alpha'(t) := 0$ otherwise. Then since $|g| \leq |h|$ on a neighborhood of t_0, it's clear that $|\alpha'(t)| \leq |\alpha(t)|$ on a neighborhood of t_0. Hence, $f(t) = \alpha'(t)h(t)$ and $\alpha'(t) \to 0$ as $t \to t_0$.

Now, why do we care about little o notation? Essentially, there are two use cases. First, if $g(t) \to \pm\infty$ as $t \to t_0$ then we can use little o notation to describe a cap on how fast something grows. After all, if g had that property and $f(t) = o(g(t))$ as $t \to t_0$, then that would be equivalent to saying that for every $\varepsilon > 0$ there exists some $\delta > 0$ such that $|f(t)| < \varepsilon|g(t)|$ for all t satisfying that $|t - t_0| < \delta$.

Meanwhile, note that if $g(t)$ is bounded on a neighborhood of t_0 and $f(t) = o(g(t))$ as $t \to t_0$, then we must have that $f(t) \to 0$ as $t \to t_0$. This points to the other use case of little o notation which is to rigorously keep track of how negligible an error term is as we approach an asymptotic case while not having to worry about what those error terms specifically are. In fact, typically for this use case we will just entirely forget what f is and write expressions like $1 + x^2 + o(x^2)$ as $x \to 0$. And if we write $o(g(t))$ (as $t \to t_0$) in an arithmetic setting without any other context, then it is implied that you can replace $o(g(t))$ with any f satisfying that $f(t) = o(g(t))$ as $t \to t_0$

One more note to make is that (according to wikipedia) usually it is assumed that $t_0 = \infty$ unless otherwise stated. Also, hopefully it is clear how little o notation can easily be applied to sequences in the case that $t_0 = \infty$.

Now Folland near the end of his next proof does some cool reasoning which unfortunately is flawed. Therefore, while I still want to show off the idea (since I think it's a cool usage of little o notation), I'm going to try and find a different proof.

Lemma: For any fixed $b \in \mathbb{R}$, we have that $\lim\limits_{n\to\infty} \left[1 + \frac{b}{n} + o(\frac{1}{n})\right]^n = e^b$.

Proof:
Let $(a_n)_{n\in\mathbb{N}}$ be any sequence in $\mathbb{C}$ with $a_n \to 0$ as $n \to \infty$ (so that $o(\frac{1}{n}) = a_n\frac{1}{n}$).
Then note that:

$$\left[1 + \tfrac{b}{n} + \tfrac{a_n}{n}\right]^n = (1 + \tfrac{b}{n})^n + \sum_{k=1}^{n}(1 + \tfrac{b}{n})^{n-k} \cdot \tfrac{n\cdots(n-k+1)}{k!} \cdot (\tfrac{a_n}{n})^k$$

It's clear that $(1 + \frac{b}{n})^n \to e^b$ as $n \to \infty$. So, we just need to show that the complicated-looking sum goes to 0 as $n \to \infty$. To do this, note that:

$$\left|\sum_{k=1}^{n}(1 + \tfrac{b}{n})^{n-k} \cdot \tfrac{n\cdots(n-k+1)}{k!} \cdot (\tfrac{a_n}{n})^k\right| \leq \sum_{k=1}^{n} \left|(1 + \tfrac{b}{n})^{n-k}\right| \cdot \tfrac{n\cdots(n-k+1)}{k!\cdot n^k} \cdot |a_n|^k$$

Now $\frac{n\cdots(n-k+1)}{k!\cdot n^k} \leq \frac{1}{k!}$. Also, when n is big enough, we can guarentee $(1 + \frac{b}{n}) \in (0, 2)$. Then $(1 + \frac{b}{n})^{n-k} = ((1 + \frac{b}{n})^n)^{1-\frac{k}{n}}$. And in the case that b is negative, we will have that $((1 + \frac{b}{n})^n)^{1-\frac{k}{n}} \leq ((1 + \frac{b}{n})^n)^0 = 1$ for all k. Meanwhile, in the case that b is nonnegative, we know that for any $C > e^b$ that $((1 + \frac{b}{n})^n)^{1-\frac{k}{n}} \leq ((1 + \frac{b}{n})^n)^1 < C$ for all k once n is big enough. Either way, this proves that there exists a constant C' such that once n is big enough, $|(1 + \frac{b}{n})^{n-k}| \leq C'$.

So, we now have for all n sufficiently large that:

$$\begin{aligned}\left|\sum_{k=1}^{n}(1 + \tfrac{b}{n})^{n-k} \cdot \tfrac{n\cdots(n-k+1)}{k!} \cdot (\tfrac{a_n}{n})^k\right| &\leq C' \sum_{k=1}^{n} \tfrac{1}{k!}|a_n|^k \\ &= C'(-1 + \sum_{k=0}^{n} \tfrac{1}{k!}|a_n|^k) \\ &\leq C'(-1 + \sum_{k=0}^{\infty} \tfrac{1}{k!}|a_n|^k) = C'(e^{|a_n|} - 1).\end{aligned}$$

And since $|a_n| \to 0$ as $n \to \infty$, we are done. ■

Theorem 10.14: Let λ be a Borel probability measure on $\mathbb{R}$ such that $\int t^2 \mathrm{d}\lambda(t) = 1$ and $\int t\mathrm{d}\lambda(t) = 0$ (side note: the finiteness of the first integral implies the existence of the second integral). For $n \in \mathbb{N}$ let $\lambda^{*n} := \lambda * \cdots * \lambda$ (i.e. λ convoluted with itself n times), and $\lambda_n(E) := \lambda^{*n}(\sqrt{n}E)$ for all $E \in \mathcal{B}_{\mathbb{R}}$ where $\sqrt{n}E := \{\sqrt{n}t : t \in E\}$. Then $\lambda_n \to \nu_0^1$ vaguely as $n \to \infty$.

As a side note: λ_n is just the measure induced by λ^{*n} and $t \mapsto n^{-1/2}t$. So λ_n is in fact a well-defined measure. This also means that $\int f(t)\mathrm{d}\lambda_n(t) = \int f(n^{-1/2}t)\mathrm{d}\lambda^{*n}(t)$ for all f where either integral exists by *proposition 10.1*.

Proof:
By applying part (b) of the *theorem on page 189* of my journal twice and then part (a) once, we can show that $\widehat{\lambda} \in C^2(\mathbb{R})$ with:

- $\widehat{\lambda}(0) = \int e^{-2\pi i(0)\cdot t}\mathrm{d}\lambda(t) = \lambda(\mathbb{R}) = 1,$
- $\widehat{\lambda}'(0) = \int -2\pi i t e^{-2\pi i(0)\cdot t}\mathrm{d}\lambda(t) = -2\pi i \int t \cdot 1\mathrm{d}\lambda(t) = 0,$
- $\widehat{\lambda}''(0) = \int (-2\pi i t)^2 e^{-2\pi i(0)\cdot t}\mathrm{d}\lambda(t) = -4\pi^2 \int t^2 \cdot 1\mathrm{d}\lambda(t) = -4\pi^2.$

It follows by applying Taylor's theorem to the real and imaginary parts separately (see my math 240c notes) that $\lambda(\xi) = 1 - 2\pi^2\xi^2 + o(\xi^2)$ as $\xi \to 0$.

Next note that: (where $o(\frac{\xi^2}{n})$ is by any path where $\frac{\xi^2}{n} \to 0$)

$$\widehat{\lambda}_n(\xi) = \int e^{-2\pi i \xi t}\mathrm{d}\lambda_n(t) = \int e^{-2\pi i \xi n^{-1/2}t}\mathrm{d}\lambda^{*n}(t)$$
$$= \widehat{\lambda^{*n}}(n^{-1/2}\xi) = \left[\widehat{\lambda}(n^{-1/2}\xi)\right]^n = \left[1 - \tfrac{2\pi^2\xi^2}{n} + o(\tfrac{\xi^2}{n})\right]^n$$

In particular, upon fixing $\xi \in \mathbb{R}$ we can apply the prior lemma to get that:

$$\left[1 - \tfrac{2\pi^2\xi^2}{n} + o(\tfrac{\xi^2}{n})\right]^n = \left[1 + \tfrac{-2\pi^2\xi^2}{n} + o(\tfrac{1}{n})\right]^n \to e^{-2\pi^2\xi^2} \text{ as } n \to \infty.$$

This shows that $\widehat{\lambda}_n \to e^{-2\pi^2\xi^2}$ pointwise as $n \to \infty$.

But now recall from Folland proposition 8.24 (which is in my math 240c notes) that if $f(x) = e^{-\pi a x^2}$ where $a > 0$, then $\widehat{f}(\xi) = a^{-1/2}e^{-\pi\xi^2/a}$. Therefore:

$$\widehat{\nu}_0^1(\xi) = (\tfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2})^\wedge(\xi) = \tfrac{1}{\sqrt{2\pi}}(e^{-\pi(\frac{1}{2\pi})x^2})^\wedge(\xi) = \tfrac{1}{\sqrt{2\pi}}(\tfrac{1}{2\pi})^{-1/2}e^{-\pi\xi^2\cdot(\frac{1}{2\pi})^{-1}} = 1 \cdot e^{-2\pi^2\xi^2}$$

Hence $\widehat{\lambda}_n \to \widehat{\nu}_0^1$ pointwise as $n \to \infty$. And since $\|\lambda_k\| = 1$ for all k, we can thus conclude by *proposition 8.50* that $\lambda_n \to \nu_0^1$ vaguely. ■

Before moving on, I want to show off what Folland tried to do. Folland notes that $\log(1+z) = z + o(z)$ as $z \to 0$. And therefore we have that:

$$\log(\widehat{\lambda}_n(\xi)) = n\log(1 - \tfrac{2\pi^2\xi^2}{n} + o(\tfrac{\xi^2}{n})) = n(-\tfrac{2\pi^2\xi^2}{n} + o(\tfrac{\xi^2}{n}) + o(-\tfrac{2\pi^2\xi^2}{n} + o(\tfrac{\xi^2}{n})))$$
$$= -2\pi^2\xi^2 + n \cdot \left(o(\tfrac{\xi^2}{n}) + o(-\tfrac{2\pi^2\xi^2}{n} + o(\tfrac{\xi^2}{n}))\right)$$
$$= -2\pi^2\xi^2 + n \cdot o(\tfrac{\xi^2}{n}) \to -2\pi^2\xi^2 \text{ as } n \to \infty.$$

My issue with this reasoning is that $\widehat{\lambda}_n(\xi)$ is a complex valued function. So, what exactly does it mean to take the logarithm of it? Also, even if we were to extend the logarithm function to the complex plane, how do we know that $\log(z^n) = n\log(z)$ when z has an imaginary component?

The Central Limit Theorem: Let $\{X_j\}_{j\in\mathbb{N}}$ be a sequence of independent identically distributed L^2 random variables with mean μ and variance σ^2. As $n\to\infty$, the distribution of $(\sigma\sqrt{n})^{-1}\sum_{j=1}^n(X_j-\mu)$ converges vaguely to the standard normal distribution ν_0^1 as $n\to\infty$, and for all $a\in\mathbb{R}$:

$$\lim_{n\to\infty} P\left(\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^n (X_j-\mu)\le a\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^a e^{\frac{-t^2}{2}}\mathrm{d}t$$

Proof:
Let λ be the common distribution of $\sigma^{-1}(X_j-\mu)$ for all j. Then λ satisfies the hypotheses last theorem. So, using the notation of the last theorem, we know that $\lambda_n\to\nu_0^1$ vaguely as $n\to\infty$. But now by applying *proposition 10.4* as well as the quick lemma on the same page, we have that:

$$\lambda_n = (\lambda^{*n})_{t\mapsto n^{-1/2}t} = (P_{\sigma^{-1}\sum_{j=1}^n(X_j-\mu)})_{t\mapsto n^{-1/2}t} = P_{\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^n(X_j-\mu)}$$

This proves the first claim. The second claim follows easily from *proposition 7.19 part (b)* after you notice that $F(a)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^a e^{\frac{-t^2}{2}}\mathrm{d}t$ is continuous at all a and that we trivially have that all of the distributions of the $(\sigma\sqrt{n})^{-1}\sum_{j=1}^n(X_j-\mu)$ are positive measures.
■

9/20/2025

To start off, there's an exercise that applies both the central limit theorem and the law of large numbers that I want to do. However, before I tackle it I need to do an exercise from chapter 2 of Folland.

Exercise 2.39: Let $(X,\mathcal{M},\mu)$ be a measure space and suppose $(f_n)_{n\in\mathbb{N}}$ is a sequence of functions on X such that $f_n\to f$ almost uniformly as $n\to\infty$ (meaning for all $\varepsilon>0$ there exists $E\in\mathcal{M}$ such that $\mu(E)<0$ and $f_n\to f$ uniformly on E^C). Then $f_n\to f$ a.e. and in measure.

Proof:
The first claim is simple. For each n let E_n be a set satisfying that $\mu(E_n)<{}^1/n$ and $f_n\to f$ uniformly as $n\to\infty$ on E_n^C. Then it is clear that $f_n\to f$ pointwise on $\bigcup_{n\in\mathbb{N}}E_n^\mathsf{C}$. And $\mu((\bigcup_{n\in\mathbb{N}}E_n^\mathsf{C})^\mathsf{C}) = \mu(\bigcap_{n\in\mathbb{N}}E_n) < \mu(E_n) = {}^1/n$ for all $n\in\mathbb{N}$. Thus it follows that $\mu((\bigcup_{n\in\mathbb{N}}E_n^\mathsf{C})^\mathsf{C}) = 0$ and $f_n\to f$ a.e.

To show the second claim, consider any fixed $\varepsilon>0$ and $\delta>0$. Then let E be a set satisfying that $\mu(E)<\delta$ and $f_n\to f$ uniformly on E^C. Then clearly there is some N such that for all $n\ge N$, $\mu(\{x : |f_n(x)-f(x)|\ge\varepsilon\})\le\mu(E)<\delta$. This proves that $f_n\to f$ in measure.

By combining the prior exercise with Egoroff's theorem, we can see that on any finite measure space, if $f_n \to f$ pointwise a.e., then $f_n \to f$ almost uniformly and therefore also in measure. This can be especially useful to keep in mind when working on spaces equipped with probability measures.

(Although now that I'm actually doing exercise 10.20, I don't think I actually need to use the prior exercise in order to do exercise 10.20.)

Exercise 2.37: Let $(X, \mathcal{M}, \mu)$ be a measure space. Then suppose $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of measurable functions on X, f is a measurable function on X, and consider any function $\phi : \mathbb{C} \to \mathbb{C}$.

(a) If ϕ is continuous and $f_n \to f$ a.e., then $\phi \circ f_n \to \phi \circ f_n$ a.e.

Suppose x satisfies that $f_n(x) \to f(x)$ as $n \to \infty$. Then since ϕ is continuous, we have that $\phi(f_n(x)) \to \phi(f(x))$ as $n \to \infty$. This shows that the set where $f_n \to f$ is a subset of the set where $\phi \circ f_n \to \phi \circ f$. Hence $\phi \circ f_n \to \phi \circ f_n$ a.e.

(b) If ϕ is uniformly continuous and $f_n \to f$ uniformly, almost uniformly, or in measure, then $\phi \circ f_n \to \phi \circ f$ uniformly, almost uniformly, or in measure respectively.

First suppose $f_n \to f$ uniformly. Then for any $\varepsilon > 0$ pick $\delta > 0$ such that $|\phi(z_1) - \phi(z_2)| < \varepsilon$ when $|z_1 - z_2| < \delta$ and $N \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \delta$ for all $n \geq N$ and $x \in X$. And now $|\phi(f_n(x)) - \phi(f(x))| < \varepsilon$ for all $n \geq N$ and $x \in X$.

This also shows that if $f_n \to f$ uniformly on some set $E^{\mathsf{C}} \subseteq X$, then $\phi \circ f_n \to \phi \circ f$ uniformly on E^{C}. And from there it's clear that $f_n \to f$ almost uniformly implies that $\phi \circ f_n \to \phi \circ f$ almost uniformly.

Finally, suppose $f_n \to f$ in measure. Then for any $\varepsilon > 0$ and $\delta > 0$ we want to show there exists $N \in \mathbb{N}$ such that $\mu(\{x : |(\phi \circ f_n)(x) - (\phi \circ f_n)(x)| \geq \delta\}) < \varepsilon$ for all $n \geq N$. To do this, pick $\kappa > 0$ such that $|\phi(z_1) - \phi(z_2)| < \delta$ whenever $|z_1 - z_2| < \kappa$. Then upon picking $N \in \mathbb{N}$ such that $\mu(\{x : |f_n(x) - f(x)| \geq \kappa\}) < \varepsilon$, we have for all $n \geq N$ that:

$$\mu(\{x : |(\phi \circ f_n)(x) - (\phi \circ f_n)(x)| \geq \delta\}) \subseteq \mu(\{x : |f_n(x) - f(x)| \geq \kappa\}) < \varepsilon. \blacksquare$$

Now to the actual exercise I wanted to do...

Exercise 10.20: If $\{X_j\}_{j\in\mathbb{N}}$ is a sequence of independent identically distributed random variables with mean 0 and variance 1, then the distributions of $\sum_{j=1}^n X_j / (\sum_{j=1}^n X_j^2)^{1/2}$ and $\sqrt{n} \sum_{j=1}^n X_j / (\sum_{j=1}^n X_j^2)$ both converge vaguely to the standard normal distribution.

Proof:

Firstly, note that:

$$\sum_{j=1}^n X_j / (\sum_{j=1}^n X_j^2)^{1/2} = n^{-1/2} \sum_{j=1}^n X_j / (\tfrac{1}{n} \sum_{j=1}^n X_j^2)^{1/2}.$$

Similarly, note that:

$$\sqrt{n} \sum_{j=1}^n X_j / (\sum_{j=1}^n X_j^2) = \sum_{j=1}^n X_j / (\tfrac{1}{\sqrt{n}} \sum_{j=1}^n X_j^2) = n^{-1/2} \sum_{j=1}^n X_j / (\tfrac{1}{n} \sum_{j=1}^n X_j^2)$$

Now by Khinchine's strong law of large numbers, we know that $\frac{1}{n} \sum_{j=1}^n X_j^2 \to \sigma^2 = 1$ almost surely as $n \to \infty$. Also, by a straightforward application of the prior exercise using $\phi(z) = \sqrt{\mathrm{Re}(z)}\chi_{(0,\infty)}(\mathrm{Re}(z))$, we know that $(\frac{1}{n} \sum_{j=1}^n X_j^2)^{1/2} \to \sqrt{1} = 1$ almost surely as $n \to \infty$.

Meanwhile, the central limit theorem says for all $a \in \mathbb{R}$ that:
$$\lim_{n\to\infty} P(n^{-1/2}\sum_{j=1}^{n} X_j < a) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-t^2/2}\mathrm{d}t.$$

So for the sake of brevity, from now on I will just write out that I'm working with a sequence of random variables $\{{}^{Y_n}/_{Z_n}\}_{n\in\mathbb{N}}$ where $Z_n \to 1$ a.s. as $n \to \infty$ and $\lim_{n\to\infty} P(Y_n < a) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-t^2/2}\mathrm{d}t$. And if we can show that the distribution of ${}^{Y_n}/_{Z_n}$ converges vaguely to the standard normal distribution, we are done.

From here our strategy is to apply *proposition 7.19*. Fortunately, the first condition of that proposition is trivial since we are working with probability measures. As for showing the second condition, let $F_n(a) := P(Y_n/Z_n \leq a)$ and $F(a) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-t^2/2}\mathrm{d}t$ for all $a \in \mathbb{R}$. Now, we need to show that $F_n \to F$ pointwise (since F is continuous everywhere).

To do this, note that $Z_n \to 1$ almost uniformly by Egoroff's theorem. And from there we know that for any $\varepsilon > 0$ and $\delta \in (0,1)$ that there exists an event E with $P(E) < \varepsilon$ as well as an integer $N \in \mathbb{N}$ such that $|Z_n(\omega) - 1| \leq \delta$ for all integers $n \geq N$ and outcomes $\omega \in E^{\mathsf{C}}$. And in turn we can say for all $n \geq N$ and $a \geq 0$ that:

- $$\begin{aligned} F_n(a) \geq P(\{\tfrac{Y_n}{Z_n} \leq a\} \cap E^{\mathsf{C}}) &\geq P(\{Y_n \leq a(1-\delta)\} \cap E^{\mathsf{C}}) \\ &\geq P(Y_n \leq a(1-\delta)) - P(E) \\ &> P(Y_n \leq a(1-\delta)) - \varepsilon \end{aligned}$$

- $$\begin{aligned} F_n(a) \leq P(\{\tfrac{Y_n}{Z_n} \leq a\} \cap E^{\mathsf{C}}) + P(E) &\leq P(\{Y_n \leq a(1+\delta)\} \cap E^{\mathsf{C}}) + P(E) \\ &\leq P(Y_n \leq a(1+\delta)) + P(E) \\ &< P(Y_n \leq a(1+\delta)) + \varepsilon \end{aligned}$$

Meanwhile if $a < 0$ and $n \geq N$ then we can say that:
$$P(Y_n \leq a(1+\delta)) - \varepsilon < F_n(a) < P(Y_n \leq a(1-\delta)) + \varepsilon.$$

However, this minute detail doesn't really affect any of the core reasoning after this. So I'm just going to assume I'm working with $a \geq 0$ from now on.

Taking $n \to \infty$, we have that:
$$F(a(1-\delta)) - \varepsilon \leq \liminf_{n\to\infty} F_n(a) \leq \limsup_{n\to\infty} F_n(a) \leq F(a(1+\delta)) + \varepsilon.$$

And since F is continuous, we can take $\varepsilon \to 0$ and $\delta \to 0$ to get that $F(a) = \lim_{n\to\infty} F_n(a)$.
■

9/21/2025

Now is the day that I return to asking about how we construct sequences of independent random variables. (see *page 199* for when I brought up this question). Unfortunately, answering this question will require taking a detour back to Folland chapter 7 to finish learning about Radon products of measures. To see where I last stopped in this chapter, go to *page 184*.

Let X and Y be LCH spaces equipped with Borel Radon measures. Then like in my paper notes for math 240a, given any set $E \subseteq X \times Y$ denote $E_x := \{y \in Y : (x,y) \in E\}$ and $E^y := \{x \in X : (x,y) \in E\}$. Also, given any function $f : X \times Y \to \mathbb{C}$, denote $f_x(\cdot) := f(x,\cdot)$ and $f^y(\cdot) := f(\cdot,y)$.

Lemma 7.23:

(a) If $E \in \mathcal{B}_{X\times Y}$, then $E_x \in \mathcal{B}_Y$ for all $x \in X$ and $E^y \in \mathcal{B}_X$ for all $y \in Y$.

Proof:
Let $\mathcal{M} := \{E \in \mathcal{B}_{X\times Y} : E_x \in \mathcal{B}_Y \text{ and } E^y \in \mathcal{B}_X \text{ for all } x \in X \text{ and } y \in Y\}$. Then we claim that $\mathcal{M}$ is a σ-algebra.

Suppose $E \in \mathcal{M}$. Then for all $x \in X$

$$\begin{aligned}(E^{\mathsf{C}})_x &= \{y \in Y : (x,y) \in E^{\mathsf{C}}\} \\ &= \{y \in Y : (x,y) \in E\}^{\mathsf{C}} = (E_x)^{\mathsf{C}} \in \mathcal{B}_Y.\end{aligned}$$

Similarly, $(E^{\mathsf{C}})^y = (E^y)^{\mathsf{C}} \in \mathcal{B}_X$ for all $y \in Y$. So $\mathcal{M}$ is closed under complements.

Next suppose $\{E_n\}_{n\in\mathbb{N}}$ is a collection in $\mathcal{M}$ and let $E = \bigcup_{n\in\mathbb{N}} E_n$. Then for all $x \in X$:

$$\begin{aligned}(E)_x &= \{y \in Y : (x,y) \in E\} \\ &= \{y \in Y : (x,y) \in \textstyle\bigcup_{n\in\mathbb{N}} E_n\} = \textstyle\bigcup_{n\in\mathbb{N}}\{y \in Y : (x,y) \in E_n\} \\ &= \textstyle\bigcup_{n\in\mathbb{N}}(E_n)_x \in \mathcal{B}_Y\end{aligned}$$

And similarly $E^y = \bigcup_{n\in\mathbb{N}}(E_n)^y \in \mathcal{B}_X$ for all $y \in Y$. So, $\mathcal{M}$ is closed under countable unions.

Also, for all $x \in X$ and $y \in Y$ the maps $x' \in X \mapsto (x',y)$ and $y' \in Y \mapsto (x,y')$ are easily checked to be continuous. Hence, we know that if $E \in \mathcal{B}_{X\times Y}$ is open, then so is E_x and E^y for all $x \in X$ and $y \in Y$. And consequently we know that $\mathcal{M}$ contains all the open sets in $X \times Y$. This proves that $\mathcal{M} = \mathcal{B}_{X\times Y}$.

(b) If $f : X \times Y \to \mathbb{C}$ is $\mathcal{B}_{X\times Y}$-measurable, then f_x is $\mathcal{B}_Y$-measurable for all $x \in X$ and f^y is $\mathcal{B}_X$-measurable for all $y \in Y$.

Proof:
Note that for any $A \in \mathcal{B}_{\mathbb{C}}$ and $x \in X$, we have that:

$$\begin{aligned}(f_x)^{-1}(A) &= \{y \in Y : f(x,y) \in A\} \\ &= \{y \in Y : (x,y) \in f^{-1}(A)\} = (f^{-1}(A))_x\end{aligned}$$

Similarly, we have for any $A \in \mathcal{B}_{\mathbb{C}}$ and $y \in Y$ that $(f^y)^{-1}(A) = (f^{-1}(A))^y$. Thus if f is $\mathcal{B}_{X\times Y}$-measurable, then it is a straightforward application of part (a) to prove part (b). ■

Lemma 7.24: If $f \in C_c(X \times Y)$ and μ and ν are Radon measures on X and Y respectively, then the functions $x \mapsto \int f_x \mathrm{d}\nu$ and $y \mapsto \int f^y \mathrm{d}\mu$ are continuous.

Proof:

Folland writes out only the proof for $x \mapsto \int f_x \mathrm{d}\nu$ so I'm only going to write out the proof for $y \mapsto \int f^y \mathrm{d}\mu$ (they're basically identical so it doesn't matter). It suffices to show that given any $y_0 \in Y$ and $\varepsilon > 0$ there is a neighborhood $V \subseteq Y$ of y_0 such that $\|f^y - f^{y_0}\|_u < \varepsilon$ for all $y \in V$ since then $|\int f^y \mathrm{d}\mu - \int f^{y_0} \mathrm{d}\mu| = |\int f^y - f^{y_0} \mathrm{d}\mu| \leq \varepsilon \cdot \mu(\pi_X(\mathrm{supp}(f)))$ for all $y \in V$ (thus showing $y \mapsto \int f^y \mathrm{d}\mu$ is continuous at any $y_0 \in Y$).

However, since f is continuous, we know that for each $x' \in \pi_X(\mathrm{supp}(f))$ there exists neighborhoods $U_{x'} \subseteq X$ and $V_{x'} \subseteq Y$ of x' and y_0 respectively such that if $(x, y) \in U_x \times V_x$, then $|f(x,y) - f(x', y_0)| < \frac{1}{2}\varepsilon$. And then since $\pi_X(\mathrm{supp}(f))$ is compact, we can choose a finite subcover $U_{x'_1}, \ldots, U_{x'_n}$ of $\pi_X(\mathrm{supp}(f))$. And by setting $V = \bigcap_{j=1}^n V_{x'_j}$, we claim that $V \subseteq Y_0$ is a neighborhood of y_0 such that for any $x \in X$ and $y \in V$, $|f(x,y) - f(x,y_0)| < \varepsilon$.

If $x \notin \pi_X(\mathrm{supp}(f))$, then clearly $|f(x,y) - f(x,y_0)| = |0 - 0| < \varepsilon$ no matter what y is. On the other hand, if $x \in \pi_X(\mathrm{supp}(f))$, then we know that $(x, y) \in U_{x'_j} \times V_{x'_j}$ for some $x'_j \in X$. And since (x, y_0) is then also in $U_{x'_j} \times V_{x'_j}$, we have that:

$|f(x,y) - f(x,y_0)| \leq |f(x,y) - f(x'_j, y_0)| + |f(x, y_0) - f(x'_j, y_0)| < \varepsilon/2 + \varepsilon/2$

This proves that $\|f^y - f^{y_0}\|_u < \varepsilon$ for all $y \in V$. ■

Proposition 7.25: Let μ and ν be Radon measures on X and Y. If U is open in $X \times Y$, then the functions $x \mapsto \nu(U_x)$ and $y \mapsto \mu(U^y)$ are Borel measurable on X and Y, and $\mu \widehat{\times} \nu(U) = \int \nu(U_x) \mathrm{d}\mu_x = \int \mu(U^y) \mathrm{d}\nu(y)$.

Proof:

Let $\mathfrak{F} \coloneqq \{f \in C_c(X \times Y) : 0 \leq f \leq \chi_U\}$. Now by Folland proposition 7.11 (a) and (e) (near the end of my math 240c notes), we know that χ_U is LSC (lower semicontinuous), and that $\chi_U = \sup\{f : f \in \mathfrak{F}\}$. It easily follows then that $\chi_{U_x} = \sup\{f_x : f \in \mathfrak{F}\}$ and that $\chi_{U^y} = \sup\{f^y : f \in \mathfrak{F}\}$.

Next, it's obvious that $\mathfrak{F}$, $\{f_x : f \in \mathfrak{F}\}$, and $\{F^y : f \in \mathfrak{F}\}$ are all families of LCS functions (since continuous function are automatically semicontinuous). And you can see that those families are directed by $\leq$ by considering that if $f_1, f_2 \in \mathfrak{F}$, then $\max(f_1, f_2) \in \mathfrak{F}$. Thus by proposition 7.12 (also near the end of my math 240c notes):

- $\mu \widehat{\times} \nu(U) = \int \chi_U \mathrm{d}\mu \widehat{\times} \nu = \sup\{\int f \mathrm{d}\mu \widehat{\times} \nu : f \in \mathfrak{F}\}$
- $\nu(U_x) = \int \chi_{U_x} \mathrm{d}\nu = \sup\{\int f_x \mathrm{d}\nu : f \in \mathfrak{F}\}$
- $\mu(U^y) = \int \chi_{U^y} \mathrm{d}\mu = \sup\{\int f^y \mathrm{d}\mu : f \in \mathfrak{F}\}$

Now by lemma 7.24 above, we know that $x \mapsto \int f_x \mathrm{d}\nu$ and $y \mapsto \int f^y \mathrm{d}\mu$ are continuous (and thus LSC) mappings for all $f \in \mathfrak{F}$. And so by applying proposition 7.11 (c), we can conclude that $x \mapsto \nu(U_x)$ and $y \mapsto \mu(U^y)$ are both LSC (and hence Borel measurable). Finally, *proposition 7.22* following by one more application of proposition 7.12 yields:

$$\mu \widehat{\times} \nu(U) = \sup\{\iint f_x \mathrm{d}\nu \mathrm{d}\mu : f \in \mathfrak{F}\} = \int \sup\{\int f_x \mathrm{d}\nu : f \in \mathfrak{F}\} \mathrm{d}\mu = \int \nu(U_x) \mathrm{d}\mu$$

Similarly, $\mu \widehat{\times} \nu(U) = \int \mu(U^y) \mathrm{d}\nu$. ■

Theorem 7.26: Suppose that μ and ν are σ-finite Radon measures on X and Y. If $E \in \mathcal{B}_{X\times Y}$, then the functions $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ (which are well-defined by lemma 7.23) are Borel measurable on X and Y respectively. Also:

$$\mu \mathbin{\widehat{\times}} \nu(E) = \int \nu(E_x)\mathrm{d}\mu(x) = \int \mu(E^y)\mathrm{d}\nu(y).$$

And finally the restriction of $\mu \mathbin{\widehat{\times}} \nu$ to $\mathcal{B}_X \otimes \mathcal{B}_Y$ is just $\mu \times \nu$.

Proof:

To start off, fix any open $U \subseteq X$ and $V \subseteq Y$ satisfying that $\mu(U) < \infty$ and $\nu(V) < \infty$. Then let $W := U \times V$ and let $\mathcal{M}$ be the collection of all $E \in \mathcal{B}_{X\times Y}$ for which $E \cap W$ satisfies the conclusions of the theorem. Then note:

1. $\mathcal{M}$ contains all open sets in $X \times Y$ by the last proposition.
2. If $E, F \in \mathcal{M}$ and $F \subseteq E$, then $E - F \in \mathcal{M}$.

   Why? Note that:
   - $\mu \mathbin{\widehat{\times}} \nu(E \cap W) = \mu \mathbin{\widehat{\times}} \nu(F \cap W) + \mu \mathbin{\widehat{\times}} \nu((E - F) \cap W)$.
   - $\nu((E \cap W)_x) = \nu((F \cap W)_x) + \nu(((E - F) \cap W)_x)$
   - $\nu((E \cap W)^y) = \nu((F \cap W)^y) + \nu(((E - F) \cap W)^y)$.

   But now since W is open, we know $\mu \mathbin{\widehat{\times}} \nu(W) = \mu(U)\nu(V) < \infty$ by the last proposition. Also $\nu(W_x) = \nu(V) < \infty$ and $\mu(W^y) = \mu(U) < \infty$. This proves the maps $x \mapsto \nu(((E - F) \cap W)_x)$ and $y \mapsto \mu(((E - F) \cap W)^y)$ are Borel measurable since they are both differences of two other Borel functions (this is why we wanted to define W). Also, we have that:

   $$\begin{aligned}\mu \mathbin{\widehat{\times}} \nu((E - F) \cap W) &= \mu \mathbin{\widehat{\times}} \nu(E \cap W) - \mu \times \widehat{\times}\, \nu(F \cap W)\\ &= \int \nu((E \cap W)_x)\mathrm{d}\mu(x) - \int \nu((F \cap W)_x)\mathrm{d}\mu(x)\\ &= \int \nu(((E - F) \cap W)_x)\mathrm{d}\mu(x)\end{aligned}$$

   And similar reasoning holds for the y-slices. This proves that $E - F \in \mathcal{M}$.

3. As a consequence of the last two bullet points, we know that if $E \in \mathcal{M}$, then so does $E^{\mathsf{C}} = X - E \in \mathcal{M}$.
4. $\mathcal{M}$ is closed under finite disjoint unions.

   Why? This is a straightforward application of the finite additivity of measures and integrals and also the fact that a finite sum of Borel measurable functions is also Borel measurable.

5. $\mathcal{M}$ is closed under countable increasing unions.

   Why? Suppose $(E_n)_{n\in\mathbb{N}}$ is an increasing sequence of sets in $\mathcal{M}$ and $E := \bigcup_{n\in\mathbb{N}} E_n$. Then by the monotonicity of measures we have that $\nu(E_x) = \lim_{n\to\infty} \nu((E_n)_x)$ and $\mu(E^y) = \lim_{n\to\infty} \mu((E_n)^y)$ for all $x \in X$ and $y \in Y$. Hence, the maps $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ are measurable (see corollary 2 on page 43 of my latex math 240a notes).

Also, by the monotone convergence theorem and the monotonicity of measures we know that:

$$\begin{aligned}\mu \mathbin{\widehat{\times}} \nu(E) &= \lim_{n\to\infty} \mu \mathbin{\widehat{\times}} \nu(E_n) \\ &= \lim_{n\to\infty} \int \nu((E_n)_x)\mathrm{d}\mu(x) = \int \nu(E_x)\mathrm{d}\mu(x)\end{aligned}$$

And similar reasoning shows that $\mu \mathbin{\widehat{\times}} \nu(E) = \int \mu(E^y)\mathrm{d}\nu(y)$. Hence $E \in \mathcal{M}$.

6. By combining points 3 and 5, we also know that $\mathcal{M}$ is closed under countable decreasing intersections. This finishes showing that $\mathcal{M}$ is a monotone class (see my math 240a paper notes).

Now let $\mathcal{E} := \{A - B : A, B \text{ are open in } X \times Y\}$, and let $\mathcal{A}$ be the collection of finite disjoint unions of sets in $\mathcal{E}$. Note that $\mathcal{E}$ is an elementary family (see my latex math 240a notes) since:

- $\emptyset - \emptyset = \emptyset \in \mathcal{E}$
- $\begin{aligned}(A_1 - B_1) \cap (A_2 - B_2) &= (A_1 \cap B_1^{\mathsf{C}}) \cap (A_2 \cap B_2^{\mathsf{C}}) \\ &= (A_1 \cap A_2) \cap (B_1^{\mathsf{C}} \cap B_2^{\mathsf{C}}) = (A_1 \cap A_2) - (B_1 \cup B_2)\end{aligned}$
- $(A - B)^{\mathsf{C}} = [(X \times Y) - A] \cup [(A \cap B) - \emptyset]$

Thus by the proposition at the top of page 15 of my math 240a latex notes, we know that $\mathcal{A}$ is an algebra. And in turn, by the monotone class lemma (see my paper math 240a notes) we know that the monotone class $\mathcal{C}(\mathcal{A})$ generated by $\mathcal{A}$ is the same as the σ-algebra generated by $\mathcal{A}$ (which is clearly $\mathcal{B}_{X\times Y}$).

But note that $\mathcal{M}$ contains $\mathcal{A}$ since for any open $A, B \subseteq X \times Y$ we have by bullet points 1 and 2 that $A, A \cap B \in \mathcal{M}$ and that $A - B = A - (A \cap B) \in \mathcal{M}$. This shows that $\mathcal{B}_{X\times Y} = \mathcal{C}(\mathcal{A}) \subseteq \mathcal{M}$. And hence, we've proven that $\mathcal{B}_{X\times Y} = \mathcal{M}$.

Next, since X and Y are σ-finite, and μ and ν are Radon, we may write $X = \bigcup_{n\in\mathbb{N}} U_n$ and $Y = \bigcup_{n\in\mathbb{N}} V_n$ where each U_n and V_n is open and has finite measure. Furthermore, we may assume $\{U_n\}_{n\in\mathbb{N}}$ and $\{V_n\}_{n\in\mathbb{N}}$ are increasing sequences of sets. If $E \in \mathcal{B}_{X\times Y}$, the preceeding argument shows that $E \cap (U_n \times V_n)$ satisfies the conclusions of our theorem for all n. And by basically an identical argument to how we showed that $\mathcal{M}$ was closed under countable increasing unions, we can thus show that E also satisfies the conclusions of our theorem.

Finally, if $E \in \mathcal{B}_X \otimes \mathcal{B}_Y$, then by Fubini's theorem we have that:

$$\mu \times \nu(E) = \int \nu(E_x)\mathrm{d}\mu(x) = \mu \mathbin{\widehat{\times}} \nu(E). \blacksquare$$

Theorem 7.27: (The Fubini-Tonelli Theorem for Radon Products)

Let μ and ν be σ-finite Radon measures on X and Y, and let f be a Borel measurable function on $X \times Y$. Then f_x and f^y are Borel measurable for every x and y. If $f \geq 0$, then $x \mapsto \int f_x \mathrm{d}\nu$ and $y \mapsto \int f^y \mathrm{d}\mu$ are Borel measurable on X and Y. Meanwhile, if $f \in L^1(\mu \mathbin{\widehat{\times}} \nu)$, then $f_x \in L^1(\nu)$ for a.e. $x \in X$, $f^y \in L^1(\mu)$ for a.e. $y \in Y$, and $x \mapsto \int f_x \mathrm{d}\nu$ and $y \mapsto \int f^y \mathrm{d}\mu$ are in $L^1(\mu)$ and $L^1(\nu)$ respectively. And in both cases, we have that $\int f \mathrm{d}(\mu \mathbin{\widehat{\times}} \nu) = \iint f \mathrm{d}\mu \mathrm{d}\nu = \iint f \mathrm{d}\nu \mathrm{d}\mu$.

Proof:
The measurability of f_x and f^y was established in *lemma 7.23(b)*. Now the rest of the proof follows identically to our proof of the original Fubini-Tonelli theorem (see my paper math 240a notes) except that we use theorem 7.26 instead of theorem 2.36 to establish that the Fubini-Tonelli theorem holds for characteristic functions. ■

Interestingly, we can extend the theory of radon products of finitely many factor to products of infinitely many factors provided that each factor measure space is compact and has total measure 1.

Specifically, let $\{X_\alpha\}_{\alpha \in A}$ be a family of compact Hausdorff spaces, and for each $\alpha \in A$ let μ_α be a Radon measure on X_α such that $\mu_\alpha(X_\alpha) = 1$. By Tychonoff's thoerem, we know that $X := \prod_{\alpha \in A} X_\alpha$ is also a compact Hausdorff space. And so, our goal is to define a Radon measure μ on X such that if E_α is a Borel on X_α for each $\alpha \in A$ and $E_\alpha = X_\alpha$ for all but finitely many $\alpha \in A$, then $\mu(\prod_{\alpha \in A} E_\alpha) = \prod_{\alpha \in A} \mu(E_\alpha)$.

As a side tangent: how do we define arbitrary products of numbers?

The intuition for this definition is that I'm just applying the definition of arbitrary sums of numbers to the collection of the logarithms of the numbers I actually care about. If $\{t_\alpha\}_{\alpha \in A} \subseteq [0, 1]$, then we define:

$$\prod_{\alpha \in A} t_\alpha := \inf\{\prod_{\alpha \in B} t_\alpha : B \subseteq A \text{ is finite}\}.$$

Meanwhile, if $\{t_\alpha\}_{\alpha \in A} \subseteq [1, \infty]$, then we define:

$$\prod_{\alpha \in A} t_\alpha := \sup\{\prod_{\alpha \in B} t_\alpha : B \subseteq A \text{ is finite}\}.$$

Theorem: If $\{t_\alpha\}_{\alpha \in A} \subseteq [0, 1]$, then $\prod_{\alpha \in A} t_\alpha > 0$ only if there exists a countable set $B \subseteq A$ with $t_\alpha = 1$ for all $\alpha \notin B$ and $t_\alpha \neq 0$ for all $\alpha \in B$. Also, in the case that the set B exists, then if $b_1, b_2, \dots$ is an enumeration of B in any order, we have that $\prod_{\alpha \in A} t_\alpha = \lim_{n \to \infty} \prod_{j=1}^{n} t_{b_j}$.

Proof:
Suppose there are uncountably many t_α with $t_\alpha < 1$. Then we know there must exist a number $c < 1$ with infinitely many $t_\alpha < c$. And in turn, we know that $\prod_{\alpha \in A} t_\alpha \leq c^n$ for all $n \in \mathbb{N}$. But $c^n \to 0$ as $n \to \infty$ since $c < 1$. Hence $\prod_{\alpha \in A} t_\alpha = 0$.

Hopefully the rest of the theorem is then obvious. I don't want to write it out since it's 1:33am and I'm tired. ■

By mirror reasoning we can also show that if $\{t_\alpha\}_{\alpha \in A} \subseteq [1, \infty]$, then $\prod_{\alpha \in A} t_\alpha < \infty$ only if there exists exists a countable set $B \subseteq A$ with $t_\alpha = 1$ for all $\alpha \notin B$ and $t_\alpha \neq \infty$ for all $\alpha \in B$. And in the case that such a subset B exists, then if $b_1, b_2, \dots$ is an enumeration of B in any order, we have that $\prod_{\alpha \in A} t_\alpha = \lim_{n \to \infty} \prod_{j=1}^{n} t_{b_j}$.

One more bit of notation we will use in the next proof is that for any $\alpha_1, \dots, \alpha_n \in A$, we'll let $\pi_{\alpha_1, \dots, \alpha_n}$ denote the natural projection of X onto $\prod_{j=1}^{n} X_{\alpha_j}$.

One more note before getting to the final theorem of chapter 7 is that we need to extend the idea of Radon products of measures to products of more than 2 spaces.

Lemma: Suppose X, Y, Z are LCH spaces and μ, ν, λ are σ-finite Radon measures on X, Y, Z. Then $(\mu \widehat{\times} \nu) \widehat{\times} \lambda = \mu \widehat{\times}(\mu \widehat{\times} \lambda)$.

Proof:
Note that $(\mu \widehat{\times} \nu) \widehat{\times} \lambda$ and $\mu \widehat{\times}(\nu \widehat{\times} \lambda)$ are the unique measures on $X \times Y \times Z$ such that for all $f \in C_c(X \times Y \times Z)$ we have that:

- $\int f \mathrm{d}((\mu \widehat{\times} \nu) \widehat{\times} \lambda) = \int f \mathrm{d}((\mu \widehat{\times} \nu) \times \lambda)$
- $\int f \mathrm{d}(\mu \widehat{\times}(\nu \widehat{\times} \lambda)) = \int f \mathrm{d}(\mu \times (\nu \widehat{\times} \lambda))$

Hence, we just need to show that $\int f \mathrm{d}((\mu \widehat{\times} \nu) \times \lambda) = \int f \mathrm{d}(\mu \times (\nu \widehat{\times} \lambda))$ for all $f \in C_c(X \times Y \times Z)$. Fortunately, we know by theorem 7.26 that:

- $\int |f| \mathrm{d}((\mu \widehat{\times} \nu) \times \lambda) \leq \|f\|_u \mu(\pi_X(\mathrm{supp}(f)))\mu(\pi_Y(\mathrm{supp}(f)))\lambda(\pi_Z(\mathrm{supp}(f)))$
- $\int |f| \mathrm{d}(\mu \times (\nu \widehat{\times} \lambda)) \leq \|f\|_u \mu(\pi_X(\mathrm{supp}(f)))\mu(\pi_Y(\mathrm{supp}(f)))\lambda(\pi_Z(\mathrm{supp}(f)))$.

This shows that for all $f \in C_c(X \times Y \times Z)$, we have that $f \in L^1((\mu \widehat{\times} \nu) \times \lambda)$ and that $f \in L^1(\mu \times (\nu \widehat{\times} \lambda))$. And thus by combining the classical Fubini theorem with the Fubini's theorem we just proved, we have that:

- $\int f \mathrm{d}((\mu \widehat{\times} \nu) \times \lambda) = \int(\int f \mathrm{d}(\mu \widehat{\times} \nu))\mathrm{d}\lambda = \int(\iint f \mathrm{d}\mu \mathrm{d}\nu)\mathrm{d}\lambda$
- $\int f \mathrm{d}(\mu \times (\nu \widehat{\times} \lambda)) = \int(\int f \mathrm{d}\mu)\mathrm{d}(\nu \widehat{\times} \lambda) = \iint(\int f \mathrm{d}\mu)\mathrm{d}\nu \mathrm{d}\lambda$

And clearly $\int(\iint f \mathrm{d}\mu \mathrm{d}\nu)\mathrm{d}\lambda = \iint(\int f \mathrm{d}\mu)\mathrm{d}\nu \mathrm{d}\lambda$ ■

Corollary: The Radon product operation on σ-finite Radon measures is associative. And hence we can unambiguously just write $\mu_1 \widehat{\times} \cdots \widehat{\times} \mu_n$ if all measures involved are σ-finite.

9/23/2025

Now we are ready to extend the notion of Radon products of measures to arbitrary collections of measure spaces.

Theorem 7.28: Suppose that, for each $\alpha \in A$, μ_α is a Radon measure on the compact Hausdorff space X_α such that $\mu_\alpha(X_\alpha) = 1$. Then there is a unique Radon measure μ on $X \coloneqq \prod_{\alpha \in A} X_\alpha$ such that for any $\alpha_1, \ldots, \alpha_n \in A$ and any Borel set $E \in \prod_{j=1}^n X_{\alpha_j}$ we have that $\mu(\pi^{-1}_{\alpha_1, \ldots, \alpha_n}(E)) = \mu_{\alpha_1} \widehat{\times} \cdots \widehat{\times} \mu_{\alpha_n}$. Also $\mu(X) = 1$.

Proof:
Let $C_F(X)$ be the collection of all $f \in C(X)$ that depend on finitely many coordinates (i.e. there exists $\alpha_1, \ldots, \alpha_n \in A$ and $g \in C(\prod_{j=1}^n X_{\alpha_j})$ with $f = g \circ \pi_{\alpha_1, \ldots, \alpha_n}$). Then if $f \in C_F(X)$, we define $I(f) \coloneqq \int g \mathrm{d}(\mu_{\alpha_1} \widehat{\times} \cdots \widehat{\times} \mu_{\alpha_n})$. Note that adding on extra coordinate to $\alpha_1, \ldots, \alpha_n$ doesn't change our definition of $I(f)$ since $\mu_\alpha(X_\alpha) = 1$ for all $\alpha \in A$. Hence I is a well-defined positive linear functional on $C_F(X)$. Also, clearly $|I(f)| \leq \|f\|_u$ for all $f \in C_F(X)$ with equality when f is a constant function.

Next, $C_F(X)$ is an algebra in $C(X)$ that separates points, contains constant functions, and is closed under complex conjugation. Hence $C_F(X)$ is uniformly dense in $C(X)$ by the Stone-Weierstrass theorem. Hence, the functional I extends uniquely to a positive linear function of norm 1 on $C(X)$, and the Riesz representation theorem then yields a unique Radon measure μ on X such that $I(f) = \int f \mathrm{d}\mu$ for all $f \in C_F(X)$. And since $\|I\|_{op} = 1$, $\mu(X) = 1$.

Finally define, $\mu_{\alpha_1,\ldots,\alpha_n}(E) := \mu \circ \pi^{-1}_{\alpha_1,\ldots,\alpha_n}(E)$ for all $E \in \mathcal{B}_{X_{\alpha_1} \times \ldots \times X_{\alpha_n}}$. Then by *proposition 10.1*, we know that $\int g \mathrm{d}\mu_{\alpha_1,\ldots,\alpha_n} = \int g \circ \pi_{\alpha_1,\ldots,\alpha_n} \mathrm{d}\mu$ whenever g is a bounded Borel measurable function on $X_{\alpha_1} \times \ldots \times X_{\alpha_n}$. In particular, by our definition of μ we know that for all $g \in C(X_{\alpha_1} \times \ldots \times X_{\alpha_n})$:

$$\int g \mathrm{d}\mu_{\alpha_1,\ldots,\alpha_n} = \int g \circ \pi_{\alpha_1,\ldots,\alpha_n} \mathrm{d}\mu = \int g \mathrm{d}(\mu_{\alpha_1} \widehat{\times} \cdots \widehat{\times}\, \mu_{\alpha_n})$$

Thus if we can show that $\mu_{\alpha_1,\ldots,\alpha_n}$ is Radon, then we will know by the uniqueness part of the Riesz representation theorem that $\mu_{\alpha_1,\ldots,\alpha_n} = \mu_{\alpha_1} \widehat{\times} \cdots \widehat{\times}\, \mu_{\alpha_n}$.

Let E be any Borel set in $\prod_{j=1}^n X_{\alpha_j}$ and write $\pi = \pi_{\alpha_1,\ldots,\alpha_n}$ for short. Since μ is regular, we know that for any $\varepsilon > 0$ there exists a compact $K \subseteq \pi^{-1}(E)$ with $\mu(K) > \mu(\pi^{-1}(E)) - \varepsilon$. Then $K' := \pi(K)$ is a compact subset of E and $\mu_{\alpha_1,\ldots,\alpha_n}(K') = \mu(\pi^{-1}(K')) \geq \mu(K)$ since $K \subseteq \pi^{-1}(K') = \pi^{-1}(\pi(K))$. And in turn $\mu_{\alpha_1,\ldots,\alpha_n}(K') > \mu(\pi^{-1}(E)) - \varepsilon = \mu_{\alpha_1,\ldots,\alpha_n}(E)$. This proves that $\mu_{\alpha_1,\ldots,\alpha_n}$ is inner regular.

Doing the same reasoning to E^{C} shows that $\mu_{\alpha_1,\ldots,\alpha_n}$ is outer regular as well. And since $\mu_{\alpha_1,\ldots,\alpha_n}$ is trivially finite on all compact sets on account of it being a finite measure, we know that $\mu_{\alpha_1,\ldots,\alpha_n}$ is Radon.

All that's left to do now is prove the uniqueness of μ. So suppose ν is another such measure on X such that for any $\alpha_1, \ldots, \alpha_n \in A$ and any Borel set $E \in \prod_{j=1}^n X_{\alpha_j}$ we have that $\nu(\pi^{-1}_{\alpha_1,\ldots,\alpha_n}(E)) = \mu_{\alpha_1} \widehat{\times} \cdots \widehat{\times}\, \mu_{\alpha_n}$. Then it's clear that ν defines the same bounded linear functional I on C_F as all the μ_α defined before. So, we can show that $I(f) = \int f \mathrm{d}\nu$ for all $f \in C_F(X)$. And by the Riesz representation theorem, this guarentees that $\mu = \nu$. ■

With this covered we are ready to start constructing sample spaces with our desired infinite sequences of random variables.

Firstly, suppose $\{X_\alpha\}_{\alpha \in A}$ is a collection of random variables (not necessarily independent) on a sample space $(\Omega, \mathcal{F}, P)$. Then for any n-tuple $\alpha_1, \ldots, \alpha_n$ of distinct elements of A (where n is any positive integer), let $P_{(\alpha_1,\ldots,\alpha_n)}$ be the joint distribution of $X_{\alpha_1}, \ldots, X_{\alpha_n}$. If $\mathcal{P}$ is the collection of such joint distributions, then any $P_{(\alpha_1,\ldots,\alpha_n)} \in \mathcal{P}$ must satisfy:

* If σ is a permutation of $\{1, \ldots, n\}$, then:

$$\int \cdot\, \mathrm{d}P_{(\alpha_{\sigma(1)},\ldots,\alpha_{\sigma(n)})}(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = \int \cdot\, \mathrm{d}P_{(\alpha_1,\ldots,\alpha_n)}(x_1, \ldots, x_n).$$

** If $k < n$ and $E \in \mathcal{B}_{\mathbb{R}^k}$, then $P_{(\alpha_1,\ldots,\alpha_k)}(E) = P_{(\alpha_1,\ldots,\alpha_n)}(E \times \mathbb{R}^{n-k})$.

Conversely, for any set A and collection $\mathcal{P}$ of measures $P_{(\alpha_1,\dots,\alpha_n)}$ (where $(\alpha_1,\dots,\alpha_n)$ is an n-tuple of disjoint elements of A and n is any integer) satisfying * and **, we shall show there exists a sample space $(\Omega, \mathcal{F}, P)$ and a family of random variables $\{X_\alpha\}_{\alpha\in A}$ such that $P_{(\alpha_1,\dots,\alpha_n)}$ is the joint distribution of $X_{\alpha_1},\dots,X_{\alpha_n}$.

Now, let $\mathbb{R}^*$ refer to the Alexandroff compactification of $\mathbb{R}$. (I.e. $\mathbb{R}^* = \mathbb{R}\cup\{\infty\}$ and a set $U\subseteq\mathbb{R}^*$ is open iff U is open in $\mathbb{R}$ or U^{C} is compact in $\mathbb{R}$). Note that any Borel measure on $\mathbb{R}^n$ can be regarded as a Borel measure on $(\mathbb{R}^*)^n$ that assigns 0 to the set $(\mathbb{R}^*)^n - \mathbb{R}^n$ and vice versa. For our construction, we will actually let the X_α be measurable functions into $\mathbb{R}^*$ and we will take our family of measures to be measures on $(\mathbb{R}^*)^n$. Thus, the X_α will be allowed to take on the value ∞ but only with probability zero (when we actually make this sample space).

> The reason why this is helpful is that by Tychonoff's theorem, we know that $(\mathbb{R}^*)^A$ is compact and Hausdorff no matter what A is. Also, $(\mathbb{R}^*)^n$ is second countable and satisfies that every open set is σ-compact for every $n\in\mathbb{N}$.

Theorem 10.18: Let A be an arbitrary nonempty set, and suppose that for each ordered n-tuple of distinct elements of A (where $n\in\mathbb{N}$) we are given a Borel probability measure $P_{(\alpha_1,\dots,\alpha_n)}$ on $\mathbb{R}^n$ satisfying * and **. Then there is a unique Radon probability measure P on the compact Hausdorff space $\Omega = (\mathbb{R}^*)^A$ such that $P_{(\alpha_1,\dots,\alpha_n)}$ extended to $(\mathbb{R}^*)^n$ in the manner described prior is the joint distribution of $X_{\alpha_1},\dots,X_{\alpha_n}$ where $X_\alpha:\Omega\to\mathbb{R}^*$ is the αth coordinate function.

Proof:

To start off, it's worth noting that this statement is slightly more involved than just being a restatement of theorem 7.28 since we didn't assume that $P_{\alpha_1,\dots,\alpha_n} = P_{\alpha_1}\times\cdots\times P_{\alpha_n}$.

> Although, if we did assume that $P_{\alpha_1,\dots,\alpha_n} = P_{\alpha_1}\times\cdots\times P_{\alpha_n}$, generally then this theorem would literally be a restatement of theorem 7.28. (Also the coordinate functions X_α would all be independent).

That said, * and ** still give enough structure to let us construct a well-defined positive linear functional I on $C_F((\mathbb{R}^*)^A)$ as we did in the proof of theorem 7.28. And then by identical logic to that of theorem 7.28, we can construct a Radon measure P on $(\mathbb{R}^*)^A$ such that for any $f = g\circ(X_{\alpha_1},\dots,X_{\alpha_n})\in C_F((\mathbb{R}^*)^A)$, we have that:

$$\int f\mathrm{d}P = \int g\mathrm{d}P_{(\alpha_1,\dots,\alpha_n)}.$$

Also it's clear that if $P'_{(\alpha_1,\dots,\alpha_n)}$ is the joint distribution of $X_{\alpha_1},\dots,X_{\alpha_n}$, then $P'_{(\alpha_1,\dots,\alpha_n)} = P_{(\alpha_1,\dots,\alpha_n)}$. After all, both are Radon measures (since they are finite on all compact sets) and for all $g\in C((\mathbb{R}^n)^A)$ we have that:

$$\int g\mathrm{d}P'_{(\alpha_1,\dots,\alpha_n)} = \int g\circ(X_{\alpha_1},\dots,X_{\alpha_n}) = I(g\circ(X_{\alpha_1},\dots,X_{\alpha_n})) = \int g\mathrm{d}P_{(\alpha_1,\dots,\alpha_n)}.$$

Thus by the Riesz representation theorem we know that $P'_{(\alpha_1,\dots,\alpha_n)} = P_{(\alpha_1,\dots,\alpha_n)}$.

Finally, uniqueness is proven in an identical manner to that done when showing theorem 7.28.■

> Side note: now that we have our probability space constructed, if you really want to have the distributions of our random variables be on $\mathbb{R}$ instead of $\mathbb{R}^*$, you can just set X_α equal to anything else when the αth coordinate is ∞.

Corollary 10.19: Suppose $\{P_\alpha\}_{\alpha \in A}$ is a family of probability measures on $\mathbb{R}$. Then there exists a sample space $(\Omega, \mathcal{F}, P)$ and independent random variables $\{X_\alpha\}_{\alpha \in A}$ on Ω such that P_α is the distribution of X_α for every $\alpha \in A$.

The tldr of this is that for any desired collection of joint probability distributions (even uncountable collections) we can always find a sample space and collection of random variables on that space with those joint distributions.

9/24/2025

Since Fall classes are about to start, I'm probably not going to have as much time to work on this journal again. Although on the other hand I'm joining a reading group soon so I might be taking notes for that and putting them here. We'll see! Anyways today I wanted to try and study distributions and generalized functions. So for the next however so while, if I write L^p, assume I'm working with the Lebesgue measure on $\mathbb{R}^n$.

To start off, how in the world do we generalize the notion of functions on $\mathbb{R}^n$?

- Well firstly it's worth noting that since we are typically working with a.e. equivalence classes of functions, we already have generalized our notion of functions a lot. Quite literally we've already bastardized functions to a point that it doesn't make sense to talk about their values at any individual point anymore, and all because we want L^p to be a metric space.
- But in all seriousness, the real strategy for further generalizing functions is to start thinking of functions as linear functionals on vector spaces of functions. For some intution on this:
    - Recall that in chapter 6 of Folland (covered in my math 240b notes), we found for any conjugate exponents p and q that that the map $g \mapsto \int \cdot\, g \mathrm{d}x$ is an isometry (and thus injection) from L^p into $(L^q)^*$. In other words, each function in L^p determines a unique bounded linear functional in $(L^q)^*$.
    - Also, since the Lebesgue measure is semifinite, we can even say that $g \in L^p$ if and only if $\phi \mapsto \int \phi g \mathrm{d}x$ is a bounded linear functional on L^q.
    - Going a step further, let $B_r(x_0)$ denote the ball of radius r about $x_0 \in \mathbb{R}^n$. Then $\phi_r := (m(B_r(x_0)))^{-1} \chi_{B_r(x_0)} \in L^q$ for any $q \geq 1$ and we have by the Lebesgue differentiation theorem that $\lim_{r \to 0} \int \phi_r g \mathrm{d}x = g(x_0)$ for a.e. x_0.
    - All of this is to say that we don't lose any information by considering any $g \in L^p$ as a linear map from L^q into $\mathbb{C}$ instead of as a function from $\mathbb{R}^n$ into $\mathbb{C}$.
- The only issue with merely copying chapter 6 of Folland is that in most cases, $g \mapsto \int \cdot\, g \mathrm{d}x$ is a surjective map from L^p into $(L^q)^*$. Yet the entire reason we want to generalize our notion of functions in the first place is so that we can work with objects that aren't functions by our old definition. Therefore, we will need to modify this approach.

This is only tangentially relevant but I still want to take a minute to formally go through this since I think it's been a point of confusion for me. So here are a bunch of theorems meant to bridge my lack of intuition on nets.

Theorem 1: If (X, ρ) is a metric space such that every Cauchy sequence converges and $\langle x_i \rangle_{i \in I}$ is any net in X such that $\langle \rho(x_{i_1}, x_{i_2}) \rangle_{(i_1, i_2) \in I \times I}$ converges to 0, then there exists $x \in X$ such that $x_i \to x$.

As a reminder: if I is a directed set, then we consider $I \times I$ to be a directed set where $(i_1, i_2) \lesssim (j_1, j_2)$ in $I \times I$ iff $i_1 \lesssim j_1$ and $i_2 \lesssim j_2$ in I.

Proof:
To start off, we need to extract a Cauchy sequence from our net. Luckily, because $\rho(x_{i_1}, x_{i_2}) \to 0$, we know that for any $\varepsilon > 0$ there exists $(j_1, j_2) \in I \times I$ such that $\rho(x_{i_1}, x_{i_2}) < \varepsilon$ for all $(i_1, i_2) \gtrsim (j_1, j_2)$. Thus, we may find a Cauchy sequence $(x_{i(k)})_{k \in \mathbb{N}}$ in $\langle x_i \rangle_{i \in I}$ as follows:

For each $k \in \mathbb{N}$, choose $(j_1, j_2) \in I$ with $\rho(x_{i_1}, x_{i_2}) < \frac{1}{k}$ for all $(i_1, i_2) \gtrsim (j_1, j_2)$. Then choose any $i(k) \in I$ such that $i(k) \gtrsim j_1$, $i(k) \gtrsim j_2$, and $i(k) \gtrsim i(k-1)$ (if $k > 1$). Then clearly $(x_{i(k)})_{k \in \mathbb{N}}$ is a sequence in our net satisfying that $\rho(x_{i(k_1)}, x_{i(k_2)}) < 1/\min(k_1, k_2)$ for all $k_1, k_2 \in \mathbb{N}$.

Now since $(x_{i(k)})_{k \in \mathbb{N}}$ is a Cauchy sequence, we know that it converges to some $x \in X$. So now we just need to show that the entire net $\langle x_i \rangle_{i \in I}$ converges to x, which is equivalent to saying that for any $\varepsilon > 0$ there exists $j \in I$ such that $\rho(x_i, x) < \varepsilon$ for all $i \gtrsim j$. To show this, pick any $k \in \mathbb{N}$ such that $1/k < \varepsilon/2$. Then for any $i \gtrsim i(k)$, we know that $\rho(x_i, x) \leq \rho(x_i, x_{i(k)}) + \rho(x_{i(k)}, x)$. Also, by how we chose $i(k)$, we know that $\rho(x_i, x_{i(k)}) < 1/k$. And since $\rho(x_{i(k)}, \cdot)$ is a (Lipschitz) continuous function from X to $\mathbb{R}$ (due to the reverse triangle inequality), $\rho(x_{i(k)}, x) = \lim_{n \to \infty} \rho(x_{i(k)}, x_{i(n)}) < 1/k$. Hence, $\rho(x_i, x) < \varepsilon/2 + \varepsilon/2$. ■

Theorem 2: If (X, ρ) is a metric space and $\langle x_i \rangle_{i \in I}$ is a net in X which converges to some $x \in X$, then the net $\langle \rho(x_{i_1}, x_{i_2}) \rangle_{(i_1, i_2) \in I \times I}$ converges to 0.

Pick any $\varepsilon > 0$. Then we know there exists $j \in I$ such that $\rho(x_i, x) < \varepsilon/2$ for all $i \gtrsim j$. And in turn, $\rho(x_{i_1}, x_{i_2}) \leq \rho(x_{i_1}, x) + \rho(x, x_{i_2}) < \varepsilon/2 + \varepsilon/2$ for all $(i_1, i_2) \gtrsim (j, j)$. ■

Theorem 3: If X is a topological space and $\langle x_\alpha \rangle_{\alpha \in A}$ is a net that converges to some $x \in X$, then every subnet of X converges to x.

Proof:
Let $\langle x_{\alpha_\beta} \rangle_{\beta \in B}$ be any subnet of $\langle x_\alpha \rangle_{\alpha \in A}$ and consider any neighborhood U of x. Since $x_\alpha \to x$, we know there exists $\alpha_0 \in A$ such that $x_\alpha \in U$ for all $\alpha \gtrsim \alpha_0$. And in turn, we know that there exists $\beta_0 \in B$ such that $x_{\alpha_\beta} \in U$ for all $\beta \gtrsim \beta_0$. This shows that $x_{\alpha_\beta} \to x$ also. ■

Corollary 4: If X is a Hausdorff topological space and $\langle x_\alpha \rangle_{\alpha \in A}$ is a net in X such that $x_\alpha \to x$ for some $x \in X$, then the only cluster point of $\langle x_\alpha \rangle_{\alpha \in A}$ is x.

Proof:
If $y \in X$ is a cluster point of $\langle x_\alpha \rangle_{\alpha \in A}$, then we know our net has a subnet $\langle x_{\alpha_\beta} \rangle_{\beta \in B}$ which converges to y. However, a result I proved for homework in math 240b is that X is Hausdorff if and only if every net converges to at most point in X. Also, by the last theorem we know that $x_{\alpha_\beta} \to x$. The only way this is possible is if $x = y$. ■

Theorem 5: If X is a first countable topological space and y is a cluster point of the net $\langle x_\alpha \rangle_{\alpha \in A}$ in X, then there exists a sequence $(x_{\alpha_n})_{n \in \mathbb{N}}$ in our net that converges to y.

Proof:
Let $\{U_n\}_{n \in \mathbb{N}}$ be a countable neighborhood base for $y \in X$. Also note that by taking intersections, we can assume without loss of generality that $\{U_n\}_{n \in \mathbb{N}}$ is a decreasing sequence of sets. Now we shall construct $(x_{\alpha_n})_{n \in \mathbb{N}}$ as follows. Start off by just letting α_0 be any element of A. Then for each $n \in \mathbb{N}$ choose α_n such that $\alpha_n \gtrsim \alpha_{n-1}$ and $x_{\alpha_n} \in U_n$. Having done that, since $U_m \subseteq U_n$ for all $m > n$ and $x_{\alpha_n} \in U_n$ for all n, we know that $(x_{\alpha_n})_{n \in \mathbb{N}}$ is eventually in every set in $\{U_n\}_{n \in \mathbb{N}}$. ■

As a result, there is very little reason to work with nets as opposed to sequences whenever X is first countable (which as an example is always the case if X is a metric space).

Theorem 6: Suppose $\langle a_i \rangle_{i \in I}$ and $\langle b_i \rangle_{i \in I}$ are two nets in $\mathbb{C}$ indexed by the same directed set such that $a_i \to A$ and $b_i \to B$. Then $(a_i + b_i) \to A + B$ and $(a_i \cdot b_i) \to A \cdot B$.

Proof:
We know that if $a_i \to A$ and $b_i \to B$ in $\mathbb{C}$, then $\langle a_i, b_i \rangle_{i \in I}$ converges to (A, B) in $\mathbb{C}^2$. (see Folland exercise 4.34 which I did for homework in math 240b). And then since the addition and multiplication maps α and m are continuous functions from $\mathbb{C}^2$ to $\mathbb{C}$, we know that $\alpha(a_i, b_i) \to \alpha(A, B) = A + B$ and $m(a_i, b_i) \to m(A, B) = A \cdot B$. ■

Given any set $E \subseteq \mathbb{R}^n$, let us denote $C_c^\infty(E)$ to be the collection of all C_c^∞ functions $f : \mathbb{R}^n \to \mathbb{C}$ satisfying that $\operatorname{supp}(f) \subseteq E$.

Proposition: If $K \subseteq \mathbb{R}^n$ is compact, then $C_c^\infty(K)$ is a Fréchet space with the topology defined by the norms $\phi \mapsto \|\partial^\alpha \phi\|_u$ (where $\alpha \in \{0, 1, 2, \ldots\}^n$).

(see my math 240c notes for info on multi-index notation...)

Proof:
To start off, $\|\partial^\alpha \phi\|_u < \infty$ for all $\phi \in C_c^\infty(K)$ since $\phi(K)$ is compact. Hence each of the (semi)norms defining the topology of $C_c^\infty(K)$ is well-defined. Also, if $\phi \neq 0$, then we know that $\|\partial^\alpha \phi\|_u > 0$ for all $\alpha \in \{0, 1, 2, \ldots\}^n$. Thus by theorems 5.14(c) and 5.16(a) (see my math 240b notes), we know that $C_c^\infty(K)$ is a Hausdorff topological vector space. And since $\{0, 1, 2, \ldots\}^n$ is countable, all that's left to do is show that $C_c^\infty(K)$ is complete.

See note afterwards on why this next bit is WAAAAY overcomplicated...
Let $\langle \phi_i \rangle_{i \in I}$ be a Cauchy net in $C_c^\infty(K)$. In other words, $\langle \phi_{i_1} - \phi_{i_2} \rangle_{(i_1, i_2) \in I \times I}$ converges to 0, which is equivalent by theorem 5.14(b) to saying that $\|\partial^\alpha \phi_{i_1} - \partial^\alpha \phi_{i_2}\|_u \to 0$ for all multi-indexes $\alpha \in \{0, 1, 2, \ldots, \}^n$. But now since $C(K')$ is a complete metric space for

any compact set K' containing K, we know that for each multi-index α there exists a unique continuous function g_α with $\partial^\alpha \phi_i \to g_\alpha$ and $\operatorname{supp}(g_\alpha) \subseteq K$.

Let f be the function such that $\phi_i \to f$. We clearly have that $\operatorname{supp}(f) \subseteq K$ and we claim that $\partial^\alpha f = g_\alpha$ for all multi-indices α (thus proving that f is C^∞).

It suffices to assume by induction that $\partial^\beta f = g_\beta$ for all multi-indexes β such that $|\beta| < |\alpha|$. This simplifies what we need to prove to the following:

If $\langle \phi_i \rangle_{i\in I}$ is a net in $C_c^\infty(K)$ such that $\|\phi_i - g\|_u \to 0$; $\|\frac{\partial}{\partial x_j}\phi_i - h\|_u \to 0$; and K contains the supports of both g and h, then $\frac{\partial}{\partial x_j} g = h$.

Fix any $x \in \mathbb{R}^n$ and set $\psi_i(t) = \frac{1}{t}(\phi_i(x + te_j) - \phi_i(x))$ for all i. Also let $\varepsilon > 0$ and j be such that $\|\frac{\partial}{\partial x_j}\phi_{i_1} - \frac{\partial}{\partial x_j}\phi_{i_2}\|_u < \varepsilon/2$ for all $i_1, i_2 \gtrsim j$. Then given any $i_1, i_2 \gtrsim j$, we have by the mean value theorem that:

$$\begin{aligned}
|\psi_{i_1}(t) - \psi_{i_2}(t)| &= \tfrac{|\phi_{i_1}(x+te_j) - \phi_{i_1}(x) - \phi_{i_2}(x+te_j) + \phi_{i_2}(x)|}{|t|} \\
&= \tfrac{|(\phi_{i_1} - \phi_{i_2})(x+te_j) - (\phi_{i_1} - \phi_{i_2})(x)|}{|t|} \\
&= \tfrac{|t|}{|t|}|\tfrac{\partial}{\partial x_j}(\phi_{i_1} - \phi_{i_2})(x + ae_j)| \text{ (where } a, b \in [0, t]) \\
&= |\tfrac{\partial}{\partial x_j}\phi_{i_1}(x + ae_j) - \tfrac{\partial}{\partial x_j}\phi_{i_2}(x + ae_j)| \\
&= |\tfrac{\partial}{\partial x_j}\phi_{i_1}(x + ae_j) - h(x + ae_j)| + |\tfrac{\partial}{\partial x_j}\phi_{i_2}(x + ae_j) - h(x + ae_j)| \\
&< \varepsilon/2 + \varepsilon/2 = \varepsilon
\end{aligned}$$

This shows that $\langle \psi_i \rangle_{i\in I}$ is uniformly Cauchy on $C(\mathbb{R} - \{0\})$. And in turn, by my tangent earlier we know that there exists a function $\psi \in C(\mathbb{R} - \{0\})$ such that $\|\psi_i - \psi\|_u \to 0$.

Now finally, we need to show that $(\lim_{t\to 0} \psi_i(t)) \to \lim_{t\to 0} \psi(t)$. To do that we just adapt the proof of proposition 116 in my math 140b notes. For reasons I will say shortly, I'm tired of this proof. So I'm not going to skip that.

Harah! We now know that $(\lim_{t\to 0} \psi_i(t)) \to \lim_{t\to 0} \psi(t)$. Or in other words, $\frac{\partial}{\partial x_j}\phi_i(x) \to (\lim_{t\to 0} \psi(t))$, and this proves that $\lim_{t\to 0} \psi(t) = h(x)$. But also note that we clearly have that $\psi(t) = \frac{1}{t}(g(x + te_j) - g(x))$. So $\lim_{t\to 0} \psi(t) = \frac{\partial}{\partial x_j} g(x)$.

With that we've now shown that there exists $f \in C_c^\infty(K)$ such that $\|\partial^\alpha \phi_i - \partial^\alpha f\|_u \to 0$ for all multi-indices α. Thus clearly $\phi_i \to f$ according to all of our norms. ■

So, I wanted to say that the proof I did above was completely unnecessary because there is an exercise I should have done back in 240b that would completely obsolete the need to work with nets as opposed to sequences. I'm going to do that exercise now.

Exercise 5.44: If $\mathcal{X}$ is a first countable topological vector space and every Cauchy sequence in $\mathcal{X}$ converges, then every Cauchy net in $\mathcal{X}$ converges.

(As a side note, a net $\langle x_i \rangle_{i\in I}$ in $\mathcal{X}$ is Cauchy iff $\langle x_{i_1} - x_{i_2} \rangle_{(i_1,i_2)\in I\times I}$ converges to $0 \in \mathcal{X}$.)

Proof:
You may note that the following proof is basically just an adaption of the proof to *theorem 1* on page 225. Let $\{U_n\}_{n\in\mathbb{N}}$ be a countable neighborhood base for $0 \in \mathcal{X}$. Then we can construct a Cauchy sequence $(x_{i(k)})_{k\in\mathbb{N}}$ in $\langle x_i\rangle_{i\in I}$ as follows:

For each $k \in \mathbb{N}$ choose $(j_1^{(k)}, j_2^{(k)}) \in I$ with $x_{i_1} - x_{i_2} \in U_k$ for all $(i_1, i_2) \gtrsim (j_1^{(k)}, j_2^{(k)})$. Then choose any $i(k) \in I$ such that $i(k) \gtrsim j_1^{(k)}$, $i(k) \gtrsim j_2^{(k)}$, and $i(k) \gtrsim i(k-1)$ (if $k > 1$). Having done this for all k, it's clear that $(x_{i(k)})_{k\in\mathbb{N}}$ is a sequence in our net satisfying that $x_{i(k_1)} - x_{i(k_2)} \in U_n$ for all $n \leq \min(k_1, k_2)$.

Now we know that there exists $x \in \mathcal{X}$ such that $x_{i(k)} \to x$ as $k \to \infty$. Hence, all that's left to show is that the net as a whole converges to x. To prove this, let $\{V_n\}_{n\in\mathbb{N}}$ be a countable neighborhood base for x.

For any $n \in \mathbb{N}$, since vector addition is a continuous map from $\mathcal{X} \times \mathcal{X}$ to $\mathcal{X}$, we know that the set $E := \{(y, z) \in \mathcal{X} \times \mathcal{X} : y + z \in V_n\}$ is a neighborhood of $(0, x) \in \mathcal{X} \times \mathcal{X}$. In turn, we know there are two open sets $W_1, W_2 \subseteq \mathcal{X}$ such that $(0, x) \in W_1 \times W_2 \subseteq E$. And going a step further there exists neighborhoods U_r of 0 and V_s of x in our countable neighborhood bases such that so long as $(y, z) \in U_r \times V_s$, then $y + z \in V_n$.

But now consider that since $x_{i(k)} \to x$ as $k \to \infty$, there exists $N \in \mathbb{N}$ such that $x_{i(k)} \in V_s$ for all $k \geq N$. Also, based on how we constructed our sequence, we know $x_i - x_{i(k)} \in U_r$ for all $k \geq r$ and $i \gtrsim i(r)$. Thus let $j = i(r)$. Then for any $i \gtrsim j$, we have that $x_i = (x_i - x_{i(\max(N,r))}) + x_{i(\max(N,r))} \in V_n$. And this proves that $\langle x_i\rangle_{i\in I}$ converges to x. ■

The significance of this result is that if a topological vector space has a topology induced by any countable collection of seminorms, then $\mathcal{X}$ is easily seen to be first countable. So, in my ill-advised proof I did to show that $C_c^\infty(K)$ is a Fréchet space (for example), there was absolutely no reason for me to talk about Cauchy nets. I should have just talked about Cauchy sequences. And then, I could have used helpful theorems like the dominated convergence theorem and the fundamental theorem of calculus to write a much neater and shorter proof.

One final note for today. If $U \subseteq \mathbb{R}^n$ is open, then $C_c^\infty(U)$ is the union of all the $C_c^\infty(K)$ where K is a compact subset of U. Now for any open $U \subseteq X$ we state the following as definitions (technically there is a topology such that all these things are true but Folland says it's not really important to know):

1. A sequence $(\phi_j)_{j\in\mathbb{N}}$ in $C_c^\infty(U)$ converges in C_c^∞ to ϕ if $\{\phi_j\}_{j\in\mathbb{N}} \subseteq C_c^\infty(K)$ for some compact $K \subseteq U$ and $\phi_j \to \phi$ in the topology of $C_c^\infty(K)$.
2. If $\mathcal{X}$ is a locally convex topological vector space and $T : C_c^\infty(U) \to \mathcal{X}$ is a linear map, T is continuous if $T|_{C_c^\infty(K)}$ is continuous for each compact set $K \subseteq U$. In other words, T is continuous if whenever $(\phi_j)_{j\in\mathbb{N}}$ is a sequence in $C_c^\infty(U)$ converging in C_c^∞ to ϕ, then $T\phi_j \to T\phi$.

3. A linear map $T : C_c^\infty(U) \to C_c^\infty(U')$ is continuous if for each compact set $K \subseteq U$ there exists a compact set $K' \subseteq U'$ such that $T(C_c^\infty(K)) \subseteq C_c^\infty(K')$ and T is continuous from $C_c^\infty(K)$ to $C_c^\infty(K')$.
4. A distribution on U is a continuous linear functional on $C_c^\infty(U)$. The space of all distributions on U is denoted by $\mathcal{D}'(U)$, and we set $\mathcal{D}' := \mathcal{D}'(\mathbb{R}^n)$. We impose the weak* topology on $\mathcal{D}'$.

9/25/2025

I've decided for the time being I'm going to try doing all my class work in my journal. Also, as I cover more math, I'm going to feel less and less need to take notes on everything I see in lecture. So expect my lecture notes here to be less organized than my independent studying I guess.

Math 241a (lecture 1):

Note that this class uses the textbook *Essential Results of Functional Analysis* by Robert J. Zimmer. So if I number any lemma, theorem, or definition in my notes for math 241a, assume it came from that book.

Let $\mathcal{X}$ denote a K-vector space where K is either $\mathbb{R}$ or $\mathbb{C}$. Also, although I won't mention it again, assume $\mathcal{X}$ isn't trivial (i.e. a singleton).

Proposition: If X is a topological space, $E \subseteq X$ is open iff for any $x \in E$ and net $\langle x_i \rangle_{i \in I}$ converging to x we have that $\langle x_i \rangle_{i \in I}$ is eventually in E.

$(\Longrightarrow)$
Suppose $x \in E$ and $\langle x_i \rangle_{i \in I}$ is a net converging to x such that x_i isn't eventually in E. Then by definition E can't be a neighborhood of x. And thus since $E - E^\circ \neq \emptyset$ since it includes x, we know that E is not open.

$(\Longleftarrow)$
Suppose E is not open. Then $E - E^\circ \neq \emptyset$. Hence, we can pick $x \in E - E^\circ$. But now $x \in (E^\circ)^{\mathsf{C}} = \overline{E^{\mathsf{C}}}$. So, there exists a net $\langle x_i \rangle_{i \in I}$ in E^{C} with $x_i \to x$ (by Folland proposition 4.18 in my math 240b notes). And clearly $\langle x_i \rangle_{i \in I}$ is not eventually in E. ■

Corollary: Let X be a set and $\mathcal{T}$ and $\mathcal{T}'$ be two topologies such that for any net $\langle x_i \rangle_{i \in I}$ in X and $x \in X$ we have that $x_i \to x$ with respect to $\mathcal{T}$ if $x_i \to x$ with respect to $\mathcal{T}'$. Then $\mathcal{T} \subseteq \mathcal{T}'$.

Proof:
Suppose $U \in \mathcal{T}$ and then consider any $x \in U$ and net $\langle x_i \rangle_{i \in I}$ such that $x_i \to x$ with respect to $\mathcal{T}'$. By assumption we also know that $x_i \to x$ with respect to $\mathcal{T}$, and thus $\langle x_i \rangle_{i \in I}$ is eventually in U. It follows by our prior lemma that $U \in \mathcal{T}'$, thus showing that $\mathcal{T} \subseteq \mathcal{T}'$. ■

In math 240b (Folland theorem 5.14), we proved that if $\mathcal{X}$ is a vector space and $\{p_\alpha\}_{\alpha \in A}$ is a family of seminorms, then the topology induced by those seminorms is locally convex, meaning there exists a base for the topology consisting entirely of convex sets. As it turns out, we can prove a converse statement.

To start off, here is a useful result about convexity. (As a reminder, $E \subseteq \mathcal{X}$ is convex if for all $x, y \in E$ and $t \in [0,1]$, $(1-t)x + ty \in E$.)

Lemma: If $\{E_\alpha\}_{\alpha \in A}$ is a collection of convex sets in a vector space $\mathcal{X}$, then $E := \bigcap_{\alpha \in A} E_\alpha$ is convex.

Proof:

Let $x, y \in E$ and $t \in [0,1]$. Then $(1-t)x + ty \in E_\alpha$ for all $\alpha \in A$. Hence for all $t \in [0,1]$ we have that $(1-t)x + ty \in \bigcap_{\alpha \in A} E_\alpha = E$. ■

It follows that given any $E \subseteq \mathcal{X}$, we can define the convex hull of E (denoted $\mathrm{conv}(E)$) to be the unique smallest convex set in $\mathcal{X}$ containing E.

Proposition: If $\mathcal{X}$ is a vector space and $E \subseteq \mathcal{X}$, then:

$$\mathrm{conv}(E) = \left\{ \sum_{j=1}^{n} \lambda_j x_j : n \in \mathbb{N}; x_1, \ldots, x_n \in E; \lambda_1, \ldots, \lambda_n > 0; \text{ and } \sum_{j=1}^{n} \lambda_j = 1 \right\}$$

Proof:

Let us denote the set on the right-hand side by C. When $n = 2$, it's clear that any $\sum_{j=1}^{n} \lambda_j x_j \in C$ must also be in $\mathrm{conv}(E)$. And by a simple enough induction argument, we can prove that $C \subseteq \mathrm{conv}(E)$.

Let $\sum_{j=1}^{n} \lambda_j x_j \in C$ (where $n > 2$). If $\lambda_n = 1$, then it's trivial that:

$$\sum_{j=1}^{n} \lambda_j x_j = x_n \in E \subseteq \mathrm{conv}(E).$$

Meanwhile, if $\lambda_n < 1$, then $(1-\lambda_n)^{-1} \sum_{j=1}^{n-1} \lambda_j = 1$ and hence we know by induction that $(1-\lambda_n)^{-1} \sum_{j=1}^{n-1} \lambda_j x_j \in \mathrm{conv}(E)$. And in turn we know that:

$$\sum_{j=1}^{n} \lambda_j x_j = (1-\lambda_n)\left((1-\lambda_n)^{-1} \sum_{j=1}^{n-1} \lambda_j x_j \right) + \lambda_n x_n \in \mathrm{conv}(E)$$

Meanwhile, if we can show that C is convex, then we must have that $\mathrm{conv}(E) \subseteq C$. Luckily consider any $\sum_{i=1}^{n} \lambda_i x_i \in C$ and $\sum_{j=1}^{m} \mu_j y_j \in C$ and fix $t \in [0,1]$. Then since $\sum_{i=1}^{n} \lambda_i = 1 = \sum_{j=1}^{m} \mu_j$, we can easily see that $(1-t)\sum_{i=1}^{n} \lambda_i + t \sum_{j=1}^{m} \mu_j = 1$. And so, we must have that:

$$(1-t)\left(\sum_{i=1}^{n} \lambda_i x_i\right) + t\left(\sum_{j=1}^{m} \mu_j y_j\right) \in C. \blacksquare$$

Next, a set $E \subseteq \mathcal{X}$ is called balanced if for all $c \in K$ with $|c| \leq 1$, we have that $cE \subseteq E$ (where $cE := \{cx \in \mathcal{X} : x \in E\}$). Concerning convexity and the property of being balanced, we have the following three lemmas:

Proposition: Suppose $\mathcal{X}$ is a topological vector space and $E \subseteq \mathcal{X}$ is balanced. Then $\mathrm{conv}(E)$ is balanced as well.

Proof:

Consider any $\sum_{j=1}^{n} \lambda_j x_j \in \mathrm{conv}(E)$ where $x_1, \ldots, x_n \in E$; $\lambda_1, \ldots, \lambda_n > 0$; and $\sum_{j=1}^{n} \lambda_j = 1$. Then let $c \in K$ with $|c| \leq 1$. Since E is balanced, we know that $cx_j \in E$ for each j. And hence $c(\sum_{j=1}^{n} \lambda_j x_j) = \sum_{j=1}^{n} \lambda_j (cx_j) \in \mathrm{conv}(E)$. ■

Proposition: Suppose $\mathcal{X}$ is a topological vector space.

(a) If N is a neighborhood of 0, there is a balanced neighborhood B of 0 with $B \subseteq N$.

Proof:

By setting $N = N^\circ$ we can without loss of generality assume that N is open. Then since scalar multiplication is continuous, we know the set $W := \{(c,x) \in K \times \mathcal{X} : cx \in N\}$ is open in $K \times \mathcal{X}$. Also, clearly $(0,0) \in W$. Hence, there are open sets $U \subseteq K$ and $V \subseteq \mathcal{X}$ such that $(0,0) \in U \times V \subseteq W$. Or in other words, there exists $\varepsilon > 0$ and an open set $V \subseteq \mathcal{X}$ such that for all $c \in K$ with $|c| < \varepsilon$ and $x \in V$, $cx \subseteq N$.

Now note that cV is open for all $c \in K$ with $c \neq 0$. This is because $x \mapsto cx$ is a continuous function on $\mathcal{X}$ with a continuous inverse function $x \mapsto x/c$. Also, since $0 \in V$, we know that $0 \in cV$ for all $c \in K$. Thus $0V \subseteq cV$ for all $c \in K$ and we don't need to worry about the edge case that $0V$ isn't open (I haven't shown it yet but it is true that singletons can't be open in a topological vector space). Consequently, we know that $B := \bigcup_{|c|<\varepsilon} cV$ is an open set in $\mathcal{X}$. And clearly $0 \in B \subseteq N$.

What's left to do now is show that B is balanced. Consider any $d \in K$ with $|d| \leq 1$. Then $|dc| \leq 1|c| < \varepsilon$ for all $c \in K$ with $|c| < \varepsilon$. Hence, for all $d \in K$ with $|d| \leq 1$:

$$dB = \{dx \in \mathcal{X} : x \in \textstyle\bigcup_{|c|<\varepsilon} cV\}$$
$$= \textstyle\bigcup_{|c|<\varepsilon}\{dx \in \mathcal{X} : x \in cV\} = \bigcup_{|c|<\varepsilon}\{dcy \in \mathcal{X} : y \in V\} = \bigcup_{|c|<\varepsilon}(dc)V \subseteq N.$$

(b) If additionally N is convex, then we can choose B to be convex.

Proof:

By the reasoning described in part (a), let $A \subseteq N$ be a balanced neighborhood of $0 \in \mathcal{X}$. Then set $B = \operatorname{conv}(A)$. Since A is balanced, we know B is also. Also, since N is convex, we know that $B \subseteq N$. And finally, since $A^\circ \subseteq B^\circ$, we know that B is also a neighborhood of 0 in $\mathcal{X}$. ■.

Proposition: Let $\mathcal{X}$ be a topological vector space.

(a) If $E \subseteq \mathcal{X}$ is balanced and $0 \in E^\circ$, then E° is balanced.

Proof:

To start off, fix any $c \in K$ with $0 < |c|$. Then note that $x \mapsto cx$ is a continuous map in $\mathcal{X}$ with a continuous inverse $x \mapsto x/c$. So if U is any open set in $\mathcal{X}$, then cU and $(1/c)U$ are also open.

If $V \subseteq E$ is open in $\mathcal{X}$. Then we clearly have the $cV \subseteq cE$ and we know from before that cV is open. Consequently, we know that $cE^\circ \subseteq (cE)^\circ$. On the other hand, if $W \subseteq cE$ is open, then $(1/c)W \subseteq E$; $(1/c)W$ is open; and $W = c((1/c)W)$ is in cE°. From this we know that $(cE)^\circ \subseteq cE^\circ$. And hence we've proven that $(cE)^\circ = cE^\circ$.

To finish off the main section of the proof, note that since E is balanced, we have that $cE \subseteq E$ whenever $|c| \leq 1$. Thus $cE^\circ = (cE)^\circ \subseteq E^\circ$. And therefore we have shown that $cE^\circ \subseteq E^\circ$ if $0 < |c| \leq 1$.

As for the annoying case of $c = 0$, note that $0E^\circ = \{0\} \subseteq E^\circ$. ■

As a side note, we do have to explicitely assume that E is a neighborhood of $0 \in \mathcal{X}$ because there are loads of examples of balanced sets which don't have 0 as an interior point. For example, if $K = \mathbb{R}$, then consider the closure of the first and third quadrants in the real number plane (equipped with the euclidean norm). There are other examples online for when $K = \mathbb{C}$.

(b) If $E \subseteq \mathcal{X}$ is convex, then so is E°.

Proof:

We first need a lemma. If U and V are open sets in $\mathcal{X}$, then we have that $U + V = \{x + y \in \mathcal{X} : x \in U, y \in V\}$ is open.

If either U or V is empty, then this is trivial. Otherwise, note that:

$$U + V = \bigcup_{x \in U}(x + V) \text{ (where } x + V = \{x + y \in \mathcal{X} : y \in V\}\text{)}.$$

Since $y \mapsto y + x$ is a continuous map with a continuous inverse $y \mapsto y - x$, we know that $x + V$ is open iff V is open. So, $U + V$ is an arbitrary union of open sets. It follows that $U + V$ is open.

Choose any $t \in [0, 1]$. If $t \in \{0, 1\}$, then we trivially have that $(1 - t)E^\circ + tE^\circ = E^\circ$ is open. Otherwise, note from the lemma we just showed plus the fact that cE° is open for any $c \in K$ with $|c| > 0$ that $(1 - t)E^\circ + tE^\circ$ is an open set. Also, by using the convexity of E we can easily check that $(1 - t)E^\circ + tE^\circ \subseteq E$. Hence, we can conclude that $(1 - t)E^\circ + tE^\circ \subseteq E^\circ$, and this proves that E° is convex since:

$$\bigcup_{t \in [0,1]} [(1 - t)E^\circ + tE^\circ] \subseteq E^\circ. \blacksquare$$

The main take away from the last three propositions is that if N is any neighborhood of 0, then we can find an open balanced set U such that $0 \in U \subseteq N$. And if additionally N is convex, we can also take U to be convex.

As for why we like working with balanced sets, consider the following fact. If $V \subseteq \mathcal{X}$ is balanced and $s \in K$, then sV is balanced.

Why? Suppose $c \in K$ with $|c| \leq 1$. Then $c(sV) = (cs)V = s(cV) \subseteq sV$.

Consequently, if $V \subseteq \mathcal{X}$ is balanced and $c_1, c_2 \in K$ with $|c_1| \leq |c_2|$, then $c_1K \subseteq c_2K$.

Why? If $c_2 = 0$, then both c_1V and c_2V equal $\{0\}$. Otherwise, we know that c_2V is balanced and that $|c_1/c_2| \leq 1$. Hence, $c_1V = (c_1/c_2)c_2V \subseteq c_2V$.

And as a side note: in the special case that $|c_1| = |c_2|$ we can apply the prior reasoning twice to say that $c_1V = c_2V$. In particular, this means that $cV = |c|V$ for all $c \in K$.

One more piece of terminology: a set E is <u>absorbing</u> if for all $x \in \mathcal{X}$ there exists a finite $r > 0$ such that for all $s \in K$ with $|s| \geq r$ we have that $x \in sE$.

<u>Lemma:</u> If $E \subseteq \mathcal{X}$ is a neighborhood of 0, then E is absorbing.

Proof:

Let $x \in \mathcal{X}$. Then since $c \mapsto cx$ is a continuous map and $0x \in E^\circ$, we know there is a neighborhood $N \subseteq K$ of 0 such that $cx \in E$ for all $c \in N$. In particular, pick $r > 0$ such that the ball $B_{1/r}(0)$ of radius $1/r$ about 0 in K is a subset of N. Then for any $s \in K$ with $|s| \geq 2r$ (so that $|s^{-1}| \leq 1/(2r) < r^{-1}$), we have that $s^{-1}x \in E$. Or in other words, for all $s \in K$ with $|s| \geq 2r$ we have that $x \in sE$. This shows that E is absorbing. $\blacksquare$

As a side note, this lemma shows that $\{0\}$ can't be open since $\{0\}$ isn't absorbing. And in turn since vector addition is continuous, we know that no singleton contained in $\mathcal{X}$ can be open. (This guarentees the topology on $\mathcal{X}$ is *not* discrete!!)

Now the next two proofs are sourced/inspired by a pdf I found online (see the *12th document* cited in my bibliography...). Let $\mathcal{X}$ be a topological vector space and $U \subseteq \mathcal{X}$ be any open balanced convex set containing 0. Then the Minkowski functional associated to U is $p_U(x) := \inf\{t \geq 0 : x \in tU\}$.

Claim: $p_U : \mathcal{X} \to [0, \infty)$ is a well-defined seminorm on $\mathcal{X}$ (i.e. $p_U(x)$ exists and is finite for all $x \in \mathcal{X}$, and p_U satisfies the two axioms of a seminorm). Also, p_U is continuous.

Proof:

Firstly, since U is absorbing, we know that for any $x \in \mathcal{X}$ there exists some finite $t > 0$ with $x \in tU$. Hence $\{t \geq 0 : x \in tU\}$ is nonempty (and contains finite values), and it easily follows that $p_U(x)$ exists and is finite for all $x \in \mathcal{X}$.

Next, we show that p_U is a seminorm.

- Suppose $c \in K$ and $x \in \mathcal{X}$. Then for any $t \geq 0$ with $x \in tU$, we have that $cx \in ctU = |c|tU$. And in turn we can easily see that:
$$p_U(cx) \leq \inf\{|c| \cdot t \geq 0 : x \in tU\} = |c| \cdot p_U(x).$$

  By similar reasoning, if $cx \in tU$ for some $t \geq 0$ and $c \neq 0$, then $x \in c^{-1}tU = |c^{-1}|tU$. So $p_U(x) \leq \inf\{|c|^{-1} \cdot t \geq 0 : cx \in tU\} = |c|^{-1}p_U(cx)$. I.e. $|c| \cdot p_U(x) \leq p_U(cx)$, and this proves that $p_U(cx) = |c|p_U(x)$ for all $c \in K$ with $c \neq 0$. As for the case that $c = 0$, then we trivially have that $p_U(0x) = p_U(0) = 0 = |0| \cdot p_U(x)$.

- To show the triangle inequality, we finally use the convexity of U. Consider any $x, y \in \mathcal{X}$ and pick $s, t > 0$ such that $x \in sU$ and $y \in tU$. Then:
$$x + y \in sU + tU = \{su + tu' : u, u' \in U\}.$$

  Now by the convexity of U, we know that $\frac{s}{s+t}u + \frac{t}{s+t}u' \in U$ for all $u, u' \in U$. Hence $su + tu' = (s+t)(\frac{s}{s+t}u + \frac{t}{s+t}u') \in (s+t)U$ and this proves that $sU + tU \subseteq (s+t)U$. So, we've shown that $p_U(x+y) \leq \inf\{s + t : s, t \geq 0, x \in sU, \text{ and } y \in tU\}$.

  Finally, we just need to show that:
$$\inf\{s + t : s, t \geq 0, x \in sU, \text{ and } y \in tU\} \leq \inf\{s \geq 0 : x \in sU\} + \inf\{t \geq 0 : y \in tU\}.$$

  As shorthand, write $\alpha = \inf\{s \geq 0 : x \in sU\}$, $\beta = \inf\{t \geq 0 : y \in tU\}$, and $\gamma = \inf\{s + t : s, t \geq 0, x \in sU, \text{ and } y \in tU\}$. Then consider any $\varepsilon > 0$ and let $s, t \geq 0$ satisfy that $x \in sU$; $y \in tU$; $s < \alpha + {}^{\varepsilon}/_{2}$; and $t < \beta + {}^{\varepsilon}/_{2}$. But now clearly $\gamma \leq s + t < \alpha + \beta + \varepsilon$. And by taking $\varepsilon \to 0$ we get that $\gamma \leq \alpha + \beta$.

Lastly, we need to show that p_U is continuous (annoyingly the paper I'm looking at just forgets about this step). To prove this, note that p_U is continuous at some $x \in \mathcal{X}$ iff for all nets $\langle x_i \rangle_{i \in I}$ converging to x, we have that $p_U(x_i) \to p_U(x)$ in K. So consider any net $\langle x_i \rangle_{i \in I}$ converging to some $x \in \mathcal{X}$. Then notice that $p_U(x_i) \to p_U(x)$ iff $|p_U(x_i) - p_U(x)| \to 0$. Also, by reverse triangle inequality we have that $|p_U(x_i) - p_U(x)| \leq p_U(x_i - x)$. Hence, we will be done once we show that for all $\varepsilon > 0$ there exists $j \in I$ such that $p_U(x_i - x) < \varepsilon$ for all $i \gtrsim j$.

Since $x_i \to x$ and vector addition is continuous, we know that $x_i - x \to 0$ in $\mathcal{X}$. Also note that if $t > 0$, then we showed while proving an *earlier proposition* that tU is also an open set containing 0. So becase $x_i - x \to 0$, we must have that $x_i - x$ is eventually in tU for all $t > 0$. This is equivalent to saying that for all $t > 0$ there exists $j \in I$ such that $p_U(x_i - x) \leq t$ for all $i \gtrsim j$. Hence, $p_U(x_i - x) \to 0$. ■

Theorem: If $\mathcal{X}$ is a topological vector space that is locally convex, then the topology of $\mathcal{X}$ is identical to the topology generated by a collection of Minkowski functionals associated with a neighborhood base of 0 consisting of convex balanced open sets.

Proof:

If $\mathcal{X}$ is locally convex, then we know there is a neighborhood base of $0 \in \mathcal{X}$ consisting of open convex sets. And by an earlier proposition we can then easily acquire a neighborhood base $\{U_\alpha\}_{\alpha \in A}$ of 0 consisting entirely of balanced open convex sets.

Now for each $\alpha \in A$ let p_{U_α} be the Minkowski functional associated to U_α. If we can prove that a net $\langle x_i \rangle_{i \in I}$ converges to x iff $p_{U_\alpha}(x_i - x) \to 0$ for all $\alpha \in A$, then we will know by the corollary at the *bottom of page 229* that the topology on $\mathcal{X}$ is equal to the topology generated by the seminorms.

$(\Longrightarrow)$

Suppose $x_i \to x$. Then that implies that $x_i - x \to 0$, which means that $x_i - x$ is eventually in every neighborhood of 0. In particular, consider any $\alpha \in A$ and $t > 0$. Then $x_i - x$ is eventually in tU_α. Or equivalently, there exists $j \in I$ such that $p_{U_\alpha}(x_i - x) \leq t$ for all $i \gtrsim j$. This proves that $p_{U_\alpha}(x_i - x) \to 0$ for all $\alpha \in A$.

$(\Longleftarrow)$

Suppose $p_{U_\alpha}(x_i - x) \to 0$ for all $\alpha \in A$ and consider any neighborhood N of 0. Since $\{U_\alpha\}_{\alpha \in A}$ is a neighborhood base of 0, we know there exists $\alpha \in A$ such that $U_\alpha \subseteq N$. Then because U_α is balanced and we eventually have that $p_{U_\alpha}(x_i - x) \leq 1$, we know that x_i is eventually in $U_\alpha \subseteq N$. This proves that $x_i - x \to 0$. And since vector addition is continuous, we in turn know that $x_i \to x$. ■

So, we've proven that a topological vector space is locally convex if and only if the topology is generated by a family of seminorms. Can we go one step further and give a necessary and sufficient condition for a topological vector space to be normable? As it turns out, yes.

If $\mathcal{X}$ is a topological vector space, a set $B \subseteq \mathcal{X}$ is called von Neumann bounded (or just bounded) if for every neighborhood N of 0 in $\mathcal{X}$ there exists some $r > 0$ such that for all $s \in K$ with $|s| \geq r$ we have that $B \subseteq sN$.

From the above definition it's obvious that if $\mathcal{X}$ is a topological vector space, $B \subseteq \mathcal{X}$ is bounded, and $E \subseteq B$, then we also have that E is bounded. Also, it's worth noting the folowing equivalent properties to boundedness.

Proposition: If N is a neighborhood of 0 in a topological vector space $\mathcal{X}$ and $B \subseteq \mathcal{X}$ is bounded, then there exists $r > 0$ such that $sB \subseteq N$ for all $s \in K$ with $|s| \leq r$.

$(\Longrightarrow)$

Note that we trivially have that $0B = \{0\} \subseteq N$. Meanwhile, if B is bounded, then suppose $B \subseteq sN$ for all $s \in K$ with $|s| \geq r > 0$. Then $s^{-1}B \subseteq N$ when $|s| \geq r$. And since for all $s' \in K$ with $0 < |s'| < 1/r$ there exists $(s')^{-1} \in K$ with $|(s')^{-1}| \geq r$, we can conclude $s'B \subseteq N$ for all $s' \in K$ with $0 < |s'| \leq 1/r$.

$(\Longleftarrow)$

The proof of this direction is analogous to that of the prior direction. So I'm omitting it. ■

Proposition: $B \subseteq \mathcal{X}$ is bounded if and only if for every balanced open set $U \subseteq \mathcal{X}$ containing 0 there exists $r \in K$ such that $B \subseteq rU$.

Proof:
The forward direction is trivial. Meanwhile, to show the backwards direction, consider any neighborhood N of 0 in $\mathcal{X}$. Then we know from earlier that there is a balanced open set $U \subseteq \mathcal{X}$ such that $0 \in U \subseteq N$. Also, by the hypothesis of our proposition we know there exists $r \in K$ such that $B \subseteq rU$. Plus, we know that $rU = |r|U$ since U is balanced. Hence, without loss of generality we can say r is real and nonnegative. Finally, we know $B \subseteq sU$ for all $s \in K$ with $|s| \geq r$ (since U is balanced). And since $U \subseteq N$, we in turn know that $B \subseteq sN$ for all $s \in K$ with $|s| \geq r$. ■

Now a question one may ask is whether this definition of boundedness is compatible with our definition of boundedness on metric spaces. Unfortunately, the answer is generally no. For example, consider $\mathbb{R}$ as a real vector space and then equip $\mathbb{R}$ with the metric $\rho(z_1, z_2) := |\arctan(x_1) - \arctan(x_2)|$.

Sanity check:
Note that ρ is in fact a metric. After all, $\arctan$ is an injective function. So $\rho(x_1, x_2) = \arctan(x_2) - \arctan(x_1) = 0$ iff $x_1 = x_2$. And if $x_1 \neq x_2$ we have that $\rho(x_1, x_2) > 0$. Also note that clearly $\rho(z_1, z_2) = \rho(z_2, z_1)$. Finally, suppose $x, y, z \in \mathbb{R}$. Then $|\arctan(x) - \arctan(y)| \leq |\arctan(x) - \arctan(z)| + |\arctan(y) - \arctan(z)|$. This shows that triangle inequality is fulfilled.

Then it is easy to see that the entire space $\mathbb{R}$ is bounded with respect to ρ according to the metric space definition of boundedness. Also, it is easy to see that ρ generates the same topology on $\mathbb{R}$ as the Euclidean metric. Hence, $(\mathbb{R}, \rho)$ is a real topological vector space. That said, if B is the Euclidean open unit ball about 0, then clearly B is balanced and contains 0 but there does not exist any $r \in \mathbb{C}$ such that $\mathbb{R} \subseteq rB$. So $\mathbb{R}$ is not von Neumann bounded in the topology induced by ρ.

Note from 10/3/2025: I originally had a different counter example here but I realized it was wrong and replaced it.

I originally wrote $\rho(z_1, z_2) = \min(|z_1, z_2|, 1)$. The issue with this metric is that it makes the closed Euclidean unit ball open with respect to ρ. So ρ doesn't generate the same topology as the Euclidean metric. Also, it's easy to check $\{z \in \mathbb{C} : |z| \leq \frac{1}{2}\}$ is not open even with respect to ρ. So ρ doesn't even define a topological vector space.

At the very least however, if $\mathcal{X}$ is a normed K-vector space (which automatically makes $\mathcal{X}$ a topological vector space), then the metric space definition of being bounded according to the norm $\|\cdot\|$ is equivalent to being von Neumann bounded.

$(\Longrightarrow)$
Let $E \subseteq \mathcal{X}$ and suppose there exists $C > 0$ such that $\|x\| < C$ for all $x \in E$. Then consider any balanced open set $U \subseteq \mathcal{X}$ containing 0. We know that there exists a ball $B_\varepsilon(0)$ of radius ε about 0 such that $B_\varepsilon(0) \subseteq U$. Also, we clearly have that $E \subseteq (C\varepsilon^{-1})B_\varepsilon(0)$ since the latter set contains all $x \in \mathcal{X}$ with $\|x\| < C$. Therefore, we also have that $E \subseteq (C\varepsilon^{-1})U$. And this shows that E is von-Neumann bounded.

$(\Longleftarrow)$
Suppose $E \subseteq \mathcal{X}$ is Von-Neumann bounded. Then the ball $B_1(0)$ of radius 1 about $0 \in \mathcal{X}$ is a neighborhood of 0 in $\mathcal{X}$. So, there exists some $r > 0$ with $E \subseteq rB_1(0) = B_r(0)$. Or in other words, $\|x\| < r$ for all $x \in E$. ■

As a side note: what sets the norm metric apart from just any metric in this proof is that the norm metric lets us say that $(C\varepsilon^{-1})B_\varepsilon(0) = B_C(0)$ and $rB_1(0) = B_r(0)$.

With that we now have all the background we need to find a necessary and sufficient condition for a topology on $\mathcal{X}$ to be a normed topology. Note that every normed vector space $\mathcal{X}$ is T_1 (in fact $\mathcal{X}$ is normal since all metric spaces are normal). Also, it's a trivial observation that every normed vector space $\mathcal{X}$ has a nonempty convex open von Neumann bounded set. For example, just take the open ball of radius 1 about $0 \in \mathcal{X}$. As it turns out, these are all the topological properties needed of a topological vector space in order to show that the space is normable.

Lemma: If $\mathcal{X}$ is a topological vector space, $y \in \mathcal{X}$, and $E \subseteq \mathcal{X}$ is von Neumann bounded, then $y + E \coloneqq \{x + y \in \mathcal{X} : x \in E\}$ is von Neumann bounded.

Proof:
For any balanced open set $U \subseteq \mathcal{X}$ containing 0, we know since E is bounded that there exists $t > 0$ such that $E \subseteq tU$. Also, since U is absorbing, we know there exists $s > 0$ such that $y \in sU$. And finally, note that $sU \subseteq tU$ if $s \leq t$ and $tU \subseteq sU$ if $t \leq s$ (this is true because U is balanced). ■

Lemma: If $\mathcal{X}$ is a topological vector space, $y \in \mathcal{X}$, and $E \subseteq \mathcal{X}$ is convex, then $y + E$ is convex.

Proof:
If E is empty or a singleton, then E is trivially convex. Otherwise, consider any two points $y + x_1$ and $y + x_2$ in $y + E$ and let $t \in [0, 1]$. Then:

$$(1 - t)(y + x_1) + t(y + x_2) = y + ((1 - t)x_1 + tx_2).$$

And since E is convex, we know that $y + ((1 - t)x_1 + tx_2) \in y + E$. ■

Theorem: Suppose $\mathcal{X}$ is a T_1 topological vector space and there exists a nonempty convex open von Neumann bounded set $U \subseteq \mathcal{X}$. Then the topology on $\mathcal{X}$ is identical to a norm-topology generated by a Minkowski functional.

Proof:
If we pick any $x \in U$, then we know by the last two lemmas plus the continuity of vector addition that $(-x)+U$ is a convex open von Neumann bounded set containing 0. So, without loss of generality we can assume U contains 0. Then by restricting U further, we can also assume without loss of generality that U is balanced.

Next, we know there is a Minkowski functional p_U associated with U which we already proved to be a seminorm and continuous on $\mathcal{X}$. We claim p_U is in fact a norm. After all, suppose $y \in \mathcal{X}$ with $y \neq 0$. Then since $\mathcal{X}$ is T_1, we know there exists an open set $V \subseteq \mathcal{X}$ with $0 \in V$ and $y \notin V$. And in turn since U is von Neumann bounded, we know there exists $r > 0$ such that $sU \subseteq V$ for all $s \in K$ with $|s| \leq r$. This proves that $p_U(y) \geq r > 0$. So, $p_U(y) = 0$ iff $y = 0$.

Finally, we need to show that a net $\langle x_i \rangle_{i \in I}$ converges to x iff $p_U(x_i - x) \to 0$. Fortunately, the *proof* that $p_U(x_i - x) \to 0$ if $x_i \to x$ which we wrote when showing our sufficient condition for $\mathcal{X}$ being seminormable still works here. As for the other implication, note by the continuity of vector addition that $x_i \to x$ iff $x_i - x \to 0$. So, all we need to show is that if $p_U(x_i - x) \to 0$ in K, then $x_i - x$ is eventually in every neighborhood N of 0 in $\mathcal{X}$. To do that, note that since U is von Neumann bounded, there exists $r > 0$ such that $sU \subseteq N$ when $|s| \leq r$. And since $p_U(x_i - x) < r$ eventually, this proves $x_i - x \in N$ eventually. ■

Anyways, all of the topology I just did is beyond the scope of our class and textbook. I only toiled at proving everything prior because I wanted to know how to put rigor behind some comments my professor made.

For example, if A has infinite cardinality, then $\mathbb{C}^A$ equipped with the topology of pointwise convergence (which as a reminder is equivalent to the product topology) is not normable. This is because any nonempty open set in $\mathbb{C}^A$ is not von Neumann bounded.

Given a family of seminorms $\{p_\alpha\}_{\alpha \in A}$ on $\mathcal{X}$, we call that family sufficient if for all $x \in \mathcal{X}$ with $x \neq 0$ there exists $\alpha \in A$ with $p_\alpha \neq 0$. Recall from math 240b that the topology generated by a family of seminorms is Hausdorff iff that family is sufficient.

Remark 1.1.9: Topologies on a vector space $\mathcal{X}$ given by a finite sufficient family of seminorms are actually norm topologies. Specifically, let $\{p_1, \ldots, p_n\}$ be a sufficient family of seminorms on $\mathcal{X}$ and let $\|\cdot\|_q$ denote the q-norm on $\mathbb{R}^n$ (where $q \in [1, \infty]$). Then $\|\cdot\| \coloneqq \|(p_1(x), \ldots, p_n(x))\|_q$ is a norm on $\mathcal{X}$ yielding the same topology as that generated by our family of seminorms.

Proof:

Firstly we check $\|\cdot\|$ is a norm. Since some $p_i(x) > 0$ whenever $x \neq 0$, we know that $\|x\| = 0$ if and only if $x = 0$. Also it's obvious that $\|x\| \geq 0$ for all $x \in \mathcal{X}$. And if $c \in K$ and $x \in \mathcal{X}$, then:

$$\begin{aligned}\|cx\| &= \|(p_1(cx), \ldots, p_n(cx))\|_q \\ &= \|(cp_1(x), \ldots, cp_n(x))\|_q - |c| \cdot \|(p_1(x), \ldots, p_n(x))\|_q = |c| \cdot \|x\|.\end{aligned}$$

Finally, note that if $(c_1, \ldots, c_n), (d_1, \ldots, d_n) \in \mathbb{R}^n$ with $0 \leq c_i \leq d_i$ for all $1 \leq i \leq n$, then $\|c_1, \ldots, c_n\|_q \leq \|d_1, \ldots, d_n\|_q$.

Why?

If $q = \infty$ then this is obvious from the fact that $\max(c_1, \ldots, c_n) \leq \max(d_1, \ldots, d_n)$. Meanwhile, if $q < 1$, then since $t \mapsto t^{1/q}$ and $t \mapsto t^q$ are strictly increasing on the domain $[0, \infty)$, we know that $(d_1^q + \ldots + d_n^q)^{1/q} \geq (c_1^q + \ldots + c_n^q)^{1/q}$.

Thus if $x, y \in \mathcal{X}$, we have that:

$$\begin{aligned}\|x + y\| &= \|(p_1(x+y), \ldots, p_n(x+y))\|_q \\ &\leq \|(p_1(x) + p_1(y), \ldots, p_n(x) + p_n(y))\|_q \\ &= \|(p_1(x), \ldots, p_n(x))\|_q + \|(p_1(y), \ldots, p_n(y))\|_q = \|x\| + \|y\|\end{aligned}$$

This proves that $\|\cdot\|$ is a norm on $\mathcal{X}$.

Next note that if $q < \infty$ and $\|\cdot\|_\infty$ is the ∞-norm, then:

- $\|(x_1, \ldots, x_n)\|_q \leq (n(\max(|x_i|))^q)^{1/q} = n^{1/q}\|(x_1, \ldots, x_n)\|_\infty$,
- $\|(x_1, \ldots, x_n)\|_\infty = (\max_{1 \leq i \leq n} |x_i|^q)^{1/q} \leq (\sum_{i=1}^n |x_i|^q)^{1/q} = \|(x_1, \ldots, x_n)\|_q$.

So in general we know there exists $C \geq 1$ such that:

$$\max_{1 \leq i \leq n} p_i(x) \leq \|x\| \leq C \cdot \max_{1 \leq i \leq n} p_i(x).$$

And from there it is easy to show that a net converges with respect to the seminorms $p_1, \ldots, p_n$ iff it converges with respect to $\|\cdot\|$. ■

As a side note: technically our textbook says that you can replace the q-norm on $\mathbb{R}^n$ with any norm on $\mathbb{R}^n$ and the prior construction still works. However, for the last 14 hours I have not for the life of me been able to prove the triangle inequality for that more general construction. Thus, since I don't have time to be wasting on this, I'm just going to move on and swallow my losses.

There are a few more things I'd like to take notes on from this lecture. However, it's now 9/28/2025 and I need to make sure I have my notes done for my other math classes. So I'll come back to my functional analysis notes later on *page 251*.

9/28/2025

Math 200a (lecture 1):

What follows is not a logical definition but instead a grammatical one. Let X be a set endowed with some sort of mathematical structure. Then we say a symmetry or automorphism of X is a bijection $f : X \to X$ such that f and f^{-1} preserve the mathemtical structure of X. Also, we write $\mathrm{Aut}(X)$ to mean the set of symmetries of X. Importantly, $\mathrm{Aut}(X)$ will be a group.

Now while the last definition is very fuzzy, it is a suitable north star to follow in order to define the following actual concrete definitions.

- If X is just a set with no additional structure, then we denote the symmetric group of X to be $\mathrm{Aut}(X) = S_X :=$ {permutations of X}.

- If V is an F-vector space of dimension n, then we denote:
$$\mathrm{Aut}(V) = \{\text{invertible linear maps on } V\}.$$
Note that if $\mathrm{GL}_n(F)$ is the set of $n \times n$ invertible matrices with coefficients in F, then $\mathrm{GL}_n(F) \cong \mathrm{Aut}(V)$. We call $\mathrm{GL}_n(F)$ the $n \times n$ general linear group on F.

- Let $G = (V, E)$ be a graph. Then we say $\theta \in \mathrm{Aut}(G)$ if and only if $\theta : V \to V$ is a bijection such that $\{v_1, v_2\} \in E$ if and only if $\{\theta(v_1), \theta(v_2)\} \in E$.

- If G is a group, then we denote $\mathrm{Aut}(G)$ to be the set of bijective group homomorphisms from G to G.

Given a group G and a symmetric set $S \subseteq G$ (meaning that $s \in S \Longleftrightarrow s^{-1} \in S$), we define the Cayley graph $\mathrm{Cay}(G, S)$ of S and G to be the graph whose vertex set is G and for which given any distinct $g_1, g_2 \in G$ we have that $\{g_1, g_2\}$ is an edge if and only if $g_1 g_2^{-1} \in S$.

Note that we require S to be symmetric so that $g_1 g_2^{-1} \in S \Longleftrightarrow g_2 g_1^{-1} \in S$. If this weren't true than we'd need to let G be a digraph instead of a graph.

Also, if G is a group and $S \subseteq G$, we denote $\langle S \rangle$ to be the smallest subgroup of G containing S and say $\langle S \rangle$ is generated by S.

Note that $\langle S \rangle$ is precisely equal to the collection H of all finite compositions of elements of S and their inverses. After all, since each finite composition of elements in S has to be in $\langle S \rangle$, we trivially know that $H \subseteq \langle S \rangle$. On the other hand, it is clear that H is closed under composition and that every element of H has an inverse element. And finally, for any $h \in H$ we have that $hh^{-1} = 1 \in H$. So, H is a group containing S. It follows that $\langle S \rangle \subseteq H$.

Proposition: Let G be a group and $S \subseteq G$ be a symmetric subset. Then $\mathrm{Cay}(G, S)$ is connected if and only if $\langle S \rangle = G$.

$(\Longleftarrow)$

Suppose $\langle S \rangle = G$. Then for any $g \in G$ there exists $s_1, \ldots, s_n \in S$ with $g = s_n \cdots s_1$. And in turn, we can easily see that the following is a walk in $\mathrm{Cay}(G, S)$ from 1 to g:

$$1, \quad s_1, \quad s_2 s_1, \quad s_3 s_2 s_1, \quad \ldots, \quad (s_{n-1} \cdots s_2 s_1), \quad g$$

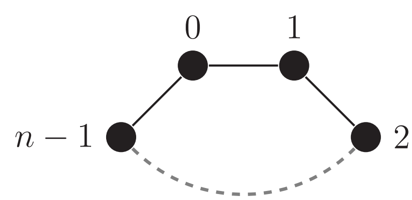
$(\Longrightarrow)$

Suppose $g \in G$ and there are $g_1, \ldots, g_n \in G$ such that $1 = g_1$, $g = g_n$, and $1, g_1, \ldots, g_n$ is a walk in $\mathrm{Cay}(G, S)$. Then $g_i(g_{i+1})^{-1} \in S$ for all $i \in \{1, \ldots, n-1\}$. Then:

$$g^{-1} = 1g^{-1} = g_1 g_n^{-1} = (g_1 g_2^{-1})(g_2 g_3^{-1}) \cdots (g_{n-1} g_n^{-1}) \in \langle S \rangle.$$

And in turn since $\langle S \rangle$ is a subgroup of G, we therefore have that $g \in \langle S \rangle$. ■

Note that an n-cycle is the Cayley graph of $\mathbb{Z}/n\mathbb{Z}$ and $\{-1, 1\}$. For if we label the vertices as shown below, we have that $\{i, j\} \in E$ if and only if $i - j = \pm 1$.



Thus it's clear that $\sigma : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ given by $\sigma(x) = x + 1$ and $\tau : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ given by $\tau(x) = -x$ are both symmetries of our graph, and that $\langle \sigma, \tau \rangle \subseteq \mathrm{Aut}(n\text{-cycle})$.

Can we show that $\langle \sigma, \tau \rangle = \mathrm{Aut}(n\text{-cycle})$? Sure.

Suppose $\theta \in \mathrm{Aut}(n\text{-cycle})$ and let $\theta(0) = i$. Then $\sigma^{-i}(x) = x - i$. Hence, we know that $\sigma^{-i} \circ \theta(0) = 0$. Next, since $\sigma^{-i} \circ \theta$ is a graph automorphism we know that $\sigma^{-i} \circ \theta(1) = \pm 1$. If $\sigma^{-i} \circ \theta(1) = -1$ then note that $\tau(0) = 0$ and $\tau(-1) = 1$. So, $\psi = \tau \circ \sigma^{-i}$ is an automorphism in $\langle \sigma, \tau \rangle$ such that $\psi \circ \theta$ fixes 0 and 1. Meanwhile, if $\sigma^{-i} \circ \theta(1) = +1$ then we can just set $\psi = \sigma^{-i}$ to get an automorphism in $\langle \sigma, \tau \rangle$ such that $\psi \circ \theta$ fixes 0 and 1.

Now it is an easy induction argument to show that $\psi \circ \theta$ fixes all of $\mathbb{Z}/n\mathbb{Z}$. Consider any $k \in \{2, \ldots, n-1\}$ and suppose that we already showed that $\psi \circ \theta(j) = j$ for all $j \in \{0, \ldots, k-1\}$. Then since $\psi \circ \theta$ is an automorphism and $\psi \circ \theta(k-1) = k-1$ we know $\psi \circ \theta(k) \in \{k-2, k\}$. However, we can't have that $\psi \circ \theta(k) = k-2$ since we already have that $\psi \circ \theta(k-2) = k-2$. Hence, we must have that $\psi \circ \theta(k) = k$.

It follows that $\psi \circ \theta = \mathrm{Id}$. Or in other words, $\theta = \psi^{-1} \in \langle \sigma, \tau \rangle$.

Note that in the prior reasoning we also proved that every $\theta \in \mathrm{Aut}(n\text{-cycle})$ can be expressed as σ^i or $\sigma^i\tau$ where $i \in \{0, \ldots, n-1\}$. Is this expression unique? Sure.

Consider the subgroup $H = \langle \sigma \rangle$ of $\mathrm{Aut}(n\text{-cycle})$. Thus $H = \{\sigma^i : i \in \mathbb{Z}\}$. And by examining that $\sigma^i(0) = i$ for all $i \in \mathbb{Z}$, we can see that $\sigma^0, \sigma^1, \ldots, \sigma^{n-1}$ are all distinct. Also, since σ^n fixes 0 and 1, we know like before that $\sigma^n = \mathrm{Id}$. Hence $H = \{\mathrm{Id}, \sigma, \sigma^2, \ldots, \sigma^{n-1}\}$.

Now importantly, $\tau \notin H$. After all, τ fixes 0 and sends 1 to -1. But the only element in H which fixes 0 is Id and that sends 1 to $+1$. Hence, we know H and $H\tau$ are disjoint cosets. Also, since all cosets of a finite group have the same order, we know that each $\tau\sigma^i$ (where $i \in \{0, \ldots, n-1\}$) is also distinct.

Given a group G, we write $o(G)$ to denote the order of G. Notably, we have now showed that $o(\mathrm{Aut}(n\text{-cycle})) = 2n$ and that:

$$\mathrm{Aut}(n\text{-cycle}) = \{\sigma^i : 0 \leq i \leq n-1\} \cup \{\sigma^i\tau : 0 \leq i \leq n-1\}.$$

We call this group the nth. dihedral group and will refer to it as D_{2n} from now on.

Note that $\tau \circ \sigma \circ \tau(x) = \tau \circ \sigma(-x) = \tau(-x+1) = x-1 = \sigma^{-1}(x)$. Hence $\tau \circ \sigma \circ \tau = \sigma^{-1}$. Also, we clearly have that $\tau^2 = \mathrm{Id}$ and that $\sigma^n = \mathrm{Id}$. We call these three identities the defining relations of the dihedral group.

We that H is a normal subgroup of G by writing $H \lhd G$. In particular, note that $\langle \sigma \rangle \lhd D_{2n}$. This is because conjugation is a group automorphism and $\tau = \tau^{-1}$, which in turn shows that $\tau \circ \sigma^j \circ \tau = (\tau \circ \sigma \circ \tau)^j = \sigma^{-j} \in \langle \sigma \rangle$ for all $j \in \mathbb{Z}$.

I'll continue with this class on *page 244*.

Math 220a (lecture 1):

In real analysis, sometimes we wanted to compactify $\mathbb{R}$ so that we could describe the limiting behavior of unbounded sequences and functions. And the common way we did that was by defining the extended real number line $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$. In complex analysis, we want to be able to do something similar. However, $\mathbb{C}$ doesn't just have two directions for an unbounded sequence to diverge in. So, we will need to modify our strategy when defining the extended complex plane / Riemann sphere.

Let $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$ denote the one-point compactification of $\mathbb{C}$. We can metrize $\mathbb{C}_\infty$ as follows.

Consider the spherical shell $S^2 = \{(x, y, z) \in \mathbb{R}^3 : \sqrt{x^2 + y^2 + z^2} = 1\} \subseteq \mathbb{R}^3$ and identify $\mathbb{C}$ with the xy-plane in the obvious way.

Now given any $z = x + iy = (x, y, 0) \in \mathbb{R}^3$, let ℓ be the line passing through z and the north pole $(0, 0, 1)$ of S^2. We claim that there is a unique other point on S^2 which ℓ crosses.

Proof:
We may parametrize ℓ by $\ell(t) = t(0,0,1) + (1-t)(x,y,0)$. Then ℓ intersects S^2 precisely when $(1-t)^2x^2 + (1-t)^2y^2 + t^2 = 1$. Or in other words, we need t to satisfy that $(1-t)^2|z|^2 + t^2 = 1$.

This then rearranges to the quadratic equation $(1+|z|^2)t^2 - 2|z|^2t + (|z|^2-1) = 0$. And by applying the quadratic formula we see that $\ell(t)$ intercepts S^2 when:

$$t = \frac{2|z|^2 \pm \sqrt{4|z|^4 - 4(1+|z|^2)(|z|^2-1)}}{2(1+|z|^2)} = \frac{|z|^2 \pm \sqrt{|z|^4 - (|z|^4-1)}}{1+|z|^2} = \frac{|z|^2 \pm 1}{|z|^2+1}$$

This shows that ℓ crosses S^2 exactly twice. Now when $t = 1$, we just have that $\ell(t) = (0,0,1)$. So, the fabled other point on S^2 which ℓ crosses and which we actually care about is:

$$\begin{aligned}\ell\left(\tfrac{|z|^2-1}{|z|^2+1}\right) &= \left((1 - \tfrac{|z|^2-1}{|z|^2+1})^2 x,\ \ (1 - \tfrac{|z|^2-1}{|z|^2+1})^2 y,\ \ \tfrac{|z|^2-1}{|z|^2+1}\right) \\ &= \left(\tfrac{|z|^2+1-(|z|^2-1)}{|z|^2+1} x,\ \ \tfrac{|z|^2+1-(|z|^2-1)}{|z|^2+1} y,\ \ \tfrac{|z|^2-1}{|z|^2+1}\right) = \left(\tfrac{2x}{|z|^2+1},\ \ \tfrac{2y}{|z|^2+1},\ \ \tfrac{|z|^2-1}{|z|^2+1}\right)\end{aligned}$$

Also, if $(c_1, c_2, c_3) \in S^2$, with $(c_1, c_2, c_3) \neq (0,0,1)$, we can go in reverse and find a unique point $z = x + iy = (x, y, 0)$ in the xy-plane which the line ℓ passing through $(0,0,1)$ and (c_1, c_2, c_3) intercepts.

Proof:
Parametrize ℓ by $\ell(t) = t(0,0,1) + (1-t)(c_1, c_2, c_3)$. Then ℓ intercepts the xy-plane precisely when $t + (1-t)c_3 = 0$. Or in other words, when $t = \frac{-c_3}{1-c_3}$.

In turn, the point at which ℓ intercepts the xy-plane is:

$$((1 + \tfrac{c_3}{1-c_3})c_1,\ \ (1+\tfrac{c_3}{1-c_3})c_2,\ \ 0) = (1+\tfrac{c_3}{1-c_3})c_1 + i(1+\tfrac{c_3}{1-c_3})c_2 = \tfrac{c_1+ic_2}{1-c_3}.$$

Hence, we have shown that there is a bijective correspondance between $\mathbb{C}$ and $S^2 - \{0,0,1\}$. And now the big observation to make is that for any sequence in $\mathbb{C}$ converging to ∞ in $\mathbb{C}_\infty$, we have that the associated sequence in $S^2 - \{(0,0,1)\}$ converges to $(0,0,1)$.

Why?
Convergence of a sequence $(z_n)_{n\in\mathbb{N}}$ to ∞ in $\mathbb{C}_\infty$ is equivalent to saying that z_n is eventually not in every compact set $K \subseteq \mathbb{C}$. And by the Heine-Borel theorem, this is equivalent to saying that $|z_n| \to \infty$ (in $\mathbb{R}$) as $n \to \infty$.

But now note that we have as $|z_n| \to \infty$ that:

- $\frac{|z_n|^2-1}{|z_n|^2+1} \to 1$,
- $\frac{2|z_n|}{|z_n|^2+1} \to \lim\limits_{t\to\infty} \frac{t}{t^2+1} = \lim\limits_{t\to\infty} \frac{1}{2t} = 0$

And since $|\mathrm{Re}(z)| \leq |z|$ and $|\mathrm{Im}(z)| \leq |z|$, we thus know that:

$$\left(\tfrac{2\mathrm{Re}(z_n)}{|z_n|^2+1},\ \ \tfrac{2\mathrm{Im}(z_n)}{|z_n|^2+1},\ \ \tfrac{|z_n|^2-1}{|z_n|^2+1}\right) \to (0,0,1) \text{ as } n \to \infty.$$

So by mapping ∞ to $(0,0,1)$, we can now say we've constructed a bijection from $\mathbb{C}_\infty$ to S^2. To finish constructing our metric on $\mathbb{C}_\infty$, just define $d(z_1, z_2)$ to be the Euclidean metric evaluated between z_1 and z_2's corresponding points on S^2.

Note that if $\mathbf{x}$ and $\mathbf{y}$ are two points on S^2, then $\|\mathbf{x}-\mathbf{y}\|_2 = (2-2(\mathbf{x}\cdot\mathbf{y}))^{1/2}$. This is because $\|\mathbf{x}-\mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 - 2(\mathbf{x}\cdot\mathbf{y}) + \|\mathbf{y}\|_2^2$. And since $\mathbf{x},\mathbf{y}\in S^2$, we know that $\|\mathbf{x}\|_2^2 = \|\mathbf{y}\|_2^2 = 1$.

This now leads us to the following formulas for $d(z_1, z_2)$ where $z_1, z_2 \in \mathbb{C}_\infty$.

- Suppose $z_1 = x_1 + iy_1 \neq \infty$ and $z_2 = x_2 + iy_2 \neq \infty$. Then:

$$\begin{aligned}
d(z_1,z_2) &= \left(2 - \tfrac{2(4x_1x_2+4y_1y_2+(|z_1|^2-1)(|z_2|^2-1))}{(|z_1|^2+1)(|z_2|^2+1)}\right)^{1/2} \\
&= \left(\tfrac{2((|z_1|^2+1)(|z_2|^2+1)-4x_1x_2-4y_1y_2-(|z_1|^2-1)(|z_2|^2-1))}{(|z_1|^2+1)(|z_2|^2+1)}\right)^{1/2} \\
&= \left(\tfrac{2(|z_1|^2|z_2|^2+|z_1|^2+|z_2|^2+1-4x_1x_2-4y_1y_2-|z_1|^2|z_2|^2+|z_1|^2+|z_2|^2-1)}{(|z_1|^2+1)(|z_2|^2+1)}\right)^{1/2} \\
&= \left(\tfrac{2(2|z_1|^2+2|z_2|^2-4x_1x_2-4y_1y_2)}{(|z_1|^2+1)(|z_2|^2+1)}\right)^{1/2} \\
&= \tfrac{2}{((|z_1|^2+1)(|z_2|^2+1))^{1/2}} \cdot \left(\|(x_1,y_1)\|_2^2 - 2((x_1,y_1)\cdot(x_2,y_2)) + \|(x_2,y_2)\|_2^2\right)^{1/2} \\
&= \tfrac{2}{((|z_1|^2+1)(|z_2|^2+1))^{1/2}} \cdot \left(\|(x_1,y_1)-(x_2,y_2)\|_2^2\right)^{1/2} = \tfrac{2|z_1-z_2|}{((|z_1|^2+1)(|z_2|^2+1))^{1/2}}.
\end{aligned}$$

- Meanwhile, suppose $z_1 \neq \infty$. Then:

$$d(z_1,\infty) = \left(2 - 2(0+0+1\tfrac{|z_1|^2-1}{|z_1|^2+1})\right)^{1/2} = \left(\tfrac{2|z_1|^2+2-2|z_1|^2+2}{|z_1|^2+1}\right)^{1/2} = \tfrac{2}{(|z_1|^2+1)^{1/2}}.$$

- And obviously, $d(\infty,\infty) = 0$.

Due to how we constructed d, it's trivial that d is a metric.

In fact, even more generally suppose X is a set, (Y,ρ) is a metric space, and $f : X \to Y$ is an injection. Then $d(x_1,x_2) = \rho(f(x_1), f(x_2))$ is a metric on X.

Why?

We know $d(x_1,x_2) = 0$ if and only if $x_1 = x_2$ because $f(x_1) = f(x_2)$ if and only if $x_1 = x_2$. Plus, it's clear since ρ is symmetric and nonnegative that $d(x_1,x_2) = d(x_2,x_1)$ and $d(x_1,x_2) \geq 0$. Finally, the triangle inequality holds since:

$$\begin{aligned}
d(x_1,x_2) &= \rho(f(x_1), f(x_2)) \\
&\leq \rho(f(x_1), f(x_3)) + \rho(f(x_3), f(x_2)) = d(x_1,x_3) + d(x_3,x_2).
\end{aligned}$$

(Since the grader for **Exercise II.1.7** didn't see the last page: another way you can see that the stereographic projection is injective is that we can literally calculate using our prior formulas that $d(z_1,z_2) \neq 0$ unless $z_1 = z_2$.)

What's less trivial is that d generates the same topology as that which is already on $\mathbb{C}_\infty$.

Let $\mathcal{T}$ equal the typical one-point compactification topology on $\mathbb{C}_\infty$ and $\mathcal{T}'$ equal the metric topology on $\mathbb{C}_\infty$ generated by d.

- Firstly, fix any $z \in \mathbb{C}$. Then note that $d(z,z') \leq 2|z-z'|$ for all $z' \in \mathbb{C}$. Also $d(z,\infty) = \frac{2}{(|z|^2+1)^{1/2}} > 0$. Thus if $U \in \mathcal{T}'$ is a d-ball of radius ε, there exists a (Euclidean) open ball $V \in \mathcal{T}$ of radius less than $\min(\frac{2}{2\cdot(|z|^2+1)^{1/2}}, \varepsilon/2)$ with $x \in V \subseteq U \cap \mathbb{C} \subseteq U$.

- Meanwhile, suppose $V \in \mathcal{T}$ is an open set containing z. Then by restricting V to $\mathbb{C}$ we can assume $V \subseteq \mathbb{C}$. So, there is a (Euclidean) ball $B \subseteq \mathbb{C}$ of radius $\varepsilon > 0$ centered at z which is contained in V.

  Next note that for any $z' \in \mathbb{C}$
  $$\begin{aligned}|z - z'| &= d(z, z')\left((|z|^2+1)(|z'|^2+1)\right)^{1/2} \\ &= d(z,z')\sqrt{|z|^2+1}\left((|z'|^2+1)^{1/2} - (|z|^2+1)^{1/2} + (|z|^2+1)^{1/2}\right) \\ &\leq d(z,z')\sqrt{|z|^2+1}\cdot\left|(|z'|^2+1)^{1/2} - (|z|^2+1)^{1/2}\right| + d(z,z')(|z|^2+1)\end{aligned}$$

  Now if $f(t) = (t^2+1)^{1/2}$ then $f'(t) = \frac{t}{\sqrt{t^2+1}}$. So $|f'(t)| \leq 1$ and we can conclude by the mean value theorem that:
  $$\left|(|z'|^2+1)^{1/2} - (|z|^2+1)^{1/2}\right| \leq 1 \cdot \left||z'| - |z|\right| \leq |z - z'|$$

  Hence $|z - z'| \leq d(z,z')|z-z'|\sqrt{|z|^2+1} + d(z,z')(|z|^2+1)$. Or equivalently, we have that:
  $$(1 - d(z,z')\sqrt{|z|^2+1})|z-z'| \leq d(z,z')(|z|^2+1).$$

  Now importantly there exists $\delta_1 > 0$ such that if $d(z,z') < \delta_1$ is small enough, then since $|z|$ is fixed we will always have that $(1 - d(z,z')\sqrt{|z|^2+1}) > 0$. Hence, we can say that $|z - z'| \leq \frac{d(z,z')(|z|^2+1)}{1-d(z,z')\sqrt{|z|^2+1}}$. And finally, since $\frac{t}{1-t} \to 0$ as $t \to 0$, we can pick $\delta_2 \in (0, \delta_1)$ such that $|z - z'| < \varepsilon$ when $d(z,z') < \delta_2$. And by shrinking δ_2 enough we can also assume that $d(z, \infty) > \delta_2$.

  Thus, we have found an open ball $U \in \mathcal{T}'$ with d-radius δ_2 about z such that $z \in U \subseteq B \subseteq V$.

What's left to show is firstly that for any $U \in \mathcal{T}$ containing ∞ there exists $V \in \mathcal{T}'$ with $\infty \in V \subseteq U$, and secondly that for any $V \in \mathcal{T}'$ containing ∞ there exists $U \in \mathcal{T}$ with $\infty \in U \subseteq V$.

- Pick any $U \in \mathcal{T}$ with $0 \in U$. Then there exists some $N > 0$ such that $z \in U$ for all $z \in \mathbb{C}$ with $|z| > N$. In turn, if $d(\infty, 2) < \frac{2}{1+N}$ then we know that:
  $$N + 1 < \frac{2}{d(\infty,z)} = \sqrt{|z|^2+1} \leq |z| + 1$$

  Hence, whenever $d(\infty, z) < \frac{2}{1+N}$ we have that $|z| > N$ and hence $z \in U$. So, we've shown that there exists $V \in \mathcal{T}'$ with $\infty \in V \subseteq U$.

- Meanwhile, pick any $V \in \mathcal{T}'$ with $\infty \in V$. Then there exists $\varepsilon > 0$ such that $z \in V$ for all $z \in \mathbb{C}$ with $d(\infty, z) < \varepsilon$. And if $|z| > {}^2/_{\varepsilon}$, then we know that:
  $$d(\infty, z) = \frac{2}{(|z|^2+1)^{1/2}} < \frac{2}{|z|} < \varepsilon.$$

  So, $U = \{z \in \mathbb{C} : |z| > {}^2/_{\varepsilon}\} \cup \{\infty\}$ is an open set in $\mathcal{T}$ with $\infty \in U \subseteq V$.

With that we've *finally* proven that $\mathcal{T} = \mathcal{T}'$. ■

As a side note, it's now clear that our bijection from $\mathbb{C}_\infty$ to S^2 is continuous. And since $\mathbb{C}_\infty$ is compact and S^2 is Hausdorff, we also know that our bijection is a homeomorphism. Hence $\mathbb{C}_\infty$ is homeomorphic to S^2.

9/29/2025

All homework problem's for math 220a will be coming from John Conway's book *functions of one complex variable* (See *item 13* in the bibliography).

Exercise I.6.4: Let Λ be a circle lying in S^2. Then (by definition of a circle) there is a unique plane P in $\mathbb{R}^3$ such that $P \cap S^2 = \Lambda$. So, take $P = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1\beta_1 + x_2\beta_2 + x_3\beta_3 = \ell\}$ where $(\beta_1, \beta_2, \beta_3)$ is a unit vector orthogonal to P and $\ell \in \mathbb{R}$. Then show that if Λ contains $(0,0,1)$, it's projection to $\mathbb{C}_\infty$ is a straight line, and meanwhile if Λ doesn't contain $(0,0,1)$ then it's projection to $\mathbb{C}_\infty$ is another circle.

To start off, any point $z = x + iy \in \mathbb{C}$ corresponds to a point in Λ iff:

$$\frac{2x}{|z|^2+1}\beta_1 + \frac{2y}{|z|^2+1}\beta_2 + \frac{|z|^2-1}{|z|^2+1}\beta^3 = \ell$$

After some rearranging this becomes $2\beta_1 x + 2\beta_2 y + (x^2+y^2-1)\beta_3 = \ell(x^2+y^2+1)$. Or in other words $(\beta_3 - \ell)x^2 + 2\beta_1 x + (\beta_3 - \ell)y^2 + 2\beta_2 y = \ell + \beta_3$. And now we break off into two cases.

1. Suppose $\beta_3 = \ell$. Then we know that $0\beta_1 + 0\beta_2 + 1\beta_3 = \ell$. So $(0,0,1) \in \Lambda$. At the same time, all the square terms in our condition cancel and we are left with the requirement $2\beta_1 x + 2\beta_2 y = \ell + \beta_3$. This is the equation of a line. Hence, we've shown that the projection of Λ onto $\mathbb{C}_\infty$ is a straight line unioned with the point at infinity.

2. Suppose $\beta_3 \neq \ell$. Then we know that $0\beta_1 + 0\beta_2 + 1\beta_3 \neq \ell$ and hence $(0,0,1) \notin \Lambda$. At the same time, we can now divide our equation from before by $\beta - \ell$ in order to get that $x^2 + 2\frac{\beta_1}{\beta_3-\ell}x + y^2 + 2\frac{\beta_2}{\beta_3-\ell}y = \frac{\ell+\beta_3}{\beta_3-\ell}$. And by completing the square we have:

$$(x + \tfrac{\beta_1}{\beta_3-\ell})^2 + (y + \tfrac{\beta_2}{\beta_3-\ell})^2 = \frac{\beta_3+\ell}{\beta_3-\ell} + \frac{\beta_1^2+\beta_2^2}{(\beta_3-\ell)^2} = \frac{1-\ell^2}{(\beta_3-\ell)^2}.$$

So, the projection of Λ onto $\mathbb{C}_\infty$ is a circle of radius $\frac{\sqrt{1-\ell^2}}{\beta_3-\ell}$ centered at $\frac{-\beta_1}{\beta_3-\ell} + i\frac{-\beta_2}{\beta_3-\ell}$.

As a side note, let $\mathbf{b} = (b_1, b_2, b_3)$ and consider any $\mathbf{y} = (y_1, y_2, y_3) \in \Lambda$. Then note by the Cauchy Schwartz inequality that $\ell^2 = (\mathbf{b}\cdot\mathbf{y})^2 \leq \|\mathbf{b}\|^2\|\mathbf{y}\|^2 = 1 \cdot 1 = 1$. Hence, $\sqrt{1-\ell^2}$ is a well defined real number.

I'll continue with this class on *page 247*.

Math 200a (lecture 2):

Given a group G and $a \in G$, we denote the order of a as $o(a) := o(\langle a \rangle)$ Now let C_n be a cyclic group of order n. In other words, $C_n = \langle a \rangle$ where $o(a) = n$. What is $\mathrm{Aut}(C_n)$?

Set 1 problem 5:

(a) Suppose $\theta \in \mathrm{Aut}(C_n)$ and pick any $m \in \mathbb{Z}$ with $\theta(a) = a^m$. Then $\gcd(m,n) = 1$.

Since θ is a bijection, we know that $\theta(a^k)$ for each $k \in \{0, \ldots, n-1\}$ is distinct. But also since θ is a group homomorphism, we know that $\theta(a^0) = a^0 = 1$ and $\theta(a^k) = (\theta(a))^k = (a^m)^k$ for $k > 0$. Hence, we must have that $o(a^m) = n$. But now recall also that $o(a^m) = \frac{o(a)}{\gcd(o(a),m)} = \frac{n}{\gcd(n,m)}$. So, we must have that $\gcd(n,m) = 1$.

(b) Suppose $m \in \mathbb{Z}$ satisfies that $\gcd(m,n) = 1$ and define $\theta_m(a) : C_n \to C_n$ by $\theta_m(a^k) = (a^k)^m$. Then $\theta_m \in \operatorname{Aut}(C_n)$.

To see that θ_m is a group homomorphism, just note that:

$$\begin{aligned}\theta_m(a^{k_1}a^{k_2}) &= \theta_m(a^{k_1+k_2}) \\ &= (a^{k_1+k_2})^m = a^{m(k_1+k_2)} \\ &= a^{mk_1+mk_2} = (a^{k_1})^m(a^{k_2})^m = \theta_m(a^{k_1})\theta_m(a^{k_2})\end{aligned}$$

Meanwhile note that $o(a^m) = \frac{o(a)}{\gcd(o(a),m)} = \frac{n}{1} = n$.

Therefore, because $\theta_m(a^k) = (\theta_m(a))^k = (a^m)^k$ since θ_m is a group homomorphism, we know by pigeonhole principle that θ_m is both injective and surjective.

(c) The mapping $(\mathbb{Z}/n\mathbb{Z})^\times \to \operatorname{Aut}(C_n)$ given by $m + n\mathbb{Z} \mapsto \theta_m$ is an isomorphism.

Suppose $m + n\mathbb{Z} \in (\mathbb{Z}/n\mathbb{Z})^\times$ and consider any $m' \in m + n\mathbb{Z}$. Then $m' = m + qn$. So for any $a^k \in C_n$, we have that:

$$\begin{aligned}\theta_m(a^k) &= (a^k)^m \\ &= (a^k)^{m'-qn} = (a^k)^{m'}(a^k)^{-qn} \\ &= (\theta_{m'}(a^k))(a^n)^{-qk} = (\theta_{m'}(a^k))1^{-qk} = \theta_{m'}(a^k).\end{aligned}$$

This proves that $m + n\mathbb{Z} \mapsto \theta_m$ is a well defined function on $(\mathbb{Z}/n\mathbb{Z})^\times$. Also, since $\gcd(m,n) = 1$ for all $m \in \mathbb{Z}$ with $m + n\mathbb{Z} \in (\mathbb{Z}/n\mathbb{Z})^\times$, we know from part (b) that each θ_m is in fact in $\operatorname{Aut}(C_n)$.

Finally, we need to show our mapping is a group isomorphism. Luckily, by part (a) we know each $\theta \in \operatorname{Aut}(C_n)$ satisfies that $\theta(a^k) = (\theta(a))^k = (a^m)^k = (a^k)^m$ where $\gcd(m,n) = 1$. And since $\gcd(m,n) = 1$ iff $m + n\mathbb{Z} \in (\mathbb{Z}/n\mathbb{Z})^\times$, this proves that our mapping into $\operatorname{Aut}(C_n)$ is a surjective map. As for showing injectivity, suppose $\theta_{m_1} = \theta_{m_2}$. Then $\theta_{m_1}(a) = a^{m_1} = a^{m_2} = \theta_{m_2}(a)$. But in turn we have that $a^{m_1 - m_2} = 1$. So $m_1 \equiv m_2 \pmod{n}$. Lastly, suppose that suppose that m_1, m_2 represent classes in $(\mathbb{Z}/n\mathbb{Z})^\times$. Then:

$$\theta_{m_1} \circ \theta_{m_2}(a) = \theta_{m_1}(a^{m_2}) = (\theta_{m_1}(a))^{m_2} = (a^{m_1})^{m_2} = a^{m_1 m_2} = \theta_{m_1 m_2}(a)$$

In turn $\theta_{m_1} \circ \theta_{m_2}(a^k) = (\theta_{m_1} \circ \theta_{m_2}(a))^k = (\theta_{m_1 m_2}(a))^k = \theta_{m_1 m_2}(a^k)$. And so we know our mapping is a group homomorphism since $\theta_{m_1} \circ \theta_{m_2} = \theta_{m_1 m_2}$. ■

Question: Given a group G and a set X, how can we treat G as a group of symmetries of X?

To answer this, let us suppose we have a function $G \times X \to X$ where we write that $(g, x) \mapsto g \cdot x$ (that way whenever we fix g we get a function from X to X).

If $x \mapsto g \cdot x$ is an automorphism for each g and G is homomorphic to $\operatorname{Aut}(X)$, then clearly we must have that $1_G \cdot x = x$ for all $x \in X$ and that $g_2 \cdot (g_1 \cdot x) = (g_2 g_1) \cdot x$ for all $g_1, g_2 \in G$ and $x \in X$.

In other words, we may describe group homomorphisms of G to $\operatorname{Aut}(X)$ using (left) group actions.

Side note: given a group action $G \times X \to X$, we say G acts on X and write $G \curvearrowright X$.

Here are some notable examples of group actions.

1. Let G be a group and $H \subseteq G$ be a subgroup. Also denote G/H as the set of left cosets of H. Then $G \curvearrowright G/H$ by left translation (i.e. $g \cdot xH := (gx)H$ is a group action).

2. If G is a group, $G \curvearrowright G$ via conjugation (i.e. where we define $g \cdot x := gxg^{-1}$).

   As a side note, we don't write $g^{-1}xg$ because that would make it so that:
   $$g_1 \cdot (g_2 \cdot x) = g_1^{-1}g_2^{-2}xg_2g_1 = (g_2g_1)^{-1}x(g_2g_1) = (g_2g_1) \cdot x.$$

   That said, if we were instead formulating everything by right group actions, then we would want to define $x \cdot g = g^{-1}xg$. After all, we'd then have that:
   $$(x \cdot g_1) \cdot g_2 = g_2^{-1}g_1^{-1}xg_1g_2 = (g_1g_2)^{-1}xg_1g_2 = x \cdot (g_1g_2).$$

   In general there is no reason we need to work with left actions as opposed to right actions. In fact, if $(g, x) \mapsto g \cdot x$ is a left action, then $(x, g) \mapsto g^{-1} \cdot x$ is a right action. Similarly, if $(x, g) \mapsto x \cdot g$ is a right action, then $(g, x) \mapsto x \cdot g^{-1}$ is a left action. For whatever reason it just happens to be more standard in english speaking countries to use left actions as opposed to right actions.

3. If $G \curvearrowright X$, then we also have that $G \curvearrowright Y^X$ (where $\mathrm{Fun}(X, Y) := Y^X$).

   How?
   For any $f \in \mathrm{Fun}(X, Y)$ define $(g \cdot f)(x) = f(g^{-1} \cdot x)$ for each $x \in X$. Then it's clear that for any $x \in X$ we have that:
   - $(1_G \cdot f)(x) = f(1_G \cdot x) = f(x)$,
   - $$\begin{aligned}(g_1 \cdot (g_2 \cdot f))(x) &= (g_2 \cdot f)(g_1^{-1} \cdot x) \\ &= f(g_2^{-1} \cdot (g_1^{-1} \cdot x)) = f((g_2^{-1}g_1^{-1}) \cdot x) \\ &= f((g_1g_2)^{-1} \cdot x) = ((g_1g_2) \cdot f)(x)\end{aligned}$$

   Hence $1_G \cdot f = f$ and $g_1 \cdot (g_2 \cdot f) = (g_1g_2) \cdot f$.

   A big reason why this is relevant is that if $Y = \mathbb{C}$ then $\mathrm{Fun}(X, Y)$ becomes a complex vector space.

Theorem: Let $\mathrm{act}(G, X)$ be the collection of group actions $a : G \times X \to X$. Also let $\mathrm{Hom}(G, S_X)$ be the set of homomorphisms from G to S_X. There there exists mappings Φ and Ψ with:

$$\begin{tikzcd} \mathrm{act}(G, X) \arrow[r, bend left, "\Phi"] & \mathrm{Hom}(G, S_x) \arrow[l, bend left, "\Psi = \Phi^{-1}"] \end{tikzcd}$$

Proof:
Firstly, given any action $(g, x) \mapsto g \cdot x$ in $\mathrm{act}(G, X)$, define:
$$\Phi((g, x) \mapsto g \cdot x) = (g \mapsto \sigma_g) \text{ where } \sigma_g(x) := g \cdot x.$$

We claim that $g \mapsto \sigma_g$ is a group homomorphism from G to S_X.

To start off, $\sigma_{1_G} = \mathrm{Id}$ since $1_G \cdot x = x$. Also note that for all $x \in X$:
$$\sigma_{g_1} \circ \sigma_{g_2}(x) = \sigma_{g_1}(g_2 \cdot x) = g_1 \cdot (g_2 \cdot x) = (g_1g_2) \cdot x = \sigma_{g_1g_2}(x).$$

This shows that the map $g \mapsto \sigma_g$ preserves compositions. In particular, note that $\sigma_g \circ \sigma_{g^{-1}} = \sigma_{1_G} = \mathrm{Id}$ and $\sigma_{g^{-1}} \circ \sigma_g = \mathrm{Id}$. Thus each σ_g has an inverse for every $g \in G$ and is therefore a bijection. So, we have proven that $g \mapsto \sigma_g$ does in fact map into S_X.

Meanwhile, given any $f \in \mathrm{Hom}(G, S_X)$ define $\Psi(f) = ((g,x) \mapsto g \cdot x)$ where $g \cdot x := (f(g))(x)$. We can easily see as follows that this is a well defined action.

- $1_G \cdot x = (f(1_G))(x) = \mathrm{Id}_X(x) = x$
- $$\begin{aligned} g_1 \cdot (g_2 \cdot x) &= g_1 \cdot (f(g_2))(x) \\ &= (f(g_1))((f(g_2))(x)) = (f(g_1) \circ f(g_2))(x) \\ &= (f(g_1 g_2))(x) = (g_1 g_2) \cdot x. \end{aligned}$$

Now we just need to show that Φ and Ψ invert each other.

- Let $(g,x) \mapsto g \cdot x$ be a group action. Then for any $g' \in G$ and $x' \in X$ we have that:
$$\begin{aligned} ((\Psi \circ \Phi)((g,x) \mapsto g \cdot x))\,(g', x') &= ((\Phi((g,x) \mapsto g \cdot x))(g'))\,(x') \\ &= ((g \mapsto (x \mapsto (g \cdot x)))(g'))\,(x') \\ &= (x \mapsto (g' \cdot x))\,(x') = g' \cdot x'. \end{aligned}$$

  Hence $\Psi \circ \Phi((g,x) \mapsto g \cdot x) = (g,x) \mapsto g \cdot x$.

- Meanwhile, suppose $f \in \mathrm{Hom}(G, S_X)$. Then note that:
$$\begin{aligned} (\Phi \circ \Psi)(f) = \Phi((g,x) \mapsto (f(g))(x)) &= [g \mapsto (x \mapsto (f(g))(x))] \\ &= [g \mapsto f(g)] = f. \blacksquare \end{aligned}$$

I'll continue with this class on *page 252*.

Math 220a (lecture 2):

There is only one theorem from this lecture that is worth writing down in my opinion. Also, the professor and Conway insist on proving this in an annoying way (imo). So, I'm going to prove it my way first.

Btw theorem numbers for this class come from Conway's *functions of one complex variable* (like in the homework).

Given any $z, w \in \mathbb{C}$, we denote the straight line from z to w as:
$$[z,w] := \{(1-t)z + tw : t \in [0,1]\}.$$

If $z_1, \ldots, z_n \in \mathbb{C}$, then the path resulting from concatenating the $[z_i, z_{i+1}]$ together will be called a polygon in this class and sometimes be denoted as $[z_1, z_2, \ldots, z_n]$.

Theorem II.2.3: Let $G \subseteq \mathbb{C}$ be an open set. Then G is connected iff for alll $a, b \in G$ there exists $a = z_1, z_2, \ldots, z_{n-1}, z_n = b$ such that $[z_1, z_2, \ldots, z_n] \subseteq G$.

$(\Longrightarrow)$

$\mathbb{C}$ is locally path connected since each open ball in $\mathbb{C}$ is convex (and thus path connected). It follows that if G is connected then G is path connected. And so, there exists a continuous function $f : [0,1] \to \mathbb{C}$ such that $f(0) = a$ and $f(1) = b$.

Now for each $t \in [0,1]$ consider that there is some $\varepsilon_t > 0$ such that the open ball $B_{\varepsilon_t}(f(t))$ of radius ε centered at $f(t) \in \mathbb{C}$ is contained in G. Then since f is continuous, there exists an open ball $B_{\delta_t}(t) \subseteq [0,1]$ such that $f(B_{\delta_t}(t)) \subseteq B_{\varepsilon_t}(f(t))$.

Importantly, the $B_{\delta_t}(t)$ as t ranges over $[0,1]$ forms an open cover of $[0,1]$. And since $[0,1]$ is compact, we can pick $0 = t_1 < t_3 < \cdots < t_{2m-1} = 1$ such that $\{B_{\delta_{t_{2i-1}}}(t_{2i-1})\}_{i=1}^m$ covers $[0,1]$ and each $B_{\delta_{t_{2i-1}}}(t_{2i-1})$ is mapped into an open ball contained in G. Also, without loss of generality we can assume no $B_{\delta_{t_{2i-1}}}(t_{2i-1})$ is entirely contained in any $B_{\delta_{t_{2j-1}}}(t_{2j-1})$ where $i \neq j$

Finally, it's easy to see for each i that $[f(s), f(t_{2i-1})] \subseteq G$ for all $s, \in B_{\delta_{t_{2i-1}}}(t_{2i-1})$. So, if for each $i \in \{1, \ldots, m-1\}$ we can find $t_{2i} \in B_{\delta_{t_{2i-1}}}(t_{2i-1}) \cap B_{\delta_{t_{2i-1}}}(t_{2i+1})$, then we will be able to set $z_i = f(t_i)$ for all $t \in \{1, \ldots 2m-1\}$ and be done.

> Fortunately, note that $B_{\delta_{t_{2i-1}}}(t_{2i-1})$ and $B_{\delta_{t_{2i-1}}}(t_{2i+1})$ are two open sets which cover $[t_{2i-1}, t_{2i+1}]$. After all, suppose $B_{\delta_{t_{2i-1}}}(t_{2i-1})$ doesn't cover all of $[t_{2i-1}, t_{2i+1}]$. Then we know that $\alpha := t_{2i-1} + \delta_{t_{2i-1}}$ is contained in some $B_{\delta_{t_{2j-1}}}(t_{2j-1})$. But that j can't be less than i since that would imply $B_{\delta_{t_{2i-1}}}(t_{2i-1})$ is entirely contained in $B_{\delta_{t_{2j-1}}}(t_{2j-1})$. For similar reasons, that j also can't be greater than $i+1$. Hence $[t_{2i-1} + \delta_{t_{2i-1}}, t_{2i+1}] \subseteq B_{\delta_{t_{2i-1}}}(t_{2i+1})$ and we've proven all of $[t_{2i-1}, t_{2i+1}]$ is contained in our two open sets.
>
> But now note that $[t_{2i-1}, t_{2i+1}]$ is connected. And since both of our open sets have a nonempty intersection with $[t_{2i-1}, t_{2i+1}]$, we thus must have that our two open sets intersect. And hence, we may now pick our t_{2i} in that intersection.

$(\Longleftarrow)$
This direction is easy because you just need to define the obvious path going from a to b to show that G is path connected and thus connected.

Meanwhile here is how Conway show this same result:

$(\Longrightarrow)$
Given a fixed $a \in G$ let $A := \{b \in G : \exists P$ a polygon going from a to $b\}$. Now we shall show that A is both open and closed.

- Suppose $b \in A$. Then because G is open, there exists a ball $B_\varepsilon(b)$ of radius ε about b such that $B_\varepsilon(b) \subseteq G$. And if P is any polygon in G going from a to b, then for any $z \in B_\varepsilon(b)$ we have that $P \cup [b, z]$ is a polygon in G going from a to z. So, $B_\varepsilon(b)$ is also contained in A, thus proving that A is open.

- Next suppose for the that $z \in \mathrm{acc}(A)$ and let $\varepsilon > 0$ be such that $B_\varepsilon(z) \subseteq G$. Then we know that there exists $b \in A$ such that $b \in B_\varepsilon(z)$. And then for any polygon $P \subseteq G$ going from a to b, we have that $P \cup [b, z]$ is a polygon going from a to z contained in G. This shows that $\mathrm{acc}(A) \subseteq A$. Hence A is closed.

Now since G is connected, the only clopen sets in G are $\emptyset$ and G. But A is clearly nonempty since it trivially contains a. So $A = G$.

$(\Longleftarrow)$
Especially now that I've realized I don't need to do the exercise associated with this proof direction, frick no I'm not copying down this direction. I don't need to reprove a weaker statement than that path connectedness implies connectedness.

Anyways time to do a bunch of busy work now for this class.

II.1.6: Prove that a set $G \subseteq X$ is open if and only if $X - G$ is closed.
The book literally defines that $X - G$ is closed iff $(X - G)^{\mathsf{C}} = G$ is open. Why was this assigned?!? Grrrr.

II.2.1:

(a) Show that a set $A \subseteq \mathbb{R}$ is connected iff for any two points $a, b \in A$ we have that $[a, b] \subseteq A$.

$(\Longrightarrow)$
Suppose there exists $a, b \in A$ such that $[a, b] \not\subseteq A$. Then pick $z \in [x, y] - A$. Importantly $(-\infty, z) \cap A$ and $(z, \infty) \cap A$ are both open in the subspace topology of A. Also, they're obviously disjoint and nonempty since they contain a and b respectively. And since $z \notin A$ and $(-\infty, z) \cup (z, \infty) = \{z\}^{\mathsf{C}}$, we know $((-\infty, z) \cap A) \cup ((z, \infty) \cap A) = A$. This shows A can be partitioned into two disjoint open sets (in the subspace topology). [I'll also mention those sets are thus clopen in the subspace topology so don't you dare take off points or I will complain a lot]. Hence A is not connected.

$(\Longleftarrow)$
Suppose A is not connected and let $U, V \subseteq A$ be open in the subspace topology such that $U \cap V = \emptyset$, $U \neq \emptyset$, $V \neq \emptyset$, and $U \cup V = A$. Next pick $a \in U$, $b \in V$, and without loss of generality assume $a < b$. If $[a, b] \subseteq A$, we must have that $U \cap [a, b]$ and $V \cap [a, b]$ are also disjoint nonempty open sets in the subspace topology of $[a, b]$. But this contradicts what we already showed in class that any interval $[a, b]$ is connected.

Suppose for the sake of contradiction that $E \subseteq [a, b]$ is clopen and $E \neq X$. Also, without loss of generality we can assume that $a \in E$ (since otherwise $a \in E^{\mathsf{C}}$ and E^{C} is clopen). Since E is open, we know there exists $\varepsilon > 0$ such that $[a, \varepsilon) \subseteq E$. Thus it is well defined to set $r := \sup\{\varepsilon > 0 : [a, \varepsilon) \subseteq E\}$.

If $r = b$, then we clearly can't have that $r \in E$ cause that contradicts that $E \neq X$. But if $r \notin E$, then we no longer have that E is closed since $[ab] - E = \{b\}$ is not open. Hence we may assume $r < b$.

Now once again suppose that $r \in E$. Then since E is open, we have that there exists $\delta > 0$ such that $(r-\delta, r+\delta) \subseteq E$. But now $(a, r+\delta) \subseteq E$, contradicting how we defined r. Hence we must have that $r \notin E$. Yet even then we have a contradiction cause we know that E^{C} is open and $r \in E^{\mathsf{C}}$. So, there exists $\delta > 0$ such that $(r - \delta, r + \delta) \subseteq E^{\mathsf{C}}$. And now this proves that $r - \delta$ is an upper bound $\{\varepsilon > 0 : [a, \varepsilon) \subseteq E\}$, thus contradicting how we defined r.

So, we can't have that $[a, b] \subseteq A$ ■

(b) Suppose $A \subseteq \mathbb{R}$ is nonempty and connected. Then A is an interval.

If A is empty or a singleton then this is trivial. Meanwhile, if A has two distinct points y_1, y_2 (where we assume without loss of generality that $y_1 < y_2$), then we know from part (a) that $[y_1, y_2] \subseteq A$. Therefore, we can choose any $x \in (y_1, y_2)$ and know $a := \inf\{t \in \mathbb{R} : [t, x] \subseteq A\} \neq x$ and $b := \sup\{s \in \mathbb{R} : [x, s] \subseteq A\} \neq x$ (note that we are letting $a = -\infty$ and $b = \infty$).

By considering a strictly decreasing sequence $(t_n)_{n \in \mathbb{N}}$ to a and a strictly increasing sequence $(s_n)_{n \in \mathbb{N}}$ to b and noting that $[t_n, s_n] \subseteq A$ by part (a), we can conclude that $(a, b) = \bigcup_{n \in \mathbb{N}} [t_n, s_n] \subseteq A$. Next suppose there exists any $t \in (-\infty, x) \cap A$ such that $t \notin [a, x]$. Then we know that $t < a$. And by part (a), we know that $[t, x] \subseteq A$. But this contradicts how we defined a. Analogous reasoning shows that there there can't exist any $s \in (x, \infty) \cap A$ such that $s \notin [x, b]$. So, $(a, b) \subseteq A \subseteq [a, b]$. This shows that A is an interval of some variety. ■

I'll continue this class on *page 274*.

10/2/2025

Math 241a (lecture 2):

To start off, let $(X, \mathcal{M}, \mu)$ be a measure space and suppose $\{E_n\}_{n \in \mathbb{N}}$ is a countable collection of sets whose union is all of X. Then given any $q \in [1, \infty]$ we can define a Fréchet space as follows:

Let $\mathcal{X}$ be the collection of measurable functions $f : X \to \mathbb{C}$ (where functions equal a.e. are identified with each other) such that $f\chi_{E_n} \in L^q(X, \mu)$ for each n. Then for each n define a seminorm $p_n(f) := \|f\chi_{E_n}\|_q$. Note that $f \in \mathcal{X}$ if and only if $\|f\|_n < \infty$ for all $n \in \mathbb{N}$.

It's obvious that $\mathcal{X}$ is a complex vector space and that each of our seminorms are in fact seminorms. We also claim that $\{p_n\}_{n \in \mathbb{N}}$ is a sufficient family. After all first suppose $q \neq \infty$ and that $f \neq 0$ a.e.. Then we know there exists $\varepsilon > 0$ and a set $A \subseteq X$ with $\mu(A) > 0$ and $|f| \geq \varepsilon$ on A. And since $0 < \mu(A) \leq \sum_{n \in \mathbb{N}} \mu(E_n \cap A)$, we know there exists $k \in \mathbb{N}$ such that:

$$p_k(f) = (\int |f|^q \chi_{E_k} \mathrm{d}\mu)^{1/q} \geq (\int \varepsilon^q \chi_{E_k \cap A})^{1/q} = \varepsilon(\mu(E_k \cap A))^{1/q} > 0$$

Meanwhile if $q = \infty$, then after choosing $k \in \mathbb{N}$ such that $\mu(E_k \cap A) > 0$, we know that $p_k(f) = \|f\chi_{E_k}\|_\infty \geq \varepsilon$.

Finally, we claim that $\mathcal{X}$ is complete. After all, note that $\mathcal{X}$ is first countable. Hence, we only need to prove that all Cauchy sequences converge in order to prove that $\mathcal{X}$ is complete. So suppose $(f_k)_{k \in \mathbb{N}}$ is a Cauchy sequence of functions in $\mathcal{X}$ (meaning that $p_n(f_{k_1} - f_{k_2}) \to 0$ as $k_1, k_2 \to \infty$ for all n).

Since $L^q(X,\mu)$ is a complete metric space, we know that for each $n \in \mathbb{N}$ there exists a function $g_n \in L^q(X,\mu)$ such that $f_k\chi_{E_n} \to g_n$ as $k \to \infty$. And since $f_k(\chi_{E_n}) = 0$ on E_n^{C} for all k, we can assume without loss of generality that $g_n = 0$ on E^{C} as well. Now if m is another integer then we know $f_k\chi_{E_n} - f_k\chi_{E_m} \to g_n - g_m$ in L^q. And by passing to a suitable subsequence we can get that $f_{k_j}\chi_{E_n} - f_{k_j}\chi_{E_m} \to g_n - g_m$ pointwise a.e.. However, $f_{k_j}\chi_{E_n} - f_{k_j}\chi_{E_m} = 0$ on $E_n \cap E_m$. This proves that the set of points in $E_n \cap E_m$ where $g_n \neq g_m$ has measure 0.

Now let $A \coloneqq \bigcup_{n\in\mathbb{N}}\bigcup_{m\in\mathbb{N}}\{x \in E_n \cap E_m : g_n(x) \neq g_m(x)\}$. Then A is a null set since it is a countable collection of null sets. At the same time, for any $n, m \in \mathbb{N}$ we have that $g_n(x) = g_m(x)$ for all $x \in E_n \cap E_m \cap A^{\mathsf{C}}$. So, there is a well defined function $g : A^{\mathsf{C}} \to \mathbb{C}$ such that $g = \bigcap_{n\in\mathbb{N}}(g_n|_{E_n\cap A^{\mathsf{C}}})$. And by setting $g = 0$ on A, we can now easily see that $g \in \mathcal{X}$ and $f_k\chi_{E_n} \to g$ in $L^q(X,\mu)$ for all $n \in \mathbb{N}$.

A typical place where the above construction is used is when we want give L^q_{loc} a topology (where $q \in [1,\infty]$). Specifically, if X is a LCH space and μ is a Radon measure on X, we shall say $f \in L^q_{\mathrm{loc}}(X,\mu)$ if for any compact set $K \subseteq X$ we have that $f\chi_K \in L^q$.

Side note: in math 240a we had defined that $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n, m)$ if f is integrable when restricted to any bounded set. That definition is a subcase of ours because in $\mathbb{R}^n$, a set is bounded iff it is contained in a compact set.

If X is σ-compact then we endow $L^q_{\mathrm{loc}}(X,\mu)$ with a topology that turns it into a Fréchet space as follows.

Let $\{U_n\}_{n\in\mathbb{N}}$ be an increasing sequence of precompact open sets in X whose union is all of X. Then define $p_n(f) = \|f\chi_{\overline{U_n}}\|_q < \infty$ for all n. If we can show that $p_n(f) < \infty$ for all n iff $f \in L^q_{\mathrm{loc}}(X,\mu)$, then we will be able to conclude by our prior reasoning that these seminorms turn $L^q_{\mathrm{loc}}(X,\mu)$ into a Fréchet space.

Fortunately it is obvious that if $\|f\chi_F\|_q < \infty$ for all compact $F \subseteq X$, then we know that $\|f\chi_{\overline{U_n}}\|_q < \infty$ for all $n \in \mathbb{N}$. As for the other direction, note that since the $\{U_n\}_{n\in\mathbb{N}}$ form an open cover of X such that $U_n \subseteq U_{n+1}$ for all n, we know that any compact $F \subseteq X$ will eventually be entirely contained in some $\overline{U_n}$. And then if $f\chi_{\overline{U_n}}$ is in L^q, we know that $\|f\chi_F\|_q < \|f\chi_{\overline{U_n}}\|_q < \infty$.

As a special case of the last construction, if $q = \infty$ and we restrict ourselves to working with continuous function, then we can express the topology of uniform convergence of continuous functions on compact sets as a Fréchet space.

All we need to do is show that if $(f_k)_{k\in\mathbb{N}} \subseteq C(X)$ is Cauchy, then the function f which the f_k converge to is also in $C(X)$. Fortunately, note that for any $x \in X$ we know there exists a precompact open ball V containing x. And since $\overline{V}$ is compact, we know there exists some $n \in \mathbb{N}$ with $V \subseteq \overline{U_n}$. So since $\|(f_k - f)\chi_{\overline{U_n}}\| \to 0$, we know that $f_k \to f$ uniformly on $V \subseteq \overline{U_n}$. This shows that f is continuous at x.

Some other Fréchet spaces I've worked with are $C_c^\infty(K)$ and the Schwartz space $\mathcal{S}$.

Given a topological vector space $\mathcal{X}$, (even if it isn't normed) we still define $\mathcal{X}^*$ to be the space of continuous linear functionals on $\mathcal{X}$. You may recall that if the topology on $\mathcal{X}$ is generated by the family $\{p_\alpha\}_{\alpha \in A}$ of seminorms, then (due to Folland proposition 5.15 in math 240b) we have that $f \in \mathcal{X}^*$ iff there exists $\alpha_1, \ldots, \alpha_n \in A$ and $C > 0$ such that $|f(x)| \leq C \sum_{k=1}^{n} p_{\alpha_k}(x)$.

Lemma 1.1.24: Let $\mathcal{X}$ be a topological vector space with a topology generated by a sufficient family $\{p_\alpha\}_{\alpha \in A}$ of seminorms. Then for all $x \in \mathcal{X} - \{0\}$ there exists $f \in \mathcal{X}^*$ such that $f(x) \neq 0$.

Proof:

For all $x \in \mathcal{X} - \{0\}$ there exists $\alpha \in A$ such that $p_\alpha(x) \neq 0$. Hence, we can definte the linear functional $g(cx) = cp_\alpha(x)$ for all $y = cx$ in the subspace $Kx \subseteq \mathcal{X}$.

If you doubt that this is well defined, suppose $c_1 x = c_2 x$. Then $(c_1 - c_2)x = 0$. And if $c_1 - c_2 \neq 0$, we could divide both sides to get that $x = 0$. But that contradicts the hypothesis of the lemma.

By the Hahn-Banach theorem we know there is linear functional $f : \mathcal{X} \to \mathbb{C}$ such that $|f(x)| \leq p_\alpha(x)$ for all $x \in \mathcal{X}$ and $f|_{Kx} = g$ (which means $f(x) = p_\alpha(x) \neq 0$). ■

Recall that the weak topology on a topological vector space $\mathcal{X}$ is defined as the weak topology generated by the continuous linear functionals in $\mathcal{X}^*$. Conveniently the weak topology is always a seminormed topology. After all, for each $f \in \mathcal{X}^*$ we may define $p_f(x) := |f(x)|$. Then it is easy to check that each p_f is in fact a seminorm. Also, it is easy to check that convergence of nets in the weak topology is equivalent to convergence with respect to each of the seminorms.

Observation: Suppose $(\mathcal{X}, \mathcal{T})$ is a topological vector space and $\mathcal{T}'$ is the weak topology on $\mathcal{X}$. Then because each $f \in \mathcal{X}^*$ is continuous, we must have that $\mathcal{T}' \subseteq \mathcal{T}$. Consequently, if $\mathcal{T}$ contains no convex sets other than $\mathcal{X}$, then we automatically know that the only continuous linear functional on $\mathcal{X}$ is the zero functional (since a nonzero functional would yield a nontrivial seminorm which would yield a convex open set in $\mathcal{T}' \subseteq \mathcal{T}$ not equal to $\mathcal{X}$).

By lemma 1.1.24, we know that if the topology on $\mathcal{X}$ is generated by a sufficient family of seminorms, then the weak topology is also generated by a sufficient family of seminorms and is thus Hausdorff.

I'll continue with this class on *page 281*.

10/3/2025

Math 200a (lecture 3):

Set 1 problem 4: Suppose a and b are non-negative integers. Prove that if $k = \frac{a^2+b^2}{1+ab}$ is an integer, then k is a perfect square.

To start off, since we are requiring a and b to be nonnegative, we know that any $k = \frac{a^2+b^2}{1+ab}$ is also nonnegative. So, we only need to worry about nonnegative solutions of k.

Next note that both $k = 0$ and $k = 1$ are perfect squares. So, we don't need to worry about those cases when solving this problem and can always assume $k \geq 2$. With that out of the way, fix any $k \geq 2$ and define $V := \{(a,b) \in \mathbb{Z}^2 : a^2 + b^2 - kab - k = 0\}$. We need to show that if V contains any pair (a_0, b_0) such that $a_0 \geq 0$ and $b_0 \geq 0$, then k is a perfect square.

To do this, first note that if $(a,b) \in V$, then we also have that $(b,a) \in V$. Also note that if $b \in \mathbb{R}$, then the sum of the roots of the quadratic $x^2 - kbx + (b^2 - k)$ must add up to kb. Also, kb is an integer if b is an integer. So, if $(a,b) \in V$, we know that $(kb - a, b) \in V$. By combining this observation with the last observation, know that $(a,b) \in V \Longrightarrow (b, kb - a) \in V$.

Or as the hint that professor Golsefidy gave says, multiplication by the matrix $\left(\begin{smallmatrix} 0 & 1 \\ -1 & k \end{smallmatrix}\right)$ is a symmetry of V.

Claim 1: If $(a,b) \in V$ and $0 \leq b \leq a$, then $kb - a < b$.

Proof:

Suppose for the sake of contradiction that $kb - a \geq b$. Then:

$$0 = a^2 + b^2 - kab - k = b^2 - k - a(kb - a).$$

Or in other words, $k + a(kb - a) = b^2$. But since $a \geq b$ and $kb - a \geq b$, we know that $k + b^2 \leq b^2$. And since $k \geq 2$, this is a contradiction.

Claim 2: If $(a,b) \in V$ and $a \geq 1$, then $b \geq 0 > -1$.

Proof:

We know that $a^2 + b^2 = k(ab + 1)$. And since $k > 0$ and $a^2 + b^2 > 0$ (because otherwise we'd have $k = 0$), this in turn means that $ab + 1 > 0$. Or in other words, $b > -\frac{1}{a}$. But since a is an integer, we know that $\frac{-1}{a} \geq -\frac{1}{1} = -1$. And since b is an integer, this means that $b \geq 0$.

Claim 3: If $(a,b) \in V$, then $a \neq b$. Consequently, if $(a,b) \in V$ where $0 \leq b \leq a$, we know that $b < a$.

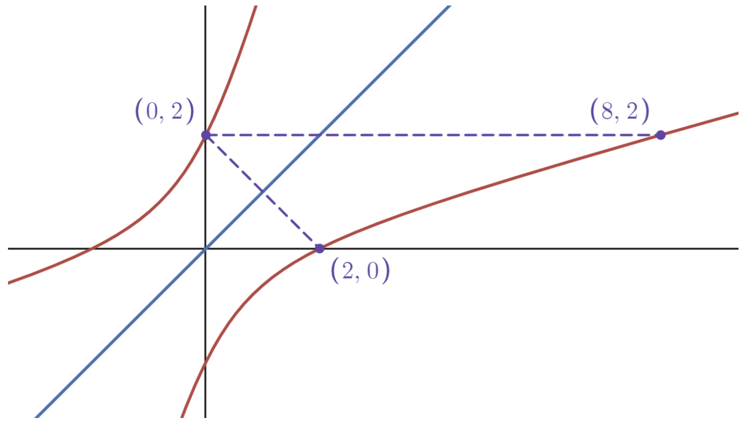
Proof:

If $b = a$ and $(a,b) \in V$, then $a^2 + a^2 - ka^2 - k = 0$. Or in other words $(2-k)a^2 = k$. But since $a^2 \geq 0$ and $2 - k \leq 0$, we know that $(2-k)a^2 \leq 0$. This contradicts that $k \geq 2$.

Now supposing there exists $(a,b) \in V$ with $a, b \geq 0$, we may choose $(a_0, b_0) \in V$ such that $0 \leq b_0 \leq a_0$ and a_0 is the least nonnegative integer for which such a b_0 exists. Then note that we must have that $b_0 = 0$. For if $b \geq 1$, then we know from the prior three claims that that $(b_0, kb_0 - a_0)$ is in V with $0 \leq kb_0 - a_0 < b_0$ and $b_0 < a_0$. But this contradicts how we chose a_0.

So if there exists nonnegative integers a, b such that $k = \frac{a^2+b^2}{1+ab}$, then we've also proven there is an integer $c \geq 0$ such that $k = \frac{c^2+0^2}{1+c(0)} = c^2$. ■

As for how someone would actually come up with this reasoning without all the hints that our professor gave, note that when $k \geq 2$, then the level set $x^2 + y^2 - kxy = k$ defines a hyperbola. And then our symmetry is hopping points via the following process:



This is a complete tangent unrelated to the class but I can't help but want to look into it briefly. Any curve graphed by the equation $Ax^2 + By^2 + Cxy + Dx + Ey = F$ in the xy-plane is called a conic. Importantly, if $C = 0$ then we already know from high school precalculus how to tell if the graphed curve is an ellipse, parabola, or hyperbola. However, the C coefficient complicates things. Is there a test we can do to still know what type of conic is graphed by our equation when $C \neq 0$?

Let $f(x, y) = Ax^2 + By^2 + Cxy + Dx + Ey$ and consider any matrix $R = \begin{pmatrix} r_1 & r_2 \\ r_3 & r_4 \end{pmatrix}$.

Then while it is messy, we can evaluate that:

$$\begin{aligned}(f \circ R)(x, y) = (Ar_1^2 + Br_3^2 + Cr_1r_3)x^2 &+ (Ar_2^2 + Br_4^2 + Cr_2r_4)y^2 \\ &+ (2Ar_1r_2 + 2Br_3r_4 + Cr_1r_4 + Cr_2r_3)xy + (Dr_1 + Er_3)x \\ &+ (Dr_2 + Er_4)y\end{aligned}$$

In particular, we care about when R is a rotation matrix. So, we can assume there exists some θ such that:

$$R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

And now we want to fix θ such that:

$$\begin{aligned}0 &= 2Ar_1r_2 + 2Br_3r_4 + Cr_1r_4 + Cr_2r_3 \\ &= -2A\cos(\theta)\sin(\theta) + 2B\sin(\theta)\cos(\theta) + C\cos^2(\theta) - C\sin^2(\theta) \\ &= 2(B - A)\cos(\theta)\sin(\theta) + C(\cos^2(\theta) - \sin^2(\theta)) \\ &= (B - A)\sin(2\theta) + C\cos(2\theta).\end{aligned}$$

Since we were originally motivated by the possibility that $B \neq 0$, for now I'm just going to assume $B \neq 0$. That way, we can explicitely calculate that $\theta = \frac{1}{2}\cot^{-1}(\frac{A-C}{B})$. And now by plugging that θ into our matrix R we can find a rotation matrix such that $(f \circ R)(x, y)$ is another degree two polynomial but with 0 for its xy coefficient. And this proves that it is always possible to rotate our conic so that the B term is 0.

Unfortunately I don't have time to continue this tangent right now. I'll continue it later on
page ___.

Before doing more 200a stuff, I unfortunately need to quickly do the rest of my math 220a homework since it is due in a few hours. So here it is:

Exercise II.2.4: Prove that if $\{D_j : j \in J\}$ is a collection of connected subsets of X and if for each j and k in J we have $D_j \cap D_k \neq \emptyset$, then $D := \bigcup_{j \in J} D_j$ is connected.

Let A be a set that is open and closed in D. Then we know that $A \cap D_j$ is both open and closed in the subspace topology of D_j. And since D_j is connected, this means that either $A \cap D_j = \emptyset$ or $A \cap D_j = D_j$ for all $j \in J$.

Now suppose that $A \neq D$. That way, we know there is some $x \in D$ such that $x \notin A$. But then in turn there is some $k \in J$ such that $x \in D_k - A$. So, we know that $A \cap D_k = \emptyset$. And since $D_j \cap D_k \neq \emptyset$ for all $j \in J$, we know that for each $j \in J$ there exists $y_j \in D_j$ such that $y_j \in D_j - A$. And hence, $A \cap D_j = \emptyset$ for all j. This shows $A = A \cap D = \emptyset$.

So since the only clopen sets in D are $\emptyset$ and D, we know that D is connected. ■

Exercise II.2.5: Let (X, d) be a metric space.

(a) Show that if $F \subseteq X$ is closed and connected, then for every pair of points $a, b \in F$ and each $\varepsilon > 0$ there are points $z_0, \ldots, z_n$ in F with $z_0 = a$, $z_n = b$, and $d(z_{k-1}, z_k) < \varepsilon$ for $1 \leq k \leq n$. Is the hypothesis that F is closed needed?

To start off, I don't know what Conway's obsession with writing the word "closed" is because it is entirely unrelated to the proof which I imagine he intended for people to write (considering I'm really just adapting Conway's proof for *theorem II.2.3*).

Fix any $a \in F$ and $\varepsilon > 0$, and let A be the set of $b \in F$ for which there exists $z_0, \ldots, z_n \in F$ as in the statement of the exercise. Clearly $A \neq \emptyset$ since $a \in A$. Meanwhile, we claim A is clopen in F.

- Suppose $b \in A$ and let $z_0, \ldots, z_n$ be a collection of points of F with $a_0 = z_0$, $b = z_n$ and $d(z_{k-1}, z_k) < \varepsilon$ for each $1 \leq k \leq n$. Then for any $w \in B_\varepsilon(b) \cap F$ (the ball of radius ε centered at b restricted to F), we know that $z_0, \ldots, z_n, w$ is another collection of points of F satisfying the conclusion of the exercise. So, $B_\varepsilon(b) \cap F \subseteq A$. And this proves that A is open in F.

- In a similar vein, if $w \in F - A$, then we must have that $B_\varepsilon(w) \cap A = \emptyset$. So $F - A$ is also open in F, and this shows that A is closed in F.

Now since F is connected and $A \neq \emptyset$ is clopen, we know that $A = F$. ■

(b) Show that even if F is closed and satisfies the conclusion the previous statement, then we still don't necessarily have that F is connected.

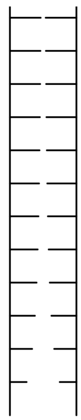
We shall construct a "broken ladder" in $\mathbb{R}^2$. Specifically, define

$$L := \{(-1, y) \in \mathbb{R}^2 : y > 0\} \cup \{(+1, y) \in \mathbb{R}^2 : y > 0\} \cup \bigcup_{k=1}^{\infty} \{(x, k) \in \mathbb{R}^2 : \tfrac{1}{k+1} \leq |x| \leq 1\}.$$

In order to see that L is closed, consider that:

$$L^{\mathsf{C}} = (((-\infty,-1)\cup(1,\infty))\times\mathbb{R})\cup(\mathbb{R}\times(-\infty,0))\cup\bigcup_{k=1}^{\infty}((-1,1)\times(k-1,k))\cup\bigcup_{k=1}^{\infty}(\tfrac{-1}{k},\tfrac{1}{k})\times(k-2,k).$$

Importantly, this shows that L^{C} is a union of open sets. So L^{C} is open and we in turn know that L is closed. Next note that L is not connected. After all, note that $A := L\cap((0,\infty)\times\mathbb{R}) = L\cap([0,\infty)\times\mathbb{R})$ is both open and closed in L. That said, $A\neq\emptyset$ and $A\neq L$.



Finally note that L does satisfy the conclusion of the problem statement in part (a). After all, $L\cap((0,\infty))$ and $L\cap((-\infty,0))$ are easily checked to be the two connected components. So if a,b are in the same component, then we know from part (a) that a collection $z_0,\dots,z_n$ of points exists as described in the problem. Meanwhile if a and b are in opposite components and $\varepsilon>0$, then we know there is some $k\in\mathbb{N}$ with $\frac{1}{k+1}<\frac{1}{2}\varepsilon$. And then $(\frac{-1}{k+1},k)$ and $(\frac{1}{k+1},k)$ are two points in the opposite components that have distance ε from each other. ■

(Also see the drawing of L to the side.)

Ok back to math 200a stuff now.

The kernel of a group action $a : G\times X\to X$ is the kernel of the induced group homomorphism $G\to S_X$ (see the *theorem at the bottom of page 246*).

Note that g is in the kernel of the group action $(g,x)\mapsto g\cdot x$ precisely when for all $x\in X$ we have that $g\cdot x = x$. Also, the kernel K of a group action is a normal subgroup and $G/K\hookrightarrow S_X$ (this arrow notation means that the induced homomorphism described in that theorem from G/K to S_X is injective / a monomorphism).

Given a group action $G\curvearrowright X$ we define:

- the fixed set of g to be $\mathrm{Fix}(g) = X^g := \{x\in X : g\cdot x = x\}$ (where g is fixed),
- the stabilizer of x to be $G_x := \{g\in G : g\cdot x = x\}$ (where x is fixed).

Lemma: G_x is a subgroup of G.

Proof:

- $1_g\cdot x = x \implies 1\in G_x$.
- Suppose $g_2\in G_x$. Then $g_2\cdot x = x\implies g_2^{-1}\cdot(g_2\cdot x) = g_2^{-1}\cdot x$. But note that $g_2^{-1}\cdot(g_2\cdot x) = (g_2^{-1}g_2)\cdot x = 1_G\cdot x = x$. So $g_2^{-1}\in G_x$.
- If $g_1,g_2\in G_x$, then $(g_1g_2^{-1})\cdot x = g_1\cdot(g_2^{-1}\cdot x) = g_1\cdot x = x$. So $g_1g_2^{-1}\in G_x$.

This proves that G_x is a subgroup of G. (which by the way we shall denote as $G_x < G$).

■

Lemma:

(a) For all $g' \in G$ we have that $\mathrm{Fix}(g'g(g')^{-1}) = g' \cdot \mathrm{Fix}(g) := \{g' \cdot x \in X : g \cdot x = x\}$.

Proof:

$$\begin{aligned} x \in \mathrm{Fix}(g'g(g')^{-1}) &\iff (g'g(g')^{-1}) \cdot x = x \\ &\iff g \cdot ((g')^{-1} \cdot x) = (g')^{-1} \cdot x \\ &\iff (g')^{-1} \cdot x \in \mathrm{Fix}(g) \iff x \in g' \cdot \mathrm{Fix}(g). \end{aligned}$$

(b) For all $g \in G$ we have that $G_{g \cdot x} = gG_x g^{-1}$

Proof:

$$\begin{aligned} g' \in G_{g\cdot x} &\iff g' \cdot (g \cdot x) = g \cdot x \\ &\iff (g^{-1}g'g) \cdot x = x \iff g^{-1}g'g \in G_x \iff g' \in gG_xg^{-1}. \blacksquare \end{aligned}$$

Corollary: Suppose $G \curvearrowright X$ and $|X| < \infty$. Then $g \mapsto |\mathrm{Fix}(g)|$ is a class function, meaning that $|\mathrm{Fix}(g'g(g')^{-1})| = |\mathrm{Fix}(g)|$ (or in other words $|\mathrm{Fix}(g)|$ is constant on any given conjugate classes).

Proof:

$|\mathrm{Fix}(g'g(g')^{-1})| = |g' \cdot \mathrm{Fix}(g)|$ by the last lemma. And since $x \mapsto g' \cdot x$ is an element of S_X, we know that $|g' \cdot \mathrm{Fix}(g)| = |\mathrm{Fix}(g)|$. $\blacksquare$

The G-orbit of $x \in X$ is the set of all points in X that are G-similar to x. Or to put into other words, we define $G \cdot x := \{g \cdot x \in X : g \in G\}$ and say that x' is G-similar to x if $x' = g \cdot x \in G \cdot x$ for some $g \in G$. Also, in that case we denote $x' \sim x$.

Lemma: $\sim$ is an equivalence relation.

Proof:

- $x \sim x$ as $1_G \cdot x = x$.
- $x \sim y \implies y \sim x$ as $x = g \cdot y \implies g^{-1} \cdot x = y$.
- If $x \sim y$ and $y \sim z$ then let $g_1, g_2 \in G$ be such that $x = g_1 \cdot y$ and $y = g_2 \cdot z$. Then $x = (g_1 g_2) \cdot z$. So $x \sim z$. $\blacksquare$

It's now clear that the G-orbit of x: $G \cdot x$, is the equivalence class of x with respect to $\sim$. Thus, we define $X/G := \{G \cdot x : x \in X\}$. Also note that X/G is a partition of X. As a result, we know that $|X| = \sum\limits_{G\cdot x \in X/G} |G \cdot x|$.

Theorem (Orbit-stabilizer): The map $G/G_x \to G \cdot x$ given by $gG_x \mapsto g \cdot x$ is a bijection. Hence $|G \cdot x| = [G : G_x]$ (where the latter is the number of left cosets of G in G_x).

Proof:

We first show this map is well-defined. Suppose $g_1 G_x = g_2 G_x$. Then $g_2 = g_1 h$ for some $h \in G_x$. And in turn $g_2 \cdot x = (g_1 h) \cdot x = g_1 \cdot (h \cdot x) = g_1 \cdot x$.

Next we show injectivity. Assume $g_1 \cdot x = g_2 \cdot x$. Then $g_2^{-1} \cdot (g_1 \cdot x) = x$. So $g_2^{-1} g_1 \in G_x$. Or in other words, $g_1 G_x = g_2 G_x$.

Finally, surjectivity is obvious from the fact that $G \cdot x$ is the set of $y \in X$ such that there exists $g \in G$ with $g \cdot x = y$. $\blacksquare$

Note that $|G \cdot x| = 1$ iff $\forall g \in G,\ \ g \cdot x = x$ iff $x \in \mathrm{Fix}(G)$ where:

$$\mathrm{Fix}(G) = X^G := \{x \in X : \forall g \in G,\ \ g \cdot x = x\}.$$

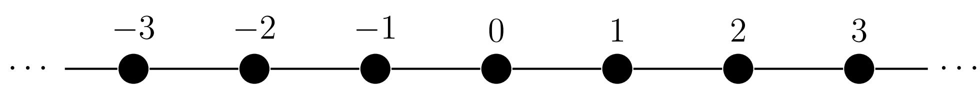
This leads to the equation:

$$|X| = \sum_{\substack{G \cdot x \in X/G \\ |G \cdot x| = 1}} |G \cdot x| + \sum_{\substack{G \cdot x \in X/G \\ |G \cdot x| > 1}} |G \cdot x| = |\mathrm{Fix}(G)| + \sum_{\substack{G \cdot x \in X/G \\ |G \cdot x| > 1}} [G : G_x].$$

I need to do the rest of the math 200a homework still. So I'm going to take a break from taking lecture notes to do the homework.

Set 1 Problem 3: Find the automorphism group of the Cayley graph of $\mathbb{Z}$ with respect to $\{-1, +1\}$.

To start off, note that $\{n, m\}$ is an edge of $\mathrm{Cay}(\mathbb{Z}, \{-1, 1\})$ iff $n - m = \pm 1$. This yields the infinite graph which I've attempted to draw below.



Now from this graph it is clear that reversing the graph is a symmetry. Specifically, define $\tau(n) = -n$. Then $\tau(n) - \tau(m) = -n - (-m) = m - n = -(n - m)$. Hence, $n - m = \pm 1$ iff $\tau(n) - \tau(m) = \mp 1$ and we thus know that τ preserves the edges of our graph and is thus a symmetry.

Another obvious symmetry of our graph are index shifts. Specifically define $\sigma(n) = n + 1$. Then $\tau(n) - \tau(m) = n - m$ for all $n, m \in \mathbb{N}$ and it is thus obvious that τ preserves the edges of graph and is a symmetry.

I glossed over this point before but technically we also need to show σ and τ are bijections. To do this, just note that σ^{-1} is given by $n \mapsto n - 1$ and $\tau^{-1} = \tau$. So, both maps are invertible.

Now we claim that every automorphism of $\mathrm{Cay}(\mathbb{Z}, \{-1, 1\})$ is some composition of τ and σ. To prove this, let θ be any arbitrary automorphism. We know that $\theta(0) = k$ for some $k \in \mathbb{Z}$. And in turn we have that $(\sigma^{-k} \circ \theta)(0) = 0$. Next note that $(\sigma^{-k} \circ \theta)(1)$ equals either $+1$ or -1. In the former case, we can trivially say that $\tau^0 \circ \sigma^{-k} \circ \theta$ fixes both 0 and 1. As for the latter case, since $\tau(0) = 0$ and $\tau(-1) = +1$, we can say that $\tau^1 \circ \sigma^{-k} \circ \theta$ fixes both 0 and 1. Either way, this shows there exists a graph automorphism $\psi = \sigma^k \circ \tau^i$ (where $k \in \{\mathbb{Z}\}$ and $i \in \{0, 1\}$) such that $\psi^{-1} \circ \theta$ fixes both 0 and 1.

Observation: If $\phi \in \mathrm{Aut}(\mathrm{Cay}(\mathbb{Z}, \{-1, 1\}))$ with $\phi(0) = 0$ and $\phi(1) = 1$, then $\phi = \mathrm{Id}$.

To prove this, we do induction separately on the positive integers and then on the negative integers.

- Suppose $n \geq 1$ and we've already shown that $\phi(k) = k$ for all $0 \leq k \leq n$. Then since ϕ is a graph automorphism, we must have that $\phi(n + 1) = \phi(n) \pm 1$. But since ϕ is a bijection and we already know that $\phi(n - 1) = n - 1 = \phi(n) - 1$, this means we can only have that $\phi(n + 1) = \phi(n) + 1 = n + 1$. By induction this means that $\phi(n) = n$ for all $n \geq 0$.

- Next suppose $n \leq 0$ and we've shown for all $k \geq n$ that $\phi(k) = k$. Then like before we must have that $\phi(n - 1) = \phi(n) \pm 1 = n \pm 1$ since ϕ is a graph automorphism. But since ϕ is a bijection and we already know $\phi(n+1) = n+1$, we can only have $\phi(n-1) = n-1$. By induction this means that $\phi(n) = n$ for all $n \in \mathbb{Z}$.

Thus $\psi^{-1} \circ \theta = \mathrm{Id}$. Or in other words $\theta = \psi = \sigma^k \tau^i$ where $k \in \mathbb{Z}$ and $i \in \{0, 1\}$. This shows that $\mathrm{Aut}(\mathrm{Cay}(\mathbb{Z}, \{-1, 1\})) = \langle \sigma, \tau \rangle$.

Now the homework sheet specifically tells us to list out all the elements of the group of automorphisms. To do this, we need to show that $\sigma^{k_1} \circ \tau^{i_1} \neq \sigma^{k_2} \circ \tau^{i_2}$ if either $k_1 \neq k_2$ or $i_1 \neq i_2$.

To start off, note that σ^{k_1} and σ^{k_2} are easily checked to not equal each other when $k_1 \neq k_2$. We merely note that $\sigma^{k_1}(0) = k_1 \neq k_2 = \sigma^{k_2}(0)$.

Also, it is easy to see that $\langle \sigma \rangle = \{\sigma^k : k \in \mathbb{Z}\}$ is a cyclic subgroup of our collection of symmetries and that τ is not in that subgroup. After all the only $k \in \mathbb{Z}$ such that $\sigma^k(0) = \tau(0)$ is $k = 0$. However, $\sigma^0(1) = 1 \neq -1 = \tau(1)$. It now follows that $\langle \sigma \rangle$ and $\langle \sigma \rangle \tau$ are two disjoint cosets which partition our collection of symmetries.

Finally, we need to show that if $k_1 \neq k_2$ then $\sigma^{k_1} \circ \tau \neq \sigma^{k_2} \circ \tau$. To do this, suppose $\sigma^{k_1} \circ \tau = \sigma^{k_2} \circ \tau$. Then by composing τ on the right side we get that $\sigma^{k_1} = \sigma^{k_2}$. And by prior work, we thus know that $k_1 = k_2$.

Thus $\mathrm{Aut}(\mathrm{Cay}(\mathbb{Z}, \{-1, 1\})) = \{\sigma^k \circ \tau^i : k \in \mathbb{Z} \text{ and } i \in \{0, 1\}\}$ and we know that the representation $\theta = \sigma^k \circ \tau^i$ is unique.

As for showing how to compose elements note that:

$$\tau \circ \sigma \circ \tau(n) = \tau \circ \sigma(-n) = \tau(-n+1) = n-1 = \sigma^{-1}(n).$$

And since conjugation is a group automorphism, we know that:

- $(\sigma^m \circ \tau) \circ \sigma^n = \sigma^m \circ (\tau \circ \sigma^n \circ \tau) \circ \tau = \sigma^m \circ (\tau \circ \sigma \circ \tau)^n \circ \tau = \sigma^m \circ \sigma^{-n} \circ \tau$ $= \sigma^{m-n} \circ \tau,$
- $(\sigma^m \circ \tau) \circ (\sigma^n \circ \tau) = \sigma^m \circ (\tau \circ \sigma^n \circ \tau) = \sigma^m \circ (\tau \circ \sigma \circ \tau)^n = \sigma^m \circ \sigma^{-n} = \sigma^{m-n},$
- $\sigma^m \circ (\sigma^n \circ \tau) = \sigma^{m+n} \circ \tau$ and $\sigma^m \circ \sigma^n = \sigma^{m+n}$. ■

Set 1 Problem 2: Suppose G is a finite group and that for every positive integer n:

$$|\{g \in G : g^n = e\}| \leq n$$

(where e is the identity element of G). Use the following steps to prove that G is a cyclic group.

(a) Prove that if there is an element of order d in G, then there are exactly $\phi(d)$ elements of order d in G where $\phi(d)$ is the Euler ϕ-function (where as a reminder $\phi(d)$ equals the number of integers between 1 and d inclusive which are coprime to d).

Suppose $g \in G$ with $o(g) = d$ and then consider the cyclic subgroup $\langle g \rangle \subseteq G$. We know that $o(g^k) = \frac{o(g)}{\gcd(o(g),k)} = \frac{d}{\gcd(d,k)} = d$ iff $\gcd(k, d) = 1$. So by considering g^k for each $k \in \{1, \ldots, d\}$ with $\gcd(d, k) = d$ we get that there are at least $\phi(d)$ distinct elements of G with order d.

That said, all g^k where $k \in \{0, \dots, d-1\}$ are distinct elements of $\{g \in G : g^d = e\}$. And since $|\{g \in G : g^d = e\}| \leq d$, this proves that $h \in G$ can satisfy that $h^d = e$ only if $h = g^k$ for some integer k. And also because h^d equaling e is a necessary condition for us to have $o(h) = d$, we know that the $\phi(d)$ elements of G we found before are the only elements of G with order d.

(b) For every positive number d, let $\psi(d)$ be the number of elements of G that have order d. Show that $\psi(d) \leq \phi(d)$ and that $\psi(d) \neq 0$ implies that $d \mid |G|$.

We know that $\phi(d) \geq 1$ for all positive d since $\gcd(1, d) = 1$. So, if $\psi(d) = 0$, then we trivially know that $\psi(d) \leq \phi(d)$. Meanwhile, if $\psi(d) > 0$ then we showed in part (a) that $\psi(d) = \phi(d)$. Hence in either case we have that $\psi(d) \leq \phi(d)$.

Also, the fact that $d \mid |G|$ if $\psi(d) \neq 0$ is just a result of Lagrange's theorem (since the order of any subgroup of G must divides $|G|$ and $\phi(d) \neq 0$ implies there is a cyclic subgroup of G with order d).

(c) Prove that $\psi(d) = \phi(d)$ if d is a positive divisor of $|G|$. Deduce that G is a cyclic group.

Let $n = |G|$ and note that $\sum_{d\,|\,n} \psi(d) = n$ since every element of G has some order dividing n. At the same time, it is a somewhat well known result that $\sum_{d\,|\,n} \phi(d) = n$ for all $n \in \mathbb{N}$.

I can't find a proof of this result anywhere in my notes so I guess I'll prove it here.

Let $S = \{1, \dots, n\}$ and define $S_d := \{k \in S : \gcd(k, n) = d\}$ for each d. Clearly, the S_d form a partition of S as we range over all the divisors of n. Also note that there is a bijective correspondence between S_d and $E_{n/d} := \{k \in \{1, \dots, \frac{n}{d}\} : \gcd(k, \frac{n}{d}) = 1\}$.

Specifically note that $\gcd(m, n) = d \implies \frac{m}{d}, \frac{n}{d} \in \mathbb{Z}$ with $\gcd(\frac{m}{d}, \frac{n}{d}) = 1$. And if we also have that $m \leq n$ then clearly $\frac{m}{d} \leq \frac{n}{d}$. So, $m \in S_d \implies \frac{m}{d} \in E_{n/d}$. Meanwhile, if $\gcd(m, \frac{n}{d}) = 1$, then we know that $\gcd(dm, n) = d$. And also if $m \leq \frac{n}{d}$, then we know that $md \leq n$ Hence $m \in E_{n/d} \implies dm \in S_d$. It now follows that the map $m \mapsto \frac{m}{d}$ is an invertible map from S_d to $E_{n/d}$.

Now $|S_d| = |E_{n/d}| = \phi(\frac{n}{d})$. Also, we know that $n = |S| = \sum_{d\,|\,n} |S_d|$. So we have shown that $n = \sum_{d\,|\,n} \phi(\frac{n}{d}) = \sum_{d\,|\,n} \phi(d)$.

Since $\psi(d) \leq \phi(d)$ for all d, we thus have that:

$$n = \sum_{d\,|\,n} \psi(d) \leq \sum_{d\,|\,n} \phi(d) = n.$$

And this proves that $\sum_{d\,|\,n} \psi(d) = \sum_{d\,|\,n} \phi(d)$. Going even further, since $0 \leq \psi(d) \leq \phi(d)$ for all d, the two sums can only equal if $\psi(d) = \phi(d)$ for all d being summed over. In particular, we must have that $\phi(n) = \psi(n) \geq 1$. So, there is some element of order $n = |G|$ in G. This is equivalent to saying that G is cyclic. ■

Set 1 Problem 1: Suppose G_1 and G_2 are two groups. We say G_1 and G_2 are algebraically independent if there are no proper normal subgroups N_1 and N_2 of G_1 and G_2 respectively such that $G_1/N_1 \cong G_2/N_2$.

(a) Prove that G_1 and G_2 are algebraically independent if and only if $G_1 \times G_2$ satisfies the following property: suppose H is a subgroup of $G_1 \times G_2$ and the projection of H to the i-th component is G_i for $i = 1, 2$. Then $H = G_1 \times G_2$.

As a reminder, the group $G_1 \times G_2$ is just the cartesian product of the two groups equipped with the law of composition that $(g_1, g_2)(g_1', g_2') = (g_1g_1', g_2g_2')$.

$(\Longrightarrow)$
Suppose G_1 and G_2 are algebraically independent and then consider any subgroup $H \subseteq G_1 \times G_2$ such that $\pi_1(H) = G_1$ and $\pi_2(H) = G_2$ (where π_1 and π_2 are the projection maps). Also let e_1 and e_2 denote the identity elements of G_1 and G_2 respectively.

To start off, let $N_1 := H \cap (\{e_1\} \times G_2)$ and $N_2 := H \cap (G_1 \times \{e_2\})$. Then set $N_1' := \pi_2(N_1)$ and $N_2' := \pi_1(N_2)$. Both N_1 and N_2 are easily seen to be subgroups of $G_1 \times G_2$ as they are both intersections of groups. From there it also easy to see that N_1' and N_2' are subgroups of G_2 and G_1 respectively on account of being images of N_2 and N_1 via the homomorphisms π_2 and π_1. And of course there are obvious group isomorphisms showing that $N_1' \cong N_1$ and $N_2' \cong N_2$.

Our first big step is to show that N_1' and N_2' are normal subgroups (which in turn means that G_1/N_2' and G_2/N_1' are well-defined quotient groups).

Suppose $g_1 \in N_2'$ and let g_1' be any element of G. Since $\pi_1(H) = G_1$, we know there is some $g_2' \in G_2$ such that $(g_1', g_2') \in H$. And since H is closed under inverses, we also know that $((g_1')^{-1}, (g_2')^{-1}) \in H$. Therefore $g_1'g(g_1')^{-1} \in N_2'$ since $(g_1'g(g_1')^{-1}, g_2'e_2(g_2')^{-1}) = (g_1'g(g_1')^{-1}, e_2) \in H$. This proves that N_2' is normal in G_1. Analogous reasoning shows that N_1' is normal in G_2.

Next we define a group homomorphism ϕ from G_1 to G_2/N_1' as follows:

Given any $g_1 \in G$, let $\phi(g_1) = g_2N_1'$ where $(g_1, g_2) \in H$.

To show this is well defined, suppose $g_2, g_2' \in G_2$ both satisfy that $(g_1, g_2) \in H$ and $(g_1, g_2') \in H$. Then $(e_1, g_2^{-1}g_2') \in H$, which in turns means that $g_2^{-1}g_2' \in N_1'$. This is equivalent to saying that $g_2^{-1}g_2'N_1' = N_1'$ which in turn is equivalent to saying that $g_2'N_1' = g_2N_1'$.

Also, to see that ϕ is a homomorphism, suppose $(g_1, g_2), (g_1', g_2') \in H$. Then $(g_1g_1', g_2g_2') \in H$ and so $\phi(g_1g_1') = g_2g_2'N_1'$. But we also have that $\phi(g_1)\phi(g_2) = g_2N_1'g_2'N_1' = g_2g_2'N_1'$. So $\phi(g_1g_2) = \phi(g_1)\phi(g_2)$.

Now we claim ϕ is surjective. After all, $\pi_2(H) = G_2$ so for all $g_2 \in G_2$ there exists $g_1 \in G_1$ such that $(g_1, g_2) \in H$. And then in turn $\phi(g_1) = g_2N_1'$. We also claim that the kernel of ϕ is N_2'. After all, suppose $\phi(g_1) = N_1'$. Then we know that there is some $g_2 \in G_2$ such that $(g_1, g_2) \in H$ and $(e_1, g_2) \in H$. But since $(e_1, g_2) \in H$, we also know that $(e_1, g_2^{-1}) \in H$, and thus $(e_1g_1, g_2^{-1}g_2) = (g_1, e_2) \in H$. So, $g_1 \in N_2'$ and we've shown that $\ker(\phi) \subseteq N_2'$. Going the other direction and showing $N_2' \subseteq \ker(\phi)$ is as simple as noting that $e_2N_1' = N_1'$.

By the first isomorphism theorem, we are thus able to conclude that $\frac{G}{N_2'} \cong \frac{G}{N_1'}$.

I ran out of time so everything after this point is not being graded...

Since G_1 and G_2 are algebraically independent, this implies that $N_2' = G_1$ and $N_1' = G_2$. But now since $G_1 \times \{e_2\}$ and $\{e_1\} \times G_2$ are both contained in H are easily seen to together generate all of $G_1 \times G_2$, we know that $H = G_1 \times G_2$. This proves the property in the problem statement.

$(\Longleftarrow)$
Suppose G_1 and G_2 are not algebraically independent and let N_1 and N_2 be proper normal subgroups of G_1 and G_2 such that $G_1/N_1 \cong G_2/N_2$. Then let $\phi : G_1/N_1 \to G_2/N_2$ be a group isomorphism.

We define the set $H := \{(g_1, g_2) \in G_1 \times G_2 : \phi(g_1N_1) = g_2N_2\}$ and claim that this is a subgroup of $G_1 \times G_2$.

- Note that $(e_1, e_2) \in H$ since we must have that $\phi(N_1) = N_2$.

- Suppose $(g_1, g_2) \in H$. Then $\phi(g_1N_1) = g_2N_2$. But note that:
$$N_2 = \phi(N_1) = \phi(g_1^{-1}g_1N_1) = \phi(g_1^{-1}N_1)\phi(g_1N_1) = \phi(g_1^{-1}N_1)g_2N_2.$$
  Therefore $\phi(g_1^{-1}N_1) = (g_2N_2)^{-1} = g_2^{-1}N_2$ and we've shown that $(g_1, g_2) \in H$.

- Suppose $(g_1, g_2), (g_1', g_2') \in H$. Then we have that $\phi(g_1N_1) = g_2N_2$ and $\phi(g_1'N_1) = g_2'N_2$. And since ϕ is a group homomorphism, we get that:
$$\phi(g_1g_1'N_1)\phi(g_1N_1)\phi(g_1'N_1) = (g_2N_2)(g_2'N_2) = g_2g_2'N_2.$$
  This shows that $(g_1g_1', g_2g_2') \in H$.

Next observe that $\pi_1(H) = G_1$. After all, for any $g_1 \in G_1$ we can just pick $g_2 \in \phi(g_1N)$ and then we'll know that $(g_1, g_2) \in H$. We also know that $\pi_2(H) = G_2$. After all, since ϕ is surjective, we know that for any $g_2 \in G_2$ there exists a coset $g_1'N_1 \in G_1/N_1$ such that $\phi(g_1'N_1) = g_2N_2$. And now by just choosing any $g_1 \in g_1'N_1$ we get that $(g_1, g_2) \in H$.

That said, $H \neq G_1 \times G_2$. To see this, just pick any $g_1 \in N_1$ and $g_2 \notin N_2$. Then $\phi(g_1N_1) \neq g_2N_2$ and we have that $(g_1, g_2) \notin H$. ■

(b) Suppose G_1 and G_2 are two finite groups and $\gcd(|G_1|, |G_2|) = 1$. Then G_1 and G_2 are algebraically independent.

Let H be any subgroup of $G_1 \times G_2$ such that $\pi_1(H) = G_1$ and $\pi_2(H) = G_2$. Since π_1 and π_2 are group homomorphisms from $G_1 \times G_2$ to G_1 and G_2 respectively, we know that both $|G_1| = |\pi_1(H)|$and $|G_2| = |\pi_2(H)|$ divide $|H|$. Hence, $\mathrm{lcm}(|G_1|, |G_2|)$ divides $|H|$. Meanwhile, we have by Lagrange's theorem that $|H|$ divides $|G_1 \times G_2| = |G_1||G_2|$.

But now because $\gcd(|G_1|, |G_2|) = 1$, we have that $\mathrm{lcm}(|G_1|, |G_2|) = |G_1||G_2|$ So, we must have $|H| = |G_1||G_2|$. And this proves that $H = G_1 \times G_2$.

By part (a), we can now conclude that G_1 and G_2 are algebraically independent. ■

I'll continue with this class on *page 269*.

10/5/2025

I agreed to help present on a book (*Ergodic Theory with a view towards Number Theory* by Einsiedler and Ward) this coming Wednesday. So today I'm going to be preparing for that.

(Definition 2.1:)

- Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be probability spaces. A map $\phi : X \to Y$ is measure preserving if it is measurable and $\mu(\phi^{-1}(B)) = \nu(B)$ for all $B \in \mathcal{C}$.
- If $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ is measure-preserving, then we call the measure space $(X, \mathcal{B}, \mu)$ T-invariant and say that $(X, \mathcal{B}, \mu, T)$ (which I will often shorthand as just (X, T)) is a measure-preserving system.

Theorem A.8: Suppose $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \mu)$ are probability spaces and that $\mathcal{E}$ is an elementary family of sets that generates the σ-algebra $\mathcal{C}$. Then a measurable map $\phi : X \to Y$ is measure-preserving if and only if $\mu(\phi^{-1}(E)) = \nu(E)$ for all $E \in \mathcal{E}$.

$(\Longrightarrow)$
This is trivial.

$(\Longleftarrow)$
Let $\mathcal{A}$ be the collection of all finite disjoint unions of elements in $\mathcal{E}$, then recall from math 240a that $\mathcal{A}$ is an algebra of sets and that by Carathéodory's theorem, any σ-finite measure on $\mathcal{C}$ is uniquely determined by it's restriction to $\mathcal{A}$.

Now importantly $\lambda : \mathcal{C} \to [0, \infty]$ defined by $\lambda(C) = \mu(\phi^{-1}(C))$ is a well-defined measure (You just need to note that $\phi^{-1}(\bigcup_{n\in\mathbb{N}} C_n) = \bigcup_{n\in\mathbb{N}} \phi^{-1}(C_n)$). And since μ is σ-finite, we know λ is. So if λ and ν agree on $\mathcal{A}$, then they must agree on all of $\mathcal{C}$. Also, note that λ and ν's values on $\mathcal{A}$ are uniquely determined by their restrictions to $\mathcal{E}$. So if $\lambda|_{\mathcal{E}} = \nu|_{\mathcal{E}}$, we have that $\lambda = \nu$. ■

As a side note: For this proof we didn't actually need $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \mu)$ to be probability measures. We just needed μ and ν to be σ-finite.

A typical application of this theorem will be that if we want to check that some map on $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ is measure-preserving then we only need to check cosets of half-intervals $[a, b)$ of $[0, 1)$.

Here is a somewhat trippy example (to me):

Consider $(\mathbb{T}, \mathcal{B}_{\mathbb{T}})$ equipped with the Lebesgue measure m and define $T_2 : \mathbb{T} \to \mathbb{T}$ by $T_2(t) = 2t \pmod 1$. We claim that T_2 is a measure preserving map.

Proof:
Consider any half-interval $[a, b) \subseteq [0, 1)$. Then we have that:

$$T_2^{-1}([a, b)) = [\frac{a}{2}, \frac{b}{2}) \cup [\frac{a}{2} + \frac{1}{2}, \frac{b}{2} + \frac{1}{2}).$$

And hence $m(T_2^{-1}([a, b))) = \frac{b}{2} - \frac{a}{2} + (\frac{b}{2} + \frac{1}{2}) - (\frac{a}{2} + \frac{1}{2}) = b - a = m([a, b))$.

The reason I find this example a bit trippy is that if $b - a \leq \frac{1}{2}$ then we know that $m(T_2([a, b))) = 2m([a, b))$. So I guess this example serves to illustrate that images of sets with respect measure preserving maps don't need to have their measure preserved.

To fully dispel my confusion, suppose X and Y are equipped with probability measures μ and ν respectively, $\phi : X \to Y$ is a measure preserving map, $x \in X$, and $y = \phi(x)$. The definition of measure preserving says that for all measurable sets $A \subseteq Y$ the probability that $y \in A$ is equal to the probability that $x \in \phi^{-1}(A)$. The last example however shows that if $B \subseteq X$ is measurable then the probability that $x \in B$ *does not* have to equal the probability that $y \in \phi(B)$.

That all said, we *do* have that if ϕ is measure-preserving and bijective with a measurable inverse, then ϕ^{-1} is also measure preserving.

This is because we then have for any measurable $B \subseteq X$ that:

$$\nu(\phi(B)) = \mu(\phi^{-1}(\phi(B))) = \mu(B).$$

If anyone in class asks, see my note on integrals of induced measures on *page 193.*
Otherwise I'll just state this without proof.

Lemma 2.6: A measure μ on X is T-invariant if and only if $\int f \mathrm{d}\mu = \int f \circ T \mathrm{d}\mu$ for all $f \in L^\infty(X, \mu)$. Moreover, if μ is T-invariant, then $\int f \mathrm{d}\mu = \int f \circ T \mathrm{d}\mu$ holds for all $f \in L^1(\mu, X)$.

(Definition 2.7:) Let $(X, \mathcal{B}_X, \mu, T)$ and $(Y, \mathcal{B}_Y, \nu, S)$ be measure-preserving systems on probability spaces.

1. The system $(Y, \mathcal{B}_Y, \nu, S)$ is a factor of $(X, \mathcal{B}_X, \mu, T)$ if there are sets $X' \in \mathcal{B}_X$ and $Y' \in \mathcal{B}_Y$ with $\mu(X') = \nu(Y') = 1$, $T(X') \subseteq X'$, and $S(Y') \subseteq Y'$, as well as a measure-preserving map $\phi : X' \to Y'$ with $\phi \circ T(x) = S \circ \phi(x)$ for all $x \in X'$.
2. If ϕ in the above definition has a measurable inverse, then we say $(Y, \mathcal{B}_Y, \nu, S)$ is isomorphic to $(X, \mathcal{B}_X, \mu, T)$.

Essentially if we let ϕ be undefined on a null set, then $(Y, \mathcal{B}_Y, \nu, S)$ is a factor of $(X, \mathcal{B}_X, \mu, T)$ if there exists a function ϕ such that the diagram below commutes:

$$\begin{tikzcd}
X \arrow[r, "T"] \arrow[d, "\phi"'] & X \arrow[d, "\phi"] \\
Y \arrow[r, "S"'] & Y
\end{tikzcd}$$

Here is my attempt at giving an example of a factor system where the two systems are not isomorphic: (not in the book)

Let $\mathbb{T}^2$ and $\mathbb{T}$ be equipped with the Lebesgue measure, and let $\phi : \mathbb{T}^2 \to \mathbb{T}$ be the projection of $\mathbb{T}^2$ onto $\mathbb{T}$. Then given any $(c_x, c_y) \in \mathbb{T}^2$ define $T : \mathbb{T}^2 \to \mathbb{T}$ by $T(x, y) = (x + c_x, y + c_y)$ and $S : \mathbb{T} \to \mathbb{T}$ by $S(x) = x + c_x$.

Then it's clear that $T \circ \phi = \phi \circ S$, and you can also somewhat easily check that T, S, and ϕ are measure-preserving.

Here is a non-trivial example (from the book) of two systems that are isomorphic. To start off, let $\{0, 1\}$ be equipped with the discrete topology (turning it into a compact Hausdorff space) as well as the coin flip measure $\mu_{1/2,1/2}$ on it's power set.

In other words $\mu_{1/2,1/2}(\{0\}) = \mu_{1/2,1/2}(\{1\}) = 1/2$.

Next let $X = \{0, 1\}^{\mathbb{N}}$ and equip X with the (Radon) product measure $\mu = \prod_{n \in \mathbb{N}} \mu_{1/2,1/2}$. Then the left shift map $\sigma : X \to X$ defined by $\sigma(x_0, x_1, \ldots) = (x_1, x_2, \ldots)$ preserves μ.

Why?

Since μ is defined on $\mathcal{B}_X$ and is trivially σ-finite (since μ is a probability measure), we know by similar reasoning as in *theorem A.8* that σ is measure preserving if and only if $\mu(\sigma^{-1}(U)) = \mu(U)$ for all open sets $U \subseteq \{0, 1\}^{\mathbb{N}}$ (in the product topology).

Now, Einsiedler and Ward call open sets in the product topology of $\{0, 1\}^{\mathbb{N}}$ cylinder sets. Specifically, $U \subseteq \{0, 1\}^{\mathbb{N}}$ is open iff there is a finite set $I \subseteq \mathbb{N}$ and a map $\mathbf{a} : I \to \{0, 1\}$ such that $U = \{(x_0, x_1, \ldots) \in X : x_j = \mathbf{a}(j) \text{ for all } j \in I\}$.

(Also Einsiedler and Ward denote this set $I(\mathbf{a})$.)

With this notation in mind, let $I(\mathbf{a})$ be an open set in $\mathcal{B}_X$. Then we clearly have that $\sigma^{-1}(I(\mathbf{a})) = I'(\mathbf{a}')$ where $I' = \{n + 1 : n \in I\}$ and and $\mathbf{a}' : I' \to \{0, 1\}$ is given by $\mathbf{a}'(n) = \mathbf{a}(n - 1)$. But also note that:

$$\mu(\sigma^{-1}(I(\mathbf{a}))) = \mu(I'(\mathbf{a}')) = (1/2)^{|I'|} = (1/2)^{|I|} = \mu(I(\mathbf{a}))$$

Also note that the map $\phi : X \to \mathbb{T}$ given by $\phi(x_0, x_1, \ldots) = \sum_{n=0}^{\infty} \frac{x_n}{2^{n+1}}$ is measure-preserving from (X, μ) to $(\mathbb{T}, m)$ (where m is the Lebesgue measure).

Why?

Note that the diadic rational numbers $\mathbb{Z}[\frac{1}{2}]$ are dense in the real numbers and include 0 and 1. And from there it is easy to see that the collection of cosets of half intervals $[a, b) \subseteq [0, 1)$ such that $a, b \in \mathbb{Z}[\frac{1}{2}]$ is an elementary family generating $\mathcal{B}_{\mathbb{T}}$. If we can show that $\phi^{-1}([a, b)) \in \mathcal{B}_X$ and that $\mu(\phi^{-1}([a, b))) = m([a, b))$ for all such half intervals, we will have thus proven what we want.

Let a and b be diadic rationals with $a < b$. Then we know that there is some $n \in \mathbb{N}$ with the binary expansion of a and b being $a = 0.a_0 a_1 \cdots a_n \overline{11}$ and $b = 0.b_0 b_1 \cdots b_n \overline{00}$. So, for all $0 \leq k \leq n$ define:

- $A_k := \{(x_0, x_1, \ldots) \in X : x_k > a_k \text{ and } x_i = a_i \text{ for all } i < k\}$,

- $B_k := \{(x_0, x_1, \dots) \in X : x_k < b_k$ and $x_i = b_i$ for all $i < k\}$.

Also define:

- $A' := \{(x_0, x_1, \dots) \in X : x_i = a_i$ for all $i \leq k\}$,
- $B' := \{(x_0, x_1, \dots) \in X : x_i = b_i$ for all $i \leq k\}$.

Then it is easy to see that $A', A_0, A_1, \dots, A_n$ are all disjoint open sets in X. And similarly it is easy to see that $B', B_0, B_1, \dots, B_n$ are all disjoint open sets in X. Plus:

$$\phi^{-1}([a,b)) = (A' \cup \bigcup_{k=0}^{n} A_k) \cap (B' \cup \bigcup_{k=0}^{n} B_k) - \{(b_0, b_1, \dots, b_n, 1, 1, \dots)\}.$$

This proves the first claim that ϕ is measurable.

As for actually calculating $\mu(\phi^{-1}([a,b)))$, let's now set $A = A' \cup (\bigcup_{k=0}^{n} A_k)$ and $B = B' \cup (\bigcup_{k=0}^{n} B_k)$. Because $\{(b_0, b_1, \dots, b_n, 1, 1, \dots)\}$ is a null set, we can ignore it. Also note that $A \cup B = X$. So we can evaluate that:

$$\begin{aligned}\mu(\phi^{-1}([a,b))) = \mu(A \cap B) &= \mu(A) + \mu(B) - 1 \\ &= (\textstyle\sum_{k=0}^{n} \mu(A_k)) + \mu(A') + (\sum_{k=0}^{n} \mu(B_k)) + \mu(B') - 1. \\ &= (\textstyle\sum_{k=0}^{n} \frac{1}{2^{k+1}}|a_k - 1|) + \frac{1}{2^{n+1}} + (\sum_{k=0}^{n} \frac{1}{2^{k+1}} b_k) + \frac{1}{2^{n+1}} - 1. \\ &= (-\textstyle\sum_{k=0}^{n} \frac{1}{2^{k+1}}(1 - a_k)) + b - 1. \\ &= 1 - a + b - 1 \\ &= b - a = m([a,b)).\end{aligned}$$

Finally, letting $T_2 : \mathbb{T} \to \mathbb{T}$ be the map from *earlier*, note that $\phi(\sigma(x)) = T_2(\phi(x))$ for all $x \in X$. And by restricting ϕ to the set $X' \subseteq X$ of elements that aren't eventually just an infinite string of zeros or of ones, we then have that ϕ is a bijection to the set $\mathbb{T} - \mathbb{Z}[\frac{1}{2}]/\mathbb{Z}$.

As for how we know that $\mu(X') = 1$, note that the measure of all singletons of X is 0 and that $X - X'$ is countable. So $\mu(X - X') = 0$.

At last this proves that $(X, \mathcal{B}_X, \mu, \sigma)$ is isomorphic to $(\mathbb{T}, \mathcal{B}_{\mathbb{T}}, m, T_2)$.

Note from 10/8/2025: Fuck I forgot to show that ϕ^{-1} is measurable.

A measure-preserving transformation $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$ is ergodic if for any $B \in \mathcal{B}$, $T^{-1}(B) = B \Longrightarrow \mu(B) = 0$ or $\mu(B) = 1$.

By the way, I Googled it and apparently the term "ergodic" was coined by Boltzmann in the 1800s.

Also, FYI the notation $T^{-n}(B)$ just means the preimage B with respect to T^n.

Proposition 2.14: The following are equivalent properties for a measure-preserving transformation T of $(X, \mathcal{B}, \mu)$.

(1) T is ergodic.

(2) For any $B \in \mathcal{B}$, $\mu(T^{-1}(B) \triangle B) = 0$ implies that $\mu(B) = 0$ or $\mu(B) = 1$.

(3) For $A \in \mathcal{B}$, $\mu(A) > 0$ implies that $\mu(\bigcup_{n=1}^{\infty} T^{-n}(A)) = 1$.

(4) For $A, B \in \mathcal{B}$, $\mu(A)\mu(B) > 0$ implies that there exists $n \geq 1$ with:
$$\mu(T^{-n}(A) \cup B) > 0.$$

$(1 \implies 2)$
Assume T is Ergodic and let B be an almost invariant measurable set (meaning that $\mu(T^{-1}(B) \triangle B) = 0$). Then consider the set
$$C = \limsup T^{-n}(B) = \bigcap_{N=0}^{\infty} \bigcup_{n=N}^{\infty} T^{-n}(B).$$

For any $N \geq 0$ we have that $B \triangle \bigcup_{n=N}^{\infty} T^{-n}(B) \subseteq \bigcup_{n=N}^{\infty}(B \triangle T^{-n}(B))$.

Sanity check since I don't typically work with symmetric differences:
$$B \triangle \bigcup_{n=N}^{\infty} T^{-n}(B) = (B - \bigcup_{n=N}^{\infty} T^{-n}(B)) \cup ((\bigcup_{n=N}^{\infty} T^{-n}(B)) - B)$$

Then the result we want follows from the fact that:
- $(\bigcup_{n=N}^{\infty} T^{-n}(B)) - B = \bigcup_{n=N}^{\infty}(T^{-n}(B) - B)$,
- $B - \bigcup_{n=N}^{\infty} T^{-n}(B) \subseteq \bigcup_{n=N}^{\infty}(B - T^{-n}(B))$.

But now note that $\mu(B \triangle T^{-n}(B)) = 0$ for all $n \geq 1$. This is because firstly,
$$B \triangle T^{-n}(B) \subseteq \bigcup_{i=0}^{n-1}(T^{-i}(B) \triangle T^{-(i+1)}(B)).$$

Why?
If $x \in T^{-n}(B) - B$, then we know there must exist some $i \in \{0, \ldots, n-1\}$ such that $T^{i+1}(x) \in B$ but $T^{i}(x) \notin B$. Similarly, if $x \in B - T^{-n}(B)$, then we know there must exist some $i \in \{0, \ldots, n-1\}$ such that $T^{i}(x) \in B$ but $T^{i+1}(x) \notin B$.

And then since T is measure-preserving we know for each i that:
$$\mu(T^{-i}(B) \triangle T^{-(i+1)}(B)) = \mu(B \triangle T^{-1}(B)) = 0.$$

This proves that for $C_n := \bigcup_{n=N}^{\infty} T^{-n}(B)$ we have that $\mu(B \triangle C_n) = 0$. And in turn since the C_n form a decreasing sequence with $C = \bigcap_{n\in\mathbb{N}} C_n$, we have by the monotonicity of measures that:
$$\mu(B \triangle C) = \mu((B - \textstyle\bigcap_{n\in\mathbb{N}} C_n)) + \mu((\bigcap_{n\in\mathbb{N}} C_n) - B)$$
$$= \lim_{n\to\infty} \mu(B - C_n) + \lim_{n\to\infty} \mu(C_n - B) = \lim_{n\to\infty} \mu(B \triangle C_n) = 0$$

This shows that $\mu(B \triangle C) = 0$. Or in other words, $\mu(B) = \mu(C)$. But now note that:
$$T^{-1}(C) = \bigcap_{N=0}^{\infty} \bigcup_{n=N}^{\infty} T^{-(n+1)}(B) = \bigcap_{N=0}^{\infty} \bigcup_{n=N+1}^{\infty} T^{-n}(B) = C$$

So by the ergodicity of T we have that $\mu(C) = \mu(B)$ is either 0 or 1.

$(2 \implies 3)$
Let A be a set with $\mu(A) > 0$ and let $B = \bigcup_{n=1}^{\infty} T^{-n}(A)$. Then:
$$T^{-1}(B) = \bigcup_{n=2}^{\infty} T^{-n}(A) \subseteq B.$$

Also, since T is measure-preserving we know that $\mu(T^{-1}B) = \mu(B)$. Therefore $\mu(T^{-1}(B) \triangle B) = 0$ and hence $\mu(B) = 0$ or 1. But since $T^{-1}(A) \subseteq B$ with $\mu(T^{-1}(A)) = \mu(A) > 0$, the former is impossible. So we must have that $\mu(B) = 1$.

$(3 \Longrightarrow 4)$
Let A and B be sets of positive measure. Then $\mu(\bigcup_{n=1}^{\infty} T^{-n}(A)) = 1$. Hence:

$$0 < \mu(B) = \mu(B \cap \bigcup_{n=1}^{\infty} T^{-n}(A)) \leq \sum_{n=1}^{\infty} \mu(B \cap T^{-n}(A)).$$

And it follows that there must be some $n \geq 1$ with $\mu(B \cap T^{-n}(A)) > 0$.

$(4 \Longrightarrow 1)$
Let A be a set with $T^{-1}(A) = A$. Then for all $n \geq 1$ we have that:

$$0 = \mu(A \cap (X - A)) = \mu(T^{-n}(A) \cap (X - A)).$$

Hence to avoid contradicting (4), we must have that $\mu(A)\mu(X - A) = 0$. And that implies that either $\mu(A) = 0$ or $\mu(A) = 1$. ■

Technically there is another equivalent condition that Einsiedler and Ward prove. However, I'm not going to continue with the proof cause I was just made aware that I read the wrong book. Oops.

The real book I was supposed to read was *Recurrence in Ergodic Theory and Combinatorial Number Theory* by Harry Furstenberg. I guess the key difference between this and the last text is that this will at first focus more on topological dynamics.

A dynamical system is a compact metric space X and a group (or semigroup) G acting on X by continuous transformations. We denote our system (X, G).

Side note: a semigroup is a set and composition operator satisfying all the group axioms except possibly the identity and inverse axioms.

If $G = \mathbb{Z}$ or $\mathbb{N}$ we denote by T the transformation on X representing the action of $1 \in G$. We then denote our dynamical system as (X, T) and call it a cyclic system.

Let $T : X \to X$ be continuous. We say $x \in X$ is recurrent for T (or for the dynamical system (X, T)) if for any neighborhood V of x there exists $n \geq 1$ with $T^n(x) \in V$. Or since X is a metric space, x is recurrent if there is an increasing sequence $(n_k)_{k \in \mathbb{N}} \subseteq \mathbb{N}$ such that $T^{n_k}(x) \to x$. (note from 10/8/2025: go *here* to see why these definitions are equivalent...)

Claim: If X is a compact metric space and $T : X \to X$ is continuous, then there always exists recurrent points.

Proof:
Let $\mathcal{F}$ be the family of nonempty closed subsets $Y \subseteq X$ with $T(Y) \subseteq Y$, and order $\mathcal{F}$ by inclusion. By the finite intersection property, we know that $\bigcap_{\alpha \in A} Y_\alpha \neq \emptyset$ and is closed for all chains $\{Y_\alpha\}_{\alpha \in A} \subseteq \mathcal{F}$. Hence by Zorn's lemma there is a minimal $Y_0 \neq \emptyset$ such $T(Y_0) \subseteq Y_0$. And we claim that each $x \in Y_0$ is recurrent for T.

To show this, let $Q(x) = \overline{\{T^n(x) : n \geq 1\}}$. Then since $T(Y_0) \subseteq Y_0$, we know that $T^n(Y_0) \subseteq Y_0$ for all $n \geq 1$. And this proves that $Q(x) \subseteq \overline{Y_0} = Y_0$. At the same time note that $T(Q(X)) \subseteq Q(X)$. After all, if $y \in Q(x)$ and $(n_k)_{k \in \mathbb{N}}$ is a sequence in $\mathbb{N}$ such that $T^{n_k}(x) \to y$, then because T is continuous, we know that $T^{n_k+1}(x) \to T(y)$.

But now by the minimality of Y_0 we have that $Q(X) = Y_0$. So, $x \in Q(x)$. ■

Let K be a compact metrized group, let $a \in K$, and define $T : K \to K$ by $T(x) = ax$. We then call the dynamical system (K, T) a Kronecker system.

As a side note, a topological group is a group equipped with a topology on which products and inverses are continuous. If that topology is also compact, we say we have a compact group.

Theorem 1.2: Every point in the space of a Kronecker system is recurrent.

Proof:

We know some point x_0 is recurrent. Now let x be any point and write $x = x_0 u$. If V is a neighborhood of x, then Vu^{-1} is a neighborhood of x_0. And $a^n x_0 \in Vu^{-1}$ implies that $a^n(x_0 u) = a^n x \in V$. ■

From now on, for the sake of shorthand I will write gx instead of $T_g(x)$ when showing the continuous transformation associated with $g \in G$ acting on x. Also, for cyclic systems I'll just write Tx instead of $T(x)$.

Let (X, G) and (Y, G) be two dynamical systems with the same group/semigroup G of operators. A homomorphism from (X, G) to (Y, G) is given by a continuous map $\phi : X \to Y$ satisfying that $\phi(gx) = g\phi(x)$ for all $x \in X$ and $g \in G$.

(Definition 1.3:) A dynamical system (Y, G) is a factor of the dynamical system (X, G) if there is a homomorphism of the latter to the former given by a map ϕ of X onto Y. In this case we also say that (X, G) is an extension of (Y, G).

Proposition 1.3: If ϕ determines a homomorphism of a cyclic system (X, T_X) to (Y, T_Y), and $x \in X$ is recurrent for (X, T_X), then $\phi(x)$ is recurrent for (Y, T_Y).

Proof:

Let $(n_k)_{k \in \mathbb{N}}$ be an increasing sequence in $\mathbb{N}$ and suppose that $T_X^{n_k}(x) \to x$. Then because ϕ is continuous, we know $\phi(T_X^{n_k}(x)) \to \phi(x)$ as $k \to \infty$. But also note that $\phi(T_X^{n_k}(x)) = T_Y^{n_k}\phi(x)$. So $T_Y^{n_k}\phi(x) \to \phi(x)$ as $k \to \infty$. ■

It's 3:30am and I need to go to sleep. I'll pick this particular subject up again probably tomorrow (technically today) late at night. See *page 277*.

10/6/2025

Math 200a (lectures 4 & 5):

Here is a small example of what we were talking about in the previous class:

Suppose $G \curvearrowright G$ by conjugation and let us denote the orbit of g as:

$$\mathrm{Cl}(g) := \{g'g(g')^{-1} : g' \in G\}.$$

The stabilizer of g is $\{g' \in G : g'g(g')^{-1} = g\} = \{g' \in G : g'g = gg'\} =: C_G(g)$.
We call this set (which is a subgroup via the *lemma on page 256*) the centralizer of g.

Side note: $C_G(g)$ consists of all elements of G that commute with g.

By the orbit-stabilizer theorem, we know $|\mathrm{Cl}(g)| = [G : C_G(g)]$. Also note that:

$$\mathrm{Fix}(G) = \{g' \in G : \forall g \in G,\ \ gg' = g'g\} =: Z(G).$$

Side note: $Z(G)$ is called the center of G. Also, $Z(G) = \bigcap_{g\in G} C_G(g)$.

So $|G| = |Z(G)| + \sum\limits_{\substack{\mathrm{Cl}(g) \\ |\mathrm{Cl}(g)|>1}} [G : C_G(g)]$.

Cauchy-Frobenius Lemma: Suppose G is finite. Then:

$$|X/G| = \mathrm{Average}_{g\in G} |\mathrm{Fix}(g)| = \frac{1}{|G|} \sum_{g\in G} |\mathrm{Fix}(g)|$$

Proof:
Let $F := \{(g, x) \in G \times X : g \cdot x = x\}$. Then clearly $|F| = \sum\limits_{g\in G} |\mathrm{Fix}(g)| = \sum\limits_{x\in X} |G_x|$.

But now recalling that $|G_{g\cdot x}| = |G_x|$ since $G_{g\cdot x} = gG_xg^{-1}$, we have that:

$$\sum_{x\in X} |G_x| = \sum_{G\cdot x\in X/G} \left(\sum_{x'\in G\cdot x} |G_{x'}| \right) = \sum_{G\cdot x\in X/G} |G\cdot x| \cdot |G_x|$$

Also from the orbit stabilizer theorem and Lagrange's theorem, we have that $|G \cdot x| = [G : G_x] = \frac{|G|}{|G_x|}$. So $|G_x| = \frac{|G|}{|G\cdot x|}$. And hence:

$$|F| = \sum_{G\cdot x\in X/G} |G\cdot x| \cdot |G_x| = \sum_{G\cdot x\in X/G} |G\cdot x| \cdot \frac{|G|}{|G\cdot x|} = |G| \cdot |G/X|$$

Thus $|G| \cdot |G/X| = \sum\limits_{g\in G} |\mathrm{Fix}(g)|$. ■

Suppose $G \curvearrowright X$. Then $G \curvearrowright X \times X$ by the action $g \cdot (x, y) = (g \cdot x, g \cdot y)$. What is $|(X \times X)/G|$?

By the last lemma we have that:

$$|(X \times X)/G| = \frac{1}{|G|} \sum_{g\in G} |\{(x, y) \in X \times X : (g \cdot x, g \cdot y) = (x, y)\}|.$$

If $\mathrm{Fix}(g)$ refers to the fixed points of the action $G \curvearrowright X$, we thus have that:

$$|\{(x, y) \in X \times X : (g \cdot x, g \cdot y) = (x, y)\}|$$
$$= |\{(x, y) \in X \times X : x, y \in \mathrm{Fix}(g)\}| = |\mathrm{Fix}(g)|^2$$

So $|(X \times X)/G| = \frac{1}{|G|} \sum_{g\in G} |\mathrm{Fix}(g)|^2$.

More generally, we can say that $|(\prod_{j=1}^{n} X)/G| = \frac{1}{|G|} \sum_{g\in G} |\mathrm{Fix}(g)|^n$.

Let $\mathrm{Sub}(G)$ denote the subgroups of G. Then $G \curvearrowright \mathrm{Sub}(G)$ via conjugation.

- We denote $\mathrm{Sub}^{\triangleleft}(G)$ as the collection of normal subgroups of G. Note that $(\mathrm{Sub}(G))^G = \mathrm{Sub}^{\triangleleft}(G)$.
- If $H \in \mathrm{Sub}(G)$, then the stabilizer of H is equal to $N_G(H) \coloneqq \{g \in G : gHg^{-1} = H\}$. We call $N_G(H)$ the normalizer of H in G.

  Note that $H \triangleleft N_G(H)$ and that $N_G(H)$ is the largest subgroup of G that has H as a normal subgroup.

- The G-orbit of H consists of all subgroups of G that are conjugate to H. The orbit-stabilizer theorem implies that the number of conjugates of a group H is $[G : N_G(H)]$.

Suppose p is prime and G is a group with $|G| = p^k$ (where k is an integer). In that case we say that G is a finite p-group.

Theorem: Suppose G is a finite p-group and $G \curvearrowright X$ where X is a finite set. Then $|X| \equiv |X^G| \pmod p$.

Proof:

Note that if $H \lneqq G$ (meaning H is a proper subgroup of G), then by Lagrange's theorem we have that $[G : H] \mid p^k$ and $[G : H] \neq 1$. Therefore, $[G : H] \equiv 0 \pmod P$. And in particular, we have that $[G : G_x] \equiv 0 \pmod p$ whenever $G_x \neq G$.

Now $|X^G| = |X| - \sum\limits_{\substack{G\cdot x \in X/G \\ |G\cdot x| > 1}} [G : G_x]$.

And since $[G : G_x] \equiv 0 \pmod p$ for all G_x with $[G : G_x] = |G \cdot x| > 1$ (so that $G \neq G_x$), we thus have that $|X^G| \equiv |X| \pmod p$. ■

Theorem (Cauchy): Suppose $p \mid |G|$ where p is prime. Then $\exists g \in G$ such that $o(g) = p$.

Proof:

We want to show that $|\{x \in G : x^p = 1\}| \equiv 0 \pmod p$.

Why? Note that if $x \neq 1$ and $x^p = 1$, then we know that $o(x) = p$. Also, by showing that $|\{x \in G : x^p = 1\}| \equiv 0 \pmod p$ we know that there exists $x \in G - \{1\}$ with $x^p = 1$.

To do that, let $X \coloneqq \{(x_0, x_1, \ldots, x_{p-1}) \in G^p : x_0 x_1 \cdots x_{p-1} = 1\}$. Then note that shifting an index is a symmetry of X. After all, if $1 = x_0 x_1 \cdots x_{p-1}$, then we have that:

$$1 = x_0^{-1} 1 x_0 = x_0^{-1}(x_0 x_1 \cdots x_{p-1}) x_0 = x_1 x_2 \cdots x_{p-1} x_0.$$

So, $C_p \curvearrowright X$ where C_p is the cyclic group of order p. And by the last theorem we have that $|X| \equiv |X^{C_p}| \pmod p$.

But now note that $X^{C_p} = \{(x, \ldots, x) \in G^p : xx\cdots x = 1\} = \{x \in G : x^p = 1\}$. Also, we have that $|X| = |G|^{p-1}$ (since you can freely choose x_0 through x_{p-2} and then x_{p-1} must be whatever the inverse of $x_0 x_1 \cdots x_{p-2}$ is). And since $p \mid |G|$, we have that $|G|^{p-1} \equiv 0 \pmod p$. Hence, $|\{x \in G : x^p = 1\}| \equiv 0 \pmod p$. ■

G is called a p-group if $o(g)$ is a power of p for all $g \in G$.

Corollary: If G is finite and p is prime, then G is a p-group if and only if $|G| = p^n$ for some integer n. (i.e. our new definition agrees with the old one)

$(\Longleftarrow)$
This is just Lagrange's theorem.

$(\Longrightarrow)$
If not, then there exists a prime factor $\ell \neq p$ such that $\ell \mid |G|$. Hence by Cauchy's theorem G has an element of order ℓ (a contradiction). ■

Proposition: Suppose P is a finite p-group and $\{1\} \neq N \lhd P$. Then $Z(P) \cap N \neq \{1\}$.

Proof:
$P \curvearrowright N$ by conjugation. So $|N| \equiv |N^P| \pmod p$. Meanwhile note that:

$$\begin{aligned} x \in N^P &\Longleftrightarrow x \in N \text{ and } \forall g \in P,\ \ g \cdot x = x \\ &\Longleftrightarrow x \in N \text{ and } \forall g \in P,\ \ gxg^{-1} = x \\ &\Longleftrightarrow x \in N \text{ and } \forall g \in P,\ \ gx = xg \Longleftrightarrow x \in Z(P) \cap N. \end{aligned}$$

So $|N| \equiv |Z(P) \cap N| \pmod p$. But finally note that $|N|$ divides $|P| = p^n$ for some n. But we also know by assumption that $|N| \neq \{1\}$. So, $|N| \equiv 0 \pmod p$ and we are done. ■

The action $G \curvearrowright X$ is transitive if $|X/G| = 1$.

Proposition: If $G \curvearrowright X$ transitively and G, X are finite with $|X| > 1$, then there exists $g \in G$ with $\mathrm{Fix}(g) = \emptyset$.

Proof:
Suppose $|\mathrm{Fix}(g)| \geq 1$ for all $g \in G$. Then:

$$|X/G| = \frac{1}{|G|}\sum_{g \in G} |\mathrm{Fix}(g)| \geq \frac{\mathrm{Fix}(1) + |G| - 1}{|G|} = \frac{|X| + |G| - 1}{|G|} > 1. \blacksquare$$

Corollary: If $|G| < \infty$ and $H \lneqq G$, then $\bigcup_{g \in G} gHg^{-1} \neq G$.

Proof:
Consider $G \curvearrowright G/H$ by left translation. This is a transitive action. And since $[G : H] > 1$, we know by the last proposition that there exists $g \in G$ such that $\mathrm{Fix}(g) = \emptyset$. Now notice that $g \cdot xH = xH \Longleftrightarrow x^{-1}gx \in H \Longleftrightarrow g \in xHx^{-1}$. So if $\mathrm{Fix}(g) = \emptyset$, then $g \notin \bigcup_{x \in G} xHx^{-1}$. ■

> Side note: the prior corollary does not hold if $|G| = \infty$. For example, consider that $\mathrm{GL}_2(\mathbb{C}) = \bigcup_{x \in \mathrm{GL}_2(\mathbb{C})} xBx^{-1}$ where:
>
> $$B := \left\{ \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} \in \mathbb{C}^{2\times 2} : a, b \neq 0 \right\}$$ is the set of all invertible upper triangular matrices.

If G is a group and $H < G$, then suppose $G \curvearrowright G/H$ by left translations. What is the kernel of this action?

Note that g is in the kernel of the action iff $gxH = xH$ for all $x \in G$. And that happens iff $\forall x \in G,\ \ g \in xHx^{-1}$ which happens iff $g \in \bigcap_{x \in G} xHx^{-1}$.

We call the kernel of the action above the normal core of H in G, and denote it as $\mathrm{core}_G(H)$ (or just $\mathrm{core}(H)$ if it's obvious what G is).

Lemma: $\mathrm{core}_G(H)$ is the largest normal subgroup of G which is a subgroup of H (i.e. $\mathrm{core}_G(H) \lhd G$, $\mathrm{core}_G(H) < H$, and if $N \lhd G$ with $N < H$ then $N < \mathrm{core}_G(H)$).

By setting $x = 1$ we can see that $\mathrm{core}_G(H) \subseteq H$. Also, since $\mathrm{core}_G(H)$ is the kernel of the induced homomorphism $G \to S_{G/H}$, we know that $\mathrm{core}_G(H)$ is a normal subgroup.

Next suppose $N < H$ and $N \lhd G$. Since $N \subseteq H$, we know that $xNx^{-1} \subseteq xHx^{-1}$ for all $x \in G$. And then since $N \lhd G$, we know that $xNx^{-1} = N$ for all $x \in G$. So, $N \subseteq xHx^{-1}$ for all $x \in G$. And this proves that $N \subseteq \bigcap_{x \in G} xHx^{-1} = \mathrm{core}_G(H)$. ■

Also note that $H \lhd G$ if and only if $\mathrm{core}_G(H) = H$. With that we're now ready to prove the following proposition...

Proposition: If $H < G$ and $[G : H] = p$ where p is the smallest prime factor of $|G|$, then $H \lhd G$.

Proof:
Let $\phi : G \to S_p$ be the induced group homomorphism of the left translation action of G on G/H. Then note that $|\phi(G)|$ divides $\gcd(|G|, |S_p|) = \gcd(|G|, p!)$. But since p is the smallest prime factor dividing $|G|$, we have that $\gcd(|G|, p!) = p$. So $|\phi(G)| \mid p$. Also note that $|\phi(G)| > 1$. After all, for any $g \notin H$ we have that $g \cdot 1H \neq 1H$ and thus g doesn't correspond to the identity permutation in S_p. So, $|\phi(G)| = p$.

But by Lagrange's theorem, we have that $[G : \mathrm{core}_G(H)] = |\phi(G)| = p$. And since $\mathrm{core}_G(H) \subseteq H$ and $[G : H] = p$, the only way this is possible is if $H = \mathrm{core}_G(H)$. So $H \lhd G$. ■

Theorem: If G is a finite group, $P < G$ is a p-group, and p divides $[G : P]$, then p divides $[N_G(P) : P]$.

Proof:
$P \curvearrowright G/P$ by left-translations. Thus $|G/P| \equiv |(G/P)^P| \pmod{p}$. But note that:

$$\begin{aligned} xP \in (G/P)^P &\iff \forall g \in P,\ \ g \cdot xP = xP \\ &\iff \forall g \in P,\ \ g \in xPx^{-1} \iff P \subseteq xPx^{-1} \end{aligned}$$

And since $|P| = |xPx^{-1}| < \infty$, we have that $P \subseteq xPx^{-1} \iff P = xPx^{-1}$. And that happens iff $x \in N_G(P)$. So $|G/P| \equiv |N_G(P)/P| \pmod{p}$.

Next since p divides $[G : P]$, we know that $0 \equiv |N_G(P)/P| \pmod{p}$. ■

Corollary: If G is a finite p-group and $H \subsetneq G$ then $H \subsetneq N_G(H)$.

Proof:
If $|G| = p^n$ and $H \lneq G$ then H is a finite p-group and p divides $[G : H]$. Then by the last theorem we have that p divides $[N_G(H) : H]$. So, $H \lneq N_G(H)$. ■

Sylow's 1st. Theorem: Suppose p^n divides $|G|$. Then there exists subgroups $P_1 < P_2 < \cdots < P_n < G$ such that $|P_i| = p^i$ for all $1 \leq i \leq n$.

Proof:
We proceed by induction on k.

The base case follows from Cauchy's theorem. Meanwhile suppose p^{k+1} divides $|G|$ and we've already found subgroups $P_1 < \cdots < P_k < G$ such that $|P_i| = p^i$ for all $1 \leq i \leq k$. Then p divides $[G : P_k]$, which means by two theorems ago that p divides $[N_G(P_k) : P_k]$. And since $P_k \lhd N_G(P_k)$ we have that $N_G(P_k)/P_k$ is a well-defined group with $p \mid |N_G(P_k)/P_k|$.

By Cauchy's theorem and the correspondance theorem (the latter is in my math 100a paper notes), we thus know there is a subgroup $P_k < P_{k+1} < N_G(P_k)$ such that P_{k+1}/P_k is a cyclic group of order p. And it follows that $|P_{k+1}| = |P_k| \cdot p = p^{k+1}$. ■

We say that $\nu_p(|G|) = k$ if $p^k \mid |G|$ and $p^{k+1} \nmid |G|$, and we call $\nu_p(|G|)$ the p-valuation of G.

$P < G$ is called a Sylow p-subgroup if $|P| = p^{\nu_p(|G|)}$. Note that P is a Sylow p-subgroup if and only if P is a finite p-group and $p \nmid [G : P]$.

Also, we let $\mathrm{Syl}_p(G)$ be the set of all Sylow p-subgroups of G.

I'll continue with this class on *page 297*.

10/7/2025

Complex Analysis Homework Assignment 2:

Right now the class is still being boring and not really covering anything new. But I do still need to do homework for the class. So…

Exercise II.3.1: Prove the following:

(a) A set A is closed iff it contains all its limits points.

(b) If $A \subseteq X$, then $\overline{A} = A \cup \{x : x \text{ is a limit point of A}\}$.

Ok actually frick whoever assigned this stupid problem. I'm not fricking type-setting problems which are at the level of a first quarter undergrad analysis class. When will this professor stop wasting everyone's time and actually get to the content that we're all paying out of our asses to learn? Hell! Math 200 and Math 240 started covering new content on the first day of class. This class is actual theft. Anyways here are my notes from math 240b:

A point x is called an accumulation point of A if $A \cap (U - \{x\}) \neq \emptyset$ for every neighborhood U of x.

Proposition 4.1: If $A \subseteq X$, let $acc(A)$ be the set of accumulation points of A. Then $\overline{A} = A \cup acc(A)$, and A is closed iff $acc(A) \subseteq A$.

Proof: If $x \notin \overline{A}$, then A^c is a neighborhood of x which doesn't intersect A. So $x \notin acc(A)$.

because $x \in (\overline{A})^c = (A^c)^\circ$.

Thus $A \cup acc(A) \subseteq \overline{A}$.

If $x \notin A \cup acc(A)$, there is an open U containing x such that $U \cap A = \emptyset$. So $\overline{A} \subseteq U^c$ and $x \notin \overline{A}$. So $\overline{A} \subseteq A \cup acc(A)$.

Finally A is closed $\Longleftrightarrow A = \overline{A} \Longleftrightarrow acc(A) \subseteq A$. ■

Proposition 4.18: If X is a topological space, $E \subseteq X$, and $x \in X$, then x is an accumulation point of E iff there is a net in $E - \{x\}$ that converges to x, and $x \in \overline{E}$ iff there is a net in E that converges to x.

Proof: If x is an accumulation point of E, let $\mathcal{N}$ be the set of neighborhoods of x directed by reverse inclusion. For each $U \in \mathcal{N}$, pick $x_U \in (U - \{x\}) \cap E$. Then $x_U \to x$.

Conversely, if $x_\alpha \in E - \{x\}$ for all $\alpha \in A$ (a directed set) and $x_\alpha \to x$, then every punctured neighborhood of x (i.e. $U - \{x\}$) contains some x_α. So $E \cap (U - \{x\}) \neq \emptyset$ for all neighborhoods U of x. Hence, x is an accumulation point of E.

Likewise, if $x_\alpha \to x$ where $x_\alpha \in E$ for all α, then $x \in \overline{E}$.

The converse follows from the fact that $\overline{E} = E \cup acc(E)$, we already showed how to find $\langle x_\alpha \rangle$ with $x_\alpha \to x$ if $x \in acc(E)$, and the case where $x \in E$ is trivial. (Choose the constant net $x_\alpha = x$ for all α.)

Exercise II.3.4: Let $z_n, z \in \mathbb{C}$ and let d be the metric on $\mathbb{C}_\infty$. Then $|z_n - z| \to 0$ iff $d(z_n - z) \to 0$.

Recall that $d(z_n, z) = \frac{2|z_n - z|}{((|z_n|^2+1)(|z|^2+1))^{1/2}}$. Hence $d(x_n, z) \leq 2|z_n - z|$ and we have that $|z_n - z| \to 0 \Longrightarrow d(z_n - z) \to 0$.

On the other hand, if $\phi : \mathbb{C}_\infty \to S^2$ is the projection of the extended complex plane to the Riemann sphere, then by definition $d(z_n, z) \to 0$ iff $\phi(z_n) \to \phi(z)$ in $\mathbb{R}^3$. And since $\phi(z) \neq (0, 0, 1)$, there must exist some $-1 < C < 1$ and $N \in \mathbb{N}$ such that the third coordinates of $\phi(z)$ and $\phi(z_n)$ are less than C for all $n \geq N$. This translates to saying that:

$$\frac{|z|^2-1}{|z|^2+1} \text{ and } \frac{|z_n|^2-1}{|z_n|^2+1} \text{ are less than } C \text{ for all } n \geq N.$$

By quadratic formula this translate to saying that $|z_n|$ and $|z|$ are less than $D := \frac{\sqrt{1-C^2}}{1-C}$ when $n \geq N$ So finally we may say that when $n \geq N$:

$$|z_n - z| = \tfrac{1}{2} d(z_n, z)((|z_n|^2 + 1)(|z|^2 + 1))^{1/2} < \tfrac{1}{2} d(z_n, z)((D^2 + 1)(D^2 + 1))^{1/2}$$

And now it's clear that $d(z_n - z) \to 0$ implies that $|z_n - z| \to 0$ as $n \to \infty$.

Also show that if $|z_n| \to \infty$, then $\{z_n\}_{n \in \mathbb{N}}$ is Cauchy in $\mathbb{C}_\infty$.

To prove Cauchy-ness, it suffices to show $\{z_n\}_{n \in \mathbb{N}}$ converges to ∞. Fortunately, note that $d(z_n, \infty) - \frac{2}{(|z_n|^2+1)^{1/2}} \to 0$ as $|z_n| \to \infty$. So $z_n \to \infty$. ■

Exercise II.3.8: Suppose $\{x_n\}_{n \in \mathbb{N}}$ is a Cauchy sequence in a metric space and $\{x_{n_k}\}_{k \in \mathbb{N}}$ is a convergent subsequence. Then $\{x_n\}$ is convergent.

Once again frick whoever is wasting all of our time with these stupid problems. Literally why I am wasting thousands of dollars to relearn content that any person getting into math grad school should already know.

Let x be the limit of $\{x_{n_k}\}_{n \in \mathbb{N}}$ and pick any $\varepsilon > 0$.

Since $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy there exists some $N \in \mathbb{N}$ such that $d(x_n, x_m) < {}^{\varepsilon}/2$ for all $n, m \geq N$. Then because $x_{n_k} \to x$, we know there exists some $k \in \mathbb{N}$ such that $d(x_{n_k}, x) < {}^{\varepsilon}/2$ and $n_k \geq N$. Thus, $d(x_n, x) \leq d(x_n, x_{n_k}) + d(x_{n_k}, x) < \varepsilon$ for all $n \geq N$. And this proves that $x_n \to x$ as $n \to \infty$. ■

Exercise II.4.1: Prove that if every collection $\mathcal{F}$ of closed subsets of K with the finite intersection property has that $\bigcap_{F\in\mathcal{F}} F \neq \emptyset$, then K is compact.

Suppose K is not compact and let $\{U_\alpha\}_{\alpha\in A}$ be an open cover of K that has no finite subcover. Then upon setting $F_\alpha = U_\alpha^{\mathsf{C}}$ for all $\alpha \in A$ we have that $\{F_\alpha\}_{\alpha\in A}$ is a collection of closed sets having the finite intersection property. After all, if there were some $F_{\alpha_1}, \ldots, F_{\alpha_n}$ such that $\bigcap_{j=1}^n F_{\alpha_j} = \emptyset$ then that would imply that $(\bigcap_{j=1}^n F_{\alpha_j})^{\mathsf{C}} = \bigcup_{j=1}^n U_{\alpha_j} = K$. But that contradicts that $\{U_\alpha\}_{\alpha\in A}$ has no finite subcover of K.

At the same time though, $\bigcap_{\alpha\in A} F_\alpha = (\bigcup_{\alpha\in A} U_\alpha)^{\mathsf{C}} = K^{\mathsf{C}} = \emptyset$. So, we have found a collection of closed sets with the finite intersection property whose intersection is empty. ■

Exercise II.4.4: Show that the union of a finite number of compact sets is compact.

Let $K_1, \ldots, K_n$ be a finite collection of compact sets and let K be their union. Now if $\{U_\alpha\}_{\alpha\in A}$ is any open cover of K, then $\{U_\alpha\}_{\alpha\in A}$ is also an open cover of each K_i. So for each i we can pick finitely many U_{α_j} covering each K_i. And by unioning the finitely many finite subcovers we get another finite subcover of all of K. ■

Exercise II.4.6: Show that the closure of a totally bounded set is totally bounded.

Let X be our metric space and suppose that $E \subseteq X$ is totally bounded. Then for any $\varepsilon > 0$ there are finitely many balls $B_1, \ldots, B_n$ of radius ${}^{\varepsilon}/2$ covering E. And in turn the union of the closures of those finitely many balls is a closed set containing E and thus also $\overline{E}$. Finally, by expanding each of our balls to have radius ε we now have a finite collection of balls of radius ε covering $\overline{E}$.

Note to cover my bases:
If $B_{\varepsilon/2}(x) := \{y \in X : d(x,y) < \varepsilon/2\}$, then $\overline{B_{\varepsilon/2}(x)} \subseteq \{y \in X : d(x,y) \leq \varepsilon/2\}$ since the latter is easily checked to be closed set (it's complement is open by triangle inequality) which contains $B_{\varepsilon/2}(x)$. Also, $\{y \in X : d(x,y) \leq \varepsilon/2\} \subseteq \{y \in X : d(x,y) < \varepsilon\}$. ■

Exercise II.5.5: Suppose $f : X \to \Omega$ is uniformly continuous (where (Ω, ρ) and (X, d) are metric spaces). If $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in X then $\{f(x_n)\}_{n\in\mathbb{N}}$ is a Cauchy sequence in Ω.

Proof:
For any $\varepsilon > 0$ pick $\delta > 0$ such that $\rho(f(x), f(y)) < \varepsilon$ whenever $d(x, y) < \varepsilon$. Since $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence there exists some $N \in \mathbb{N}$ such that $d(x_n, x_m) < \delta$ for all $n, m \geq N$. And in turn $\rho(f(x_n), f(x_m)) < \varepsilon$ for all $n, m \geq N$. This proves that $\{f(x_n)\}_{n\in\mathbb{N}}$ is a Cauchy sequence in Ω.

Does the prior statement still hold if f is only assumed to be continuous but not uniformly continuous.

No. For example, consider the sequence $\{\frac{1}{n}\}_{n\in\mathbb{N}}$ in $(0, \infty)$. This sequence is Cauchy since for every $\varepsilon > 0$ we may choose some N such that $\frac{1}{N} < \varepsilon$. And then for all $n, m \geq N$ we have that $|\frac{1}{n} - \frac{1}{m}| < \max(\frac{1}{n}, \frac{1}{m}) < \frac{1}{N} < \varepsilon$.

Next let $f(x) = \frac{1}{x}$. To show that f is continuous on $(0, \infty)$, note that both 1 and x are continuous on $(0, \infty)$. After all, just set $\delta = \varepsilon$ when doing the ε-δ-proof. And since $x \neq 0$ on $(0, \infty)$, we have by proposition 5.5 in Conway that $1/x$ is continuous on $(0, \infty)$.

But now $\{\frac{1}{(1/n)}\}_{n \in \mathbb{N}}$ is not Cauchy. After all, for all $n, m \in \mathbb{N}$ with $n \neq m$ we have that:

$$|\frac{1}{(1/n)} - \frac{1}{(1/m)}| = |n - m| \geq 1. \blacksquare$$

That's all my my math 220a homework for this week. Yay I'm done (finally). God I hope this class stops being a giant waste of money at some point. Anyways the next notes for this class will be on *page 294*.

My next order of business for tonight is to continue taking notes for the reading group tomorrow. I'm picking up from *here*.

Let (Y, T) be a dynamical system, K be a metrized compact group, and $\psi : Y \to K$ be a continuous mapping. Then define $X = Y \times K$ and $T' : X \to X$ by $T'(y, k) = (Ty, \psi(y)k)$. The resulting system is called a group extension or skew product of (Y, T) with K.

Sanity check:

1. T' is a continuous map. After all, it's continuous iff its two output coordinates are individually continuous. And since we're already requiring T and ψ to be continuous and we know $(k_1, k_2) \mapsto k_1 k_2$ is continuous from $K \times K$ to K, we can easily show both coordinates are continuous.

2. If we define $\phi : X \to Y$ by $\phi(y, k) = y$, then ϕ is continuous (this is just a defining property of product topologies) and:

   $$\phi(T'(y, k)) = \phi(Ty, \psi(y)k) = Ty = T\phi(y, k).$$

   So, ϕ defines a homomorphism of (X, T') to (Y, T) and we've shown that (X, T') does extend (Y, T) by our prior definition.

Note that in our above system, $X \curvearrowleft K$ by right translation. Specifically, for each $k' \in K$ there exists a homeomorphism $R_{k'} : X \to X$ given by $R_{k'}(y, k) = (y, kk')$. Also, R_k commutes with T' since:

$$R_{k'}(T'(y, k)) = R_{k'}(Ty, \phi(y)k) = (Ty, \phi(y)kk') = T'(y, kk') = T'(R_{k'}(y, k))$$

Thus, each $R_{k'}$ is an automorphism of the dynamical system (X, T').

Side note:
I think the author's choice to use right translations instead of left translations is really weird because it makes it so $R_{k_2}(R_{k_1}(y, k)) = (y, kk_1k_2) = R_{k_1k_2}(y, k)$. To fix this I would have instead set $T'(y, k) = (Ty, k\psi(y))$ when defining the group extension, and then I'd have made $K \curvearrowright X$ by left translation. That way, we still have that (X, T') is a dynamical system, ϕ is still a homomorphism from (X, T') to (Y, T), but now additionally we have that $R_{k_2} \circ R_{k_1} = R_{k_2k_1}$. Unfortunately, I don't want to confuse other presenters later on so I'll stick to the conventions of the book. :(

Also just to be clear: I shall denote $X \curvearrowleft G$ instead of $G \curvearrowright X$ when G is acting from the right on X instead of from the left.

Before moving on to the next theorem, I need to clear up some confusion and write a few theorems.

Firstly, recall on page 268 when I said a point x is recurrent for (X, T) if for any neighborhood V of x there exists $n \geq 1$ with $T^n x \in V$. And then I said that that is equivalent to saying that there is some increasing sequence $(n_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}$ such that $T^{n_k} x \to x$. Why are these two definitions equivalent?

$(\Longleftarrow)$
This direction is obvious.

$(\Longrightarrow)$
If $T^n(x) = x$ for some n, then we know that $T^{kn}(x) = x$ for all $k \in \mathbb{N}$ and then we trivially have that $T^{kn}(x) \to x$ as $k \to \infty$. Meanwhile, suppose $T^n(x) \neq x$ for any $n \in \mathbb{N}$ and let $\{U_k\}_{k\in\mathbb{N}}$ be a countable decreasing neighborhood base for $x \in X$. Then we may construct a subsequence $\{T^{n_k}\}_{k\in\mathbb{N}}$ converging to k as follows:

To start off, just pick any $n_1 \in \mathbb{N}$ such that $T^{n_1} x \in U_1$. Then we proceed by recursive definition.

Suppose we have already chosen $n_1 < \cdots < n_k$ in $\mathbb{N}$ and that $T^{n_i} x \in V_i$ for all $1 \leq i \leq k$. If X is Hausdorff (which it will always be if X is a metric space), then we can find pairs of disjoint open sets $V_j, W_j \subseteq X$ for all $1 \leq j \leq n_k$ such that $x \in V_j$ and $T^j x \in W_j$. And by setting $V := \bigcap_{j=1}^{n_k} V_j$ we have that V is an open set containing x such that $T^j x \notin V$ for all $1 \leq j \leq n_k$.

Now pick $m \geq k+1$ such that $U_m \subseteq V$. Then by assumption, there is some $n_{k+1} \in \mathbb{N}$ such that $T^{n_{k+1}} x \in U_m$. And also we must have that $n_{k+1} > n_k$. ■

More generally, given a <u>dynamical system</u> (X, T) and any $x \in X$, define the <u>forward orbit closure</u> of x as $Q(x) := \overline{\{T^n x : n \geq 1\}}$. Note that x is recurrent for (X, T) iff $x \in Q(x)$. Also, by slightly modified reasoning to before, we know $z \in Q(x)$ iff $z = T^n x$ for some n or either there is some increasing sequence $\{n_k\}_{k\in\mathbb{N}}$ in $\mathbb{N}$ such that $T^{n_k} \to z$.

<u>Lemma A:</u> Let (X, T) be a dynamical system and suppose that $y \in Q(x)$ and $z \in Q(y)$. Then $z \in Q(x)$.

Proof:
It suffices to show $Q(y) \subseteq Q(x)$. Fortunately, if $y = T^n x$ for some $n \in \mathbb{N}$, then it's obvious that $Q(y) \subseteq Q(x)$. So we can now assume without loss of generality that there is some sequence $T^{n_i} x \to y$ as $i \to \infty$.

Next if $z = T^m y$ for some $m \in \mathbb{N}$, then it's clear that $T^m(T^{n_i} x) = T^{m+n_i} x \to z$ as $i \to \infty$, thus showing that $z \in Q(x)$. So, we may assume now that there is a sequence $T^{n_j} y \to z$ as $j \to \infty$.

Let $\{U_k\}_{k\in\mathbb{N}}$ be a countable decreasing neighborhood base for $z \in X$. We can construct a sequence $\{T^{n_k} x\}_{k\in\mathbb{N}}$ converging to 0 as follows:

We'll want to keep track of subsequences of $\{n_{j(k)}\}_{k\in\mathbb{N}}$ and $\{n_{i(k)}\}_{n\in\mathbb{N}}$ of $\{n_j\}_{j\in\mathbb{N}}$ and $\{n_k\}_k \in \mathbb{N}$ respectively as we are constructing our actual desired sequence.

For any k pick $j(k) \in \mathbb{N}$ such that $T^{n_{j(k)}}y \in U_k$ and $j(k) > j(k-1)$ (if $k > 1$). Then set $V = (T^{n_j})^{-1}(U_k)$. Because V is an open neighborhood of y, we know that there is some $I \in \mathbb{N}$ such that $T^{n_i}x \in V$ whenever $i \geq I$. So pick $i(k)$ such that $T^{n_{i(k)}}x \in V$ and $i(k) > i(k-1)$ (if $k > 1$). Now $T^{n_{j(k)}+n_{i(k)}}x \in U_k$.

Doing this for all $K \in \mathbb{N}$, we get that $T^{n_{j(k)}+n_{i(k)}}x \to z$ as $k \to \infty$ and that $\{n_{j(k)} + n_{i(k)}\}_{k\in\mathbb{N}}$ is increasing. ■

Lemma B: Suppose (X,T) is a dynamical system and R is an automorphism of (X,T) (meaning $R \circ T = T \circ R$ and $R : X \to X$ is continuous). Then for any $x \in X$ we have that $R(Q(x)) \subseteq Q(Rx)$.

Proof:
Suppose $z \in Q(x)$. If $z = T^n x$ for some n, then $R(z) = R(T^n x) = T^n(Rx) \in Q(Rx)$. Otherwise, let $T^{n_k}x \to z$. Then $R(z) = \lim_{k\to\infty} R(T^{n_k}x) = \lim_{k\to\infty} T^{n_k}(Rx)$. So $R(z) \in Q(Rx)$ and we've shown that $R(Q(x)) \subseteq Q(Rx)$. ■

Lemma C: Suppose K is a compact metrized group and $k \in K$. Then the sequence $\{k^n\}_{n\in\mathbb{N}}$ has the identity e of K as a subsequential limit.

Proof:
We know that (K,T) where $T(k') = k'k$ is a Kronecker system. So $e \in K$ is recurrent and we know there is some increasing sequence $(n_i)_{i\in\mathbb{N}}$ of integers with $T^{n_i}e = k^{n_i-1} \to e$ as $i \to \infty$. ■

I think before next week I will try to read up on topological groups some more cause I feel like I have a knowledge gap right now.

Theorem 1.4: If $y_0 \in Y$ is recurrent for (Y,T) and (X,T') is a skew extension of (Y,T), then (y_0,k_0) is a recurrent point of (X,T') for all $k_0 \in K$.

Proof:
Let e denote the identity of K. We shall show (y_0,e) is a recurrent point of (X,T) since it then follows from *proposition 1.3:* that each $R_{k_0}(y_0,e) = (y_0,k_0)$ is recurrent.

For any $x \in X$ let $Q(x) = \overline{\{(T')^n x : n \geq 1\}}$ denote the forward orbit closure of x. Then x is recurrent for (X,T') iff $x \in Q(x)$. Now since y_0 is recurrent for (Y,T), there is some $k_1 \in K$ such that: $(y_0,k_1) \in Q(y_0,e)$.

Why (since Furstenberg doesn't explain this)?

Note that $(T')^n(y_0,e) = (T^n y_0, k^{(n)})$ where:

$$k^{(n)} = \phi(T^{n-1}y_0)\phi(T^{n-2}y_0)\cdots\phi(Ty_0)\phi(y_0).$$

Now because y_0 is recurrent for (Y,T), we know that there is a subsequence $\{T^{n_i}y\}_{i\in\mathbb{N}}$ such that $T^{n_i}y \to y$ as $i \to \infty$. And since K is compact, we have that $\{k^{n_i}\}_{i\in\mathbb{N}}$ has a subsequential limit point k_1. And finally, by passing to another subsequence $\{n_{i_j}\}_{j\in\mathbb{N}}$ we get that $(T')^{n_{i_j}}(y_0,e) \to (y_0,k_1)$ as $j \to \infty$.

But now note by induction that $(y_0, k_1^n) \in Q(y_0, e)$ for all $n \in \mathbb{N}$.

Why?

We know $(y_0, k_1) \in Q(y_0, e)$. Now by induction assume that $(y_0, k_1^n) \in Q(y_0, e)$. Then $(y_0, k_1^{n+1}) = R_{k_1}(y_0, k_1^n) \in Q(R(y_0, e)) = Q(y_0, k_1)$ by lemma B. But we also have that $(y_0, k_1) \in Q(y_0, e)$. So $(y_0, k_1^{n+1}) \in Q(y_0, e)$ by lemma A.

In turn by lemma C we have that there is an increasing subsequence $\{n_j\}_{j \in \mathbb{N}}$ such that $(y_0, k^{n_j}) \to (y_0, e)$ as $j \to \infty$. And since $Q(y_0, e)$ is closed, we thus have shown that $(y_0, e) \in Q(y_0, e)$. ■

By using the prior theorems we can inductively obtain examples of non-Kronecker dynamical systems where every point is recurrent.

For example, let $T : \mathbb{T}^2 \to \mathbb{T}^2$ be given by $T(\theta, \phi) = (\theta + a, \phi + 2\theta + a)$. Then $(\mathbb{T}^2, T)$ is a group extension of the Kronecker system $(\mathbb{T}, T')$ where $T'\theta = \theta + a$ and $\psi(\theta) = 2\theta + a$. By theorems 1.2. and 1.4 we know that every point in $(\mathbb{T}^2, T)$ is recurrent.
In the prior system note that the orbit of $(0, 0) \in \mathbb{T}^2$ is:

$$\begin{aligned} &(0,0) \to (a,a) \to (2a,4a) \to \cdots \to (na, n^2a) \\ &\qquad \to (na + a, n^2a + 2na + a) = ((n+1)a, (n^2 + 2n + 1)a) \\ &\qquad = ((n+1)a, (n+1)^2a) \to \cdots \end{aligned}$$

This leads to the following proposition:

Proposition 1.5: For any $\alpha \in \mathbb{R}$ and $\varepsilon > 0$ we can solve the diophantine inequality $|\alpha n^2 - m| < \varepsilon$ (where $n > 0$).

Proof:

We know there is some increasing sequence $\{n_k\}_{k \in \mathbb{N}}$ in $\mathbb{N}$ such that $(n_k a, n_k^2 a) \to (0, 0) \in \mathbb{T}^2$. In particular, since $n_k^2 a \to 0 \in \mathbb{T}$ we know that if k is fixed large enough, then there exists $m \in \mathbb{N}$ such that $|n_k^2 - m| < \varepsilon$. ■

We can extend this to higher degree polynomials too. Let $p(x)$ be a polynomial of degree d with real coefficients and write $p_d(x) := p(x)$ as well as:

$$p_{d-1}(x) := p_d(x+1) - p_d(x), \quad p_{d-2}(x) := p_{d-1}(x+1) - p_{d-1}(x), \text{ etc.}$$

Each $p_i(x)$ is of degree i. Let α be the constant $p_0(x)$. Then define a transformation $T : \mathbb{T}^d \to \mathbb{T}^d$ by $T(\theta_1, \theta_2, \theta_3 \ldots, \theta_d) := (\theta_1 + \alpha, \theta_2 + \theta_1, \theta_3 + \theta_2, \ldots, \theta_d - \theta_{d-1})$.

By induction, we can see that this is a group extension of a dynamical system on the $(d-1)$-torus which in turn is a group extension of a dynamical system on the $(d-2)$-torus, and so on and so forth down to the 1-torus. So we can conclude that each point in $\mathbb{T}^d$ is recurrent.

I'll do a step of the induction cause why not.

Suppose we've shown that $(\mathbb{T}^k, T_k)$ is a dynamical system where:

$$T_k(\theta_1, \theta_2, \ldots, \theta_k) = (\theta_1 + \alpha, \theta_2 + \theta_1, \theta_3 + \theta_2, \ldots, \theta_k - \theta_{k-1}).$$

Then by letting $\psi : \mathbb{T}^k \to \mathbb{T}$ be given by $\psi(\theta_1, \theta_2, \ldots, \theta_k) = \theta_k$ we have that $(\mathbb{T}^k, T_{k+1})$ is a group extension where:

$$\begin{aligned} T_{k+1}(\theta_1, \ldots, \theta_{k+1}) &= (T_k(\theta_1, \ldots, \theta_k), \theta_{k+1} + \psi(\theta_1, \ldots, \theta_k)) \\ &= ((\theta_1 + \alpha, \theta_2 + \theta_1, \ldots, \theta_k + \theta_{k-1}), \theta_{k+1} + \theta_k). \end{aligned}$$

Now compute the orbit of $(p_1(0), \ldots, p_d(0))$. Since $p_{i-1}(n) + p_i(n) = p_i(n+1)$ we find that $T(p_1(n), p_2(n), \ldots, p_d(n)) = (p_1(n+1), p_2(n+1), \ldots, p_d(n+1))$.

And thus $T^n(p_1(0), p_2(0), \ldots, p_d(0)) = (p_1(n), p_2(n), \ldots, p_d(n))$.

We conclude that $p_d(n) = p(n)$ must get arbitrarily close to $p(0)$ modulo 1. Hence the following theorem:

> Theorem 1.6: If $p(x)$ is any real polynomial with $p(0) = 0$, then for any $\varepsilon > 0$ we can solve the diophantine inequality $|p(n) - m| < \varepsilon$ (where $n > 0$).

I'll study more related to this reading group on *page* ___.

10/8/2025

Math 200a Homework:

Set 2 Problem 1: Suppose G is a simple group (meaning the only normal subgroups of G are $\{1\}$ and G), and it has a subgroup H of index n where n is an integer more than 1. Prove that G can be embedded into the symmetric group S_n.

> Consider the action $G \curvearrowright G/H$ by left translation, and let $\phi : G \to S_{G/H}$ be the group homomorphism induced by this action. Note that since $[G : H] = n$, we know $S_{G/H} \cong S_n$. So, as long as we can prove that ϕ is injective then we will be done.
>
> Suppose $x \in \ker(\phi)$. Then we know that $xgH = gH$ for all $g \in G$. But that's true iff $x \in gHg^{-1}$ for all $g \in G$. Or in other words, we must have that $x \in \mathrm{core}_G(H)$. This proves that $\{1\} < \ker(\phi) < \mathrm{core}_G(H)$. Next note that because $H \neq G$ (which we know since $n > 1$), $\mathrm{core}_G(H) \lhd G$, $\mathrm{core}_G(H) < H$, and G is simple, we must have that $\mathrm{core}_G(H) = \{1\}$. So, we've shown that $\ker(\phi) = \{1\}$. And this proves that ϕ is injective. ■

Set 2 Problem 7: Suppose N is a finite cyclic normal subgroup of G. Prove that every subgroup of N is normal in G.

> Let $N = \langle g_0 \rangle$ and then consider any integer $k \in \mathbb{Z}$. Our goal is to show that for any $x \in G$ we have that $xg_0^k x^{-1}$ equals some power of g_0^k. After all, if $H < N$ and $g_0^k \in H$, then we must have that $(g_0^k)^m \in H$ for all $m \in \mathbb{Z}$. And since every element of H will have the form g_0^k where k is some integer, this will have proved that every subgroup of N is normal.
>
> > Consider any $x \in G$ and let $r \in \mathbb{Z}$ be such that $xg_0x^{-1} = g_0^r$. Note that such an r exists because N is normal and cyclic. Then $xg_0^k x^{-1} = (xg_0x^{-1})^k = (g_0^r)^k = (g_0^k)^r$. ■

Math 241a (lecture 3-5):

If $\mathcal{X}$ is a general topological K-vector space, then the weak* topology on $\mathcal{X}^*$ is defined as the weak topology generated by the linear functionals $f \mapsto f(x)$ as we range over all $x \in \mathcal{X}$. Like before on *page 252*, we can reframe this definition as being of a topology generated by the family of seminorms $\{p_x(f) := |f(x)| : x \in \mathcal{X}\}$.

Note that this family will always be sufficient because if $p_x(f) = 0$ for all $x \in \mathcal{X}$, then we must have that $f = 0$. So the weak* topology is always Hausdorff.

Also, note that since $\mathcal{X}$ is not assumed to have a norm, there is no induced operator norm topology on $\mathcal{X}^*$. Hence, until after we define an actual topology on $\mathcal{X}^*$ it doesn't make sense to talk about whether a linear functional on $\mathcal{X}^*$ is continuous.

(Note: in this class we use $M(X)$ to refer to the collection of probability measures on X where X is a compact metric space. I forgot to prove it until *page 439* but all $\mu \in M(X)$ are Radon).

Corollary 1.1.29: Let X be a compact metric space. Then $M(X)$ is compact in the weak* topology of $C(X)^*$.

Proof:
We already know by Alaoglu's theorem (see my math 240b notes) that the ball $B = \{\lambda \in C(X)^* : \|\lambda\|_{\mathrm{op}} \leq 1\}$ is compact in the weak* topology on $C(X)^*$. Also, by the Riesz representation theorem we know that $M(X) \subseteq B$. And just before we noted that the weak* topology is Hausdorff. Hence B is closed. So, if we can show that $M(X)$ is a closed subset of $C(X)^*$, then we will have successfully proven that $M(X)$ is compact.

Fortunately, note that $M(X) = \{\lambda \in C(X)^* : \lambda(1) = 1 \text{ and } \lambda(f) \geq 0 \text{ when } f \geq 0\}$. Now we know that $\{\lambda \in C(X)^* : \lambda(f) \geq 0\}$ is closed in the weak* topology for all $f \in C(X)$. Also, we know that $\{\lambda \in C(X)^* : \lambda(1) = 1\}$ is closed in the weak* topology. So, $M(X) = \bigcap_{f \geq 0} \{\lambda \in C(X)^* : \lambda(f) \geq 0\} \cap \{\lambda \in C(X)^* : \lambda(1) = 1\}$ is closed. ■

Theorem: Suppose F is any field, $\mathcal{X}, \mathcal{Y}$ are F-vector spaces (which are allowed to be infinite dimensional), that $\mathcal{M} \subseteq \mathcal{X}$ is a vector subspace, and $f : \mathcal{M} \to \mathcal{Y}$ is a linear map. Then there exists a linear map $g : \mathcal{X} \to \mathcal{Y}$ such that $g|_{\mathcal{M}} = f$.

Proof:
Let $\mathcal{F}$ be the collection of linear maps $h : W \to \mathcal{Y}$ where $W \subseteq \mathcal{X}$ is a vector subspace containing $\mathcal{M}$ and $h|_{\mathcal{M}} = f$. Then $\mathcal{F}$ when considered as a collection of sets in $\mathcal{X} \times \mathcal{Y}$ is partially ordered by set inclusion. Furthermore, given a simply ordered subset $\{h_\alpha\}_{\alpha \in A}$ of $\mathcal{F}$ we can easily see that $\bigcup_{\alpha \in A} h_\alpha$ is a well defined linear map in $\mathcal{F}$. Hence, by Zorn's lemma we can conclude that there is a maximal linear map $g \in \mathcal{F}$ extending f.

If V is the domain of g, we claim that $V = \mathcal{X}$. After all, suppose there exists $x_0 \in \mathcal{X} - V$. Then $V + Fx_0 := \{y + cx_0 : y \in V, c \in F\}$ is another vector subspace of $\mathcal{X}$ which satisfies that if $y' + c'x_0 = y + cx_0$ then we have that $y = y'$ and $c = c'$. And so, we may define the linear map $h : V + Fx_0 \to \mathcal{Y}$ by $h(y + cx_0) = g(y)$. But now $h \in \mathcal{F}$ and $g \subset h$, thus contradicting the maximality of g. ■

Theorem: Suppose F is any field, $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are F-vector spaces (which are allowed to be infinite dimensional), and that $\varphi : \mathcal{X} \to \mathcal{Y}$ and $f : \mathcal{X} \to \mathcal{Z}$ are linear maps such that $f(x) = 0 \implies \varphi(x) = 0$. Then there exists a linear map $\tilde{\varphi} : \mathcal{Z} \to \mathcal{Y}$ such that $\tilde{\varphi} \circ f = \varphi$. Also $\tilde{\varphi}$ is unique if f is surjective.

Proof:
To start off, for all $z \in f(\mathcal{X})$ define $\tilde{\varphi} : \mathcal{Z} \to \mathcal{Y}$ by $g(z) = \varphi(x)$ where $x \in f^{-1}(\{z\})$. To see that $\tilde{\varphi}$ is well-defined, note that if $f(x_1) = f(x_2) = z$ then we have that $f(x_1 - x_2) = 0$. So, $\varphi(x_1 - x_2) = 0$ and in turn we have that $\varphi(x_1) = \varphi(x_2)$. Also, to see that $\tilde{\varphi}$ is linear on $f(\mathcal{X})$, suppose $c_1, c_2 \in F$ and $z_1, z_2 \in \mathcal{Z}$. Then upon letting $x_1, x_2 \in \mathcal{X}$ such that $f(x_1) = z_1$ and $f(x_2) = z_2$, we know by the linearity of f that $f(c_1x_1 + c_2x_2) = c_1z_1 + c_2z_2$. So:

$$\tilde{\varphi}(c_1z_1 + c_2z_2) = \varphi(c_1x_1 + c_2x_2) = c_1\varphi(x_1) + c_2\varphi(x_2) = c_1\tilde{\varphi}(z_1) + c_2\tilde{\varphi}(z_2).$$

Now by how we defined $\tilde{\varphi}$, we know that $\tilde{\varphi}$ is the unique map from $f(\mathcal{X})$ to $\mathcal{Y}$ such that $\tilde{\varphi} \circ f = \varphi$. And if f is surjective, we are done. Meanwhile, if f isn't surjective, then we can use the previous theorem to extend the domain of $\tilde{\varphi}$ to all of $\mathcal{Z}$. ■.

Proposition 1.1.30: Let $\mathcal{X}$ be a topological vector space whose topology is defined by a sufficient family of seminorms. Then any element of $(\mathcal{X}^*$, weak* topology$)^*$ has the form $\lambda \mapsto \lambda(x)$ where $x \in \mathcal{X}$.

Proof:
Suppose $\varphi : \mathcal{X}^* \to K$ is weak* continuous and linear. Then $\varphi^{-1}(\{t \in K : |t| < 1\})$ is open in $\mathcal{X}^*$. And since the zero functional in $\mathcal{X}^*$ is in that preimage, we thus know there are finitely many elements $x_1, \ldots, x_n \in \mathcal{X}$ and $r_1, \ldots, r_n > 0$ such that:

$$\bigcap_{i=1}^{n} \{\lambda \in \mathcal{X}^* : |\lambda(x_i)| < r_i\} \subseteq \varphi^{-1}(\{t \in K : |t| < 1\}).$$

Now if $\lambda \in \mathcal{X}^*$ with $\lambda(x_i) = 0$ for all i, then $c\lambda(x_i) = 0$ for all $c \in K$ and each i. It follows that $|\varphi(c\lambda)| < 1$ for all $c \in K$, and the only way that is possible is if $\varphi(\lambda) = 0$. To phrase this another way, let $f : \mathcal{X}^* \to K^n$ be given by $f(\lambda) \coloneqq (\lambda(x_1), \ldots, \lambda(x_n))$. Then we have that $f(\lambda) = 0 \Longrightarrow \varphi(\lambda) = 0$.

It now follows by the two prior theorems that there is a linear map $\tilde{\varphi} : K^n \to K$ such that $\tilde{\varphi} \circ f = \varphi$. And in particular, there must exist $c_1, \ldots, c_n \in K$ such that $\tilde{\varphi}(a_1, \ldots, a_n) = \sum_{i=1}^{n} c_i a_i$ for all $(a_1, \ldots, a_n) \in K^n$. So, for any $\lambda \in \mathcal{X}^*$ we have that $\varphi(\lambda) = \tilde{\varphi}(f(\lambda)) = \sum_{i=1}^{n} c_i\lambda(x_i) = \lambda(\sum_{i=1}^{n} c_i x_i)$. ■

Note that by the definition of the weak* topology we already knew each linear functional $\lambda \mapsto \lambda(x)$ is continuous. The significance of this proposition though is that those linear functionals are the *only* continuous linear functionals according to the weak* topology.

If $\mathcal{X}$ and $\mathcal{Y}$ are topological vector spaces, we shall write $B(\mathcal{X}, \mathcal{Y})$ to denote the space of continuous linear maps from $\mathcal{X}$ to $\mathcal{Y}$. If $T \in B(\mathcal{X}, \mathcal{Y})$ is bijective with a continuous inverse, then T is called an isomorphism of $\mathcal{X}$ and $\mathcal{Y}$. And if additionally $\mathcal{X} = \mathcal{Y}$, then T is called an automorphism.

We let $\mathrm{Aut}(\mathcal{X}) \subseteq B(\mathcal{X}) \coloneqq B(\mathcal{X}, \mathcal{X})$ be the set of automorphisms of $\mathcal{X}$. If $\mathcal{X}$ is a normed vector space, we let $\mathrm{Iso}(\mathcal{X}) \subseteq \mathrm{Aut}(\mathcal{X})$ be the set of isometric automorphisms.

Here are some examples of linear operators (i.e. linear maps on spaces of functions):

- (Example 1.2.1): Let (X, μ) be a semifinite measure space and consider any $\varphi \in L^\infty(X)$. For any $p \in [1, \infty]$ define $M_\varphi : L^p(X) \to L^p(X)$ by $M_\varphi f = \varphi f$. Then M_φ is a bounded linear map with $\|M_\varphi\|_{\mathrm{op}} = \|\varphi\|_\infty$.

  It's clear that $\|\varphi f\|_p \leq \|\varphi\|_\infty \|f\|_p$ for all $f \in L^p(X)$. Hence $\|M_\varphi\|_{\mathrm{op}} \leq \|\varphi\|_\infty$. On the other hand, let $A \subseteq X$ be a measurable set such that $0 < \mu(A) < \infty$ and $|\varphi(x)| \geq \|\varphi\|_\infty - \varepsilon$ for all $x \in A$. Then $\chi_A \in L^p(X)$ and:
  $$\|\varphi \chi_A\|_p \geq (\|\varphi\|_\infty - \varepsilon)\|\chi_A\|_p.$$

  In other words, the linear map $M : L^\infty(X) \to B(L^p(X))$ (given by $\varphi \mapsto M_\varphi$) is an isometry.

- (Example 1.2.4): If X is a set and $\varphi : X \to X$ is a bijection, then φ can often to be taken to define a "translation operator" T_φ on spaces of functions on X given by $(T_\varphi f)(x) = f(\varphi^{-1}(x))$.

  (a) Let X be a topological space and $\varphi \in \mathrm{Homeo}(X)$ (where $\mathrm{Homeo}(X)$ is the group of homeomorphisms of X). Then $T_\varphi : BC(X) \to BC(X)$ is an isometric isomorphism with $(T_\varphi)^{-1} = T_{\varphi^{-1}}$.

  Since φ^{-1} is continuous we know that $f \in BC(X) \implies f \circ \varphi^{-1} \in BC(X)$ as well. And T_φ is clearly an isometry since for all $f \in BC(X)$ we have that $\|T_\varphi f\|_u = \|f \circ \varphi^{-1}(x)\|_u = \|f\|_u$.To see the other claim, just note that:
  $$(T_{\varphi^{-1}} T_\varphi f)(x) = (f \circ \varphi^{-1} \circ \varphi)(x) = f(x) = (f \circ \varphi \circ \varphi^{-1})(x) = (T_\varphi T_{\varphi^{-1}} f)(x).$$

  (b) Let X be a locally compact separable metric space (which implies that X is σ-compact as I showed in my math 240c notes). Then bestow $C(X)$ with the Fréchet space topology of uniform convergence on compact sets (described on *page 251*). If $\varphi \in \mathrm{Homeo}(X)$ then $T_\varphi : C(X) \to C(X)$ is an isomorphism with inverse $T_{\varphi^{-1}}$.

I need to take a break from functional analysis to do the rest of my algebra homework. I guess you can click *here* to skip ahead to that.

Math 200a homework continued:

Set 2 Problem 2: For a group G let $\mathrm{Aut}(G)$ be the group of automorphisms of G. Let $c : G \to \mathrm{Aut}(G)$ be such that $c(g) = c_g : G \to G$ where $c_g(x) = gxg^{-1}$ for every $x \in G$.

(a) Prove that each c_g is an automorphism of G and c is a group homomorphism.

To see that c_g is a group homomorphism, just note that for any $x, y \in G$ we have that $c_g(xy) = gxyg^{-1} = gxg^{-1}gyg^{-1} = c_g(x)c_g(y)$. Also, to see c_g is injective, suppose $c_g(x) = gxg^{-1} = e$ where e is the identity on G. Then $x = g^{-1}g = e$. Finally, to prove surjectivity consider any $y \in G$ and let $x = g^{-1}yg$. Then $c_g(x) = y$. This finishes the proof that c_g is an automorphism of G.

Next to see that c is a group homomorphism, consider any $g_1, g_2 \in G$. Then we have for each $x \in G$ that $c_{g_1} \circ c_{g_2}(x) = g_1(g_2 x g_2^{-1})g_1^{-1} = (g_1g_2)x(g_1g_2)^{-1} = c_{g_1g_2}(x)$.

(b) Prove that $\ker(c) = Z(G)$ where $Z(G) := \{g \in G : \forall x \in G, gx = xg\}$ is the center of G.

$g \in \ker(c)$ iff $c_g(x) = gxg^{-1} = x$ for all $x \in G$. Or in other words, $g \in \ker(c)$ iff $gx = xg$ for all $x \in G$. This is the same as saying that $g \in Z(G)$.

(c) The image of c is called the group of inner automorphisms of G and it is denoted by $\mathrm{Inn}(G)$. Prove that $\mathrm{Inn}(G)$ is a normal subgroup of $\mathrm{Aut}(G)$.

Suppose f is any automorphism of G and $g \in G$. Then for any $x \in G$ we have that:

$$\begin{aligned}(f \circ c_g \circ f^{-1})(x) &= f(gf^{-1}(x)g^{-1}) \\ &= f(g)f(f^{-1}(x))f(g^{-1}) = f(g)x(f(g))^{-1} = c_{f(g)}(x).\end{aligned}$$

So $f \circ c_g \circ f^{-1} = c_{f(g)}$. And this shows that $\mathrm{Inn}(G) \lhd \mathrm{Aut}(G)$.

(d) Prove that $|Z(\mathrm{Aut}(G))| \leq |\mathrm{Hom}(G, Z(G))|$. In particular, if either $Z(G) = \{e\}$ or G is perfect (meaning G is equal to it's derived subgroup $[G,G]$), then $Z(\mathrm{Aut}(G)) = \{\mathrm{Id}\}$.

Suppose $\phi \in Z(\mathrm{Aut}(G))$. That means that $\phi \circ f = f \circ \phi$ for all $f \in \mathrm{Aut}(G)$. In particular, this means that $c_g \circ \phi = \phi \circ c_g$ for each $g \in G$. Also, $\phi \circ c_g = c_{\phi(g)} \circ \phi$ due to the fact that for any $x \in G$ we have $\phi(gxg^{-1}) = \phi(g)\phi(x)\phi(g)^{-1}$. So, we've shown that $c_g \circ \phi = c_{\phi(g)} \circ \phi$. And by composing ϕ^{-1} on the right side we've proven that $c_g = c_{\phi(g)}$ for all $g \in G$.

Thus we know that $gxg^{-1} = \phi(g)x\phi(g)^{-1}$ for all $g, x \in G$, and some rearranging then yields that $\phi(g) = g(xg^{-1}\phi(g)x^{-1})$ for all $x \in G$. If we define $\eta(g) := g^{-1}\phi(g)$ for all g, we then have that $\eta(g) \in Z(G)$ (since $\eta(g) = x\eta(g)x^{-1}$ for all $x \in G$) and that $\phi(g) = g\eta(g)$ for all g.

Finally, we claim $\eta : G \to Z(G)$ is a group homomorphism. After all, for any $g, h \in G$ we know $gh\eta(gh) = \phi(gh) = \phi(g)\phi(h) = g\eta(g)h\eta(h)$. And since $\eta(g) \in Z(G)$, we have that $\eta(g)h = h\eta(g)$. So, $gh\eta(gh) = gh\eta(g)\eta(h)$. Or in other words, $\eta(gh) = \eta(g)\eta(h)$.

This proves that for any $\phi \in Z(\mathrm{Aut}(G))$ there must exist $\eta \in \mathrm{Hom}(G, Z(G))$ such that $\phi(g) = \mathrm{Id}(g)\eta(g)$. In particular, the map $Z(\mathrm{Aut}(G)) \to \mathrm{Hom}(G, Z(G))$ given by $\phi \mapsto \eta$ is injective and we thus have that $|Z(\mathrm{Aut}(G))| \leq |\mathrm{Hom}(G, Z(G))|$.

Note on what is the derived subgroup:
If $g, h \in G$, then the commutator of g with h is $[g,h] := ghg^{-1}h^{-1}$. The derived subgroup of G, denoted as $[G,G]$ is the group generated by all the commutators in G.

As for why we should care about if $G = [G,G]$, consider that if $\eta \in \mathrm{Hom}(G, Z(G))$, then:

$$\begin{aligned}\eta([g,h]) = \eta(ghg^{-1}h^{-1}) &= \eta(g)(\eta(h)\eta(g^{-1}))\eta(h^{-1}) \\ &= \eta(g)(\eta(g^{-1})\eta(h))\eta(h^{-1}) = \eta(gg^{-1})\eta(hh^{-1}) = e.\end{aligned}$$

In other words, $[G, G] \subseteq \ker(\eta)$ for all $\eta \in \mathrm{Hom}(G, Z(G))$. So, if $[G, G] = G$ then the only homomorphism from G to $Z(G)$ Is the trivial homomorphism.

(I'll also note that having $Z(G) = \{e\}$ is another way of making it so that the only $\eta \in \mathrm{Hom}(G, Z(G))$ is the trivial map...)

Set 2 Problem 3: Let $\mathrm{SL}_2(\mathbb{R})$ be the set of real 2×2 matrices with determinant 1. For any $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2$ and $z \in H := \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$, let: $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \cdot z := \frac{az+b}{cz+d}$.

How do we know this is well-defined? Suppose $cz + d = 0$. If $z = x + iy$, then since both c and d are real we must have that $cx + d = 0$ and $cy = 0$. But we know that $y > 0$. So this would imply that $c = 0$, and in turn that $d = 0$ also. But that contradicts that $\det(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)) = 1$. So, we know that we will always have that $cz + d \neq 0$ if our matrix is in $\mathrm{SL}_2(\mathbb{R})$ and $\mathrm{Im}(z) > 0$.

(a) Prove that $\mathrm{Im}\left(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \cdot z\right) = \frac{\mathrm{Im}(z)}{|cz+d|^2}$ for any $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$.

Note that $\frac{az+b}{cz+d} = \frac{(az+b)\overline{(cz+d)}}{|cz+d|^2}$. Also, if we set $z = x + iy$, then we can see that:

$$\begin{aligned} \mathrm{Im}((a(x+iy)+b)\overline{(c(x+iy)+d)}) &= \mathrm{Im}((ax+iay+b)(cx-icy+d)) \\ &= -acxy + acxy + adiy - bciy \\ &= \det(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right))y = 1 \cdot \mathrm{Im}(z) \end{aligned}$$

So $\mathrm{Im}\left(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \cdot z\right) = \mathrm{Im}(\frac{az+b}{cz+d}) = \frac{\mathrm{Im}(z)}{|cz+d|^2}$.

(b) Prove that $\cdot$ is an action of $\mathrm{SL}_2(\mathbb{R})$ on H.

By part (a) we know that $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \cdot z \in H$ for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ and $z \in H$. Also note that $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \cdot z = \frac{1z+0}{0z+1} = z$. Finally, if $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right)$, then note that $AB = \left(\begin{smallmatrix} aa'+bc' & ab'+bd' \\ ca'+dc' & cb'+dd' \end{smallmatrix}\right)$. Therefore:

- $(AB) \cdot z = \frac{(aa'+bc')z+ab'+bd'}{(ca'+dc')z+cb'+dd'}$.
- $A \cdot (B \cdot z) = g \cdot \frac{a'z+b'}{c'z+d'} = \frac{a\frac{a'z+b'}{c'z+d'}+b}{c\frac{a'z+b'}{c'z+d'}+d} = \frac{aa'z+ab'+bc'z+bd'}{ca'z+cb'+dc'z+dd'} = \frac{(aa'+bc')z+ab'+bd'}{(ca'+dc')z+cb'+dd'}$.

And this shows that $(AB) \cdot z = A \cdot (B \cdot z)$. ■

If $\mathcal{X}, \mathcal{Y}$ are both real vector spaces, we say that $f : \mathcal{X} \to \mathcal{X}$ is an <u>affine</u> transformation or map if for all $t \in [0, 1]$ and $x_1, x_2 \in \mathcal{X}$ we have that:

$$f(tx_1 + (1-t)x_2) = tf(x_1) + (1-t)f(x_2).$$

In other words, a transformation is affine if it maps line segments to line segments and in a way where "ratios of lengths" are preserved.

<u>Lemma:</u> A map $f : \mathcal{X} \to \mathcal{Y}$ is affine iff $f(x) = g(x) - b$ where $b \in \mathcal{Y}$ and g is a linear map from $\mathcal{X}$ to $\mathcal{Y}$.

($\Longleftarrow$) Suppose $f(x) = g(x) + b$ for all $x \in \mathcal{X}$. Then we clearly have for all $x_1, x_2 \in \mathcal{X}$

and $t \in [0,1]$ that:

$$\begin{aligned} f(tx_1 + (1-t)x_2) &= g(tx_1 + (1-t)x_2) + b \\ &= tg(x_1) + (1-t)g(x_2) + (t + (1-t))b \\ &= t(g(x_1) + b) + (1-t)(g(x_2) + b) = tf(x_1) + (1-t)f(x_2) \end{aligned}$$

$(\Longrightarrow)$
Fix $b = f(0)$ and then define $g(x) \coloneqq f(x) - b$. Then it is clear that $f(x) = g(x) + b$. And if we can prove that g is linear we will be done.

- Suppose $c \in [0,1]$ and $x \in \mathcal{X}$. Then:
  $$g(cx) = f(cx) - b = cf(x) + (1-c)b - b = cf(x) - cb = cg(x).$$

- Next suppose $c \in (1,\infty)$ and $x \in \mathcal{X}$. Then: $c^{-1} \in (0,1)$ and we know from before that $c^{-1}g(cx) = g(c^{-1}cx) = g(x)$. Or in other words, $g(cx) = cg(x)$.

- Now suppose $x_1, x_2 \in \mathcal{X}$. Then we know that:
  $$\begin{aligned} g(x_1 + x_2) &= 2g(\tfrac{1}{2}x_1 + \tfrac{1}{2}x_2) \\ &= 2(f(\tfrac{1}{2}x_1 + \tfrac{1}{2}x_2) - b) = 2(\tfrac{1}{2}f(x_1) + \tfrac{1}{2}f(x_2) - \tfrac{1}{2}b - \tfrac{1}{2}b) \\ &= 2(\tfrac{1}{2}g(x_1) + \tfrac{1}{2}g(x_2)) = g(x_1) + g(x_2). \end{aligned}$$

- Finally, suppose $x \in \mathcal{X}$. Then note that:
  $$g(-x) = g(\tfrac{2}{3}(-2x) + \tfrac{1}{3}x) = \tfrac{2}{3}g(-2x) + \tfrac{1}{3}g(x).$$

  Therefore, $\frac{1}{3}g(x) = g(-x) - \frac{2}{3}g(-2x) = g(-x) - \frac{4}{3}g(-x) = -\frac{1}{3}g(-x)$. Or in other words, $-g(x) = g(-x)$.

The four above points put together show that g is real-linear. ■

A corollary that is now easy to see is that for any $x_1, \ldots, x_n \in \mathcal{X}$ and nonnegative scalars $t_1, \ldots, t_n$ such that $\sum_{j=1}^n t_j = 1$, if f is affine then:

$$f(\textstyle\sum_{j=1}^n t_j x_j) = \sum_{j=1}^n t_j f(x_j).$$

Set 2 Problem 4: Suppose G is a finite group, $C \subseteq \mathbb{R}^n$ is nonempty and convex, and G acts on C by affine transformations. Then prove that G has a fixed point. Or in other words, prove there exists $x \in C$ such that $g \cdot x = x$ for all $g \in G$.

Based on the provided hints for the problem, I think the professor wants me to explicitely prove these facts first.

Lemma 1: If $c_1, \ldots, c_n \in C$, then $\frac{c_1 + \ldots + c_n}{n} \in C$.

Proof:
We proceed by induction on n. The base case of $n = 1$ is trivial. Meanwhile if $n > 1$, then:

$$\begin{aligned} \tfrac{c_1 + \ldots + c_n}{n} &= \tfrac{1}{n}c_n + \tfrac{1}{n}\textstyle\sum_{j=1}^{n-1} c_j \\ &= \tfrac{1}{n}c_n + \tfrac{n-1}{n}\textstyle\sum_{j=1}^{n-1} \tfrac{1}{n-1}c_j = \tfrac{1}{n}c_n + (1 - \tfrac{1}{n})\tfrac{c_1 + \ldots + c_{n-1}}{n-1} \end{aligned}$$

By our inductive hypothesis plus the definition of convexity, we thus know that $\frac{c_1+\ldots+c_n}{n} \in C$.

Lemma 2: If $x_1, \ldots, x_n \in \mathbb{R}^n$ and f is an affine map, then $f(\frac{x_1+\ldots+x_n}{n}) = \frac{f(x_1)+\ldots+f(x_n)}{n}$.

Proof:

Like before we proceed by induction on n. The base case $n = 1$ is trivial. Meanwhile, if $n > 1$, we know that $f(\frac{c_1+\ldots+c_n}{n}) = f(\frac{1}{n}c_n + (1-\frac{1}{n})\frac{c_1+\ldots+c_{n-1}}{n-1})$. Then by the definition of an affine map, we know:

$$f(\tfrac{1}{n}c_n + (1-\tfrac{1}{n})\tfrac{c_1+\ldots+c_{n-1}}{n-1}) = \tfrac{1}{n}f(c_n) + (1-\tfrac{1}{n})f(\tfrac{c_1+\ldots+c_{n-1}}{n-1}).$$

And by our inductive hypothesis we have that:

$$\tfrac{n-1}{n}f(\tfrac{c_1+\ldots+c_{n-1}}{n-1}) = \tfrac{n-1}{n} \cdot \tfrac{f(c_1)+\ldots+f(c_{n-1})}{n-1}$$

So, $f(\frac{c_1+\ldots+c_n}{n}) = \frac{1}{n}f(c_n) + \frac{f(c_1)+\ldots+f(c_{n-1})}{n}$.

Now pick any $y \in C$ and consider $A_G(y) = \frac{1}{|G|}\sum_{g\in G}(g \cdot y)$. By our first lemma, we know that $A_G(y) \in C$. Also, for any $h \in G$, we have by our second lemma that:

$$h \cdot A_G(y) = h \cdot (\tfrac{1}{|G|}\textstyle\sum_{g\in G}(g \cdot y)) = \tfrac{1}{|G|}\sum_{g\in G}(h \cdot (g \cdot y)) = \tfrac{1}{|G|}\sum_{g\in G}((hg) \cdot y)$$

But note that the map $G \to G$ given by $g \mapsto hg$ is a bijection. Therefore, we have that $\sum_{g\in G}((hg) \cdot y) = \sum_{g\in G}(g \cdot y)$ and this proves that $h \cdot A_G(y) = A_G(y)$ for all $h \in G$. ■

Set 2 Problem 5: Suppose G is a finite subgroup of the group $\mathrm{GL}_n(\mathbb{R})$ of $n \times n$ invertible real matrices. Prove that there is a G-invariant inner product on $\mathbb{R}^n$ (meaning that $\langle gv, gw\rangle = \langle v, w\rangle$ for all $g \in G$).

Given any $v, w \in \mathbb{R}^n$ define $\langle x, y\rangle = \frac{1}{|G|}\sum_{g\in G}(gv) \cdot (gw)$ where $\cdot$ is the dot product on $\mathbb{R}^n$. We claim this is a G-invariant inner product.

- Suppose $c_1, c_2 \in \mathbb{R}$ and $v_1, v_2, w \in \mathbb{R}^n$. Then:

$$\begin{aligned}\langle c_1v_1 + c_2v_2, w\rangle &= \tfrac{1}{|G|}\sum_{g\in G}(g(c_1v_1 + c_2v_2)) \cdot (gw) \\ &= \tfrac{1}{|G|}\sum_{g\in G}(c_1gv_1 + c_2gv_2) \cdot (gw) \\ &= \tfrac{1}{|G|}\sum_{g\in G}(c_1((gv_1) \cdot (gw)) + c_2((gv_2) \cdot (gw))) \\ &= \tfrac{c_1}{|G|}\left(\sum_{g\in G}(gv_1) \cdot (gw)\right) + \tfrac{c_2}{|G|}\left(\sum_{g\in G}(gv_2) \cdot (gw)\right) \\ &= c_1\langle v_1, w\rangle + c_2\langle v_2, w\rangle\end{aligned}$$

  Similarly, we can show that $\langle w, c_1v_1+c_2v_2\rangle = c_1\langle w, v_1\rangle + c_2\langle w, v_2\rangle$. So $\langle\,,\,\rangle$ is bilinear.

- Note that $\langle\,,\,\rangle$ is symmetric since for any $v, w \in \mathbb{R}^n$:

$$\langle v, w\rangle = \tfrac{1}{|G|}\sum_{g\in G}(gv) \cdot (gw) = \tfrac{1}{|G|}\sum_{g\in G}(gw) \cdot (gv) = \langle w, v\rangle.$$

- Suppose $v \in \mathbb{R}^n$. By the bilinearity we proved earlier, we can already see that $\langle v, v\rangle = 0$ if $v = 0$. Meanwhile, suppose $v \neq 0$. Then because each $g \in G$ is bijective, we know that $gv \neq 0$ and therefore $(gv) \cdot (gv) > 0$ for all $g \in G$. It easily follows that $\langle v, v\rangle = \frac{1}{|G|}\sum_{g\in G}(gv) \cdot (gv) > 0$.

- Let $v, w \in \mathbb{R}^n$ and consider any $h \in G$. Then since the map $G \to G$ given by $g \mapsto gh$ is a bijection, we know that:
$$\langle hv, hw\rangle = \tfrac{1}{|G|}\sum_{g\in G}(g(hv))\cdot(g(hw))$$
$$= \tfrac{1}{|G|}\sum_{g\in G}((gh)v)\cdot((gh)w) = \tfrac{1}{|G|}\sum_{g'\in G}(g'v)\cdot(g'w) = \langle v, w\rangle. \blacksquare$$

Set 2 Problem 6: Let G be a group and suppose $H < G$. We denote:

$$C_G(H) \coloneqq \{x \in G : xh = hx \text{ for all } h \in H\} \text{ and } N_G(H) \coloneqq \{x \in G : xHx^{-1} = H\}$$

to be the centralizer and normalizer of H in G respectively.

Note that $C_G(H)$ and $N_G(H)$ are both subgroups of G. To see this, note that $C_G(H) = \bigcap_{h\in H} C_G(h)$ where each $C_G(h)$ is the stabilizer of h with respect to the group action $G \curvearrowright G$ via conjugation. Meanwhile, $N_G(H)$ is the stabilizer of H with respect to the group action $G \curvearrowright \mathrm{Sub}(G)$ via conjugation. And since stabilizers are subgroups and intersections of subgroups are subgroups, we know that $C_G(H)$ and $N_G(H)$ are both subgroups of G.

Clearly $H < C_G(H) < N_G(H)$. Now prove that $N_G(H)/C_G(H)$ can be embedded into $\mathrm{Aut}(H)$ as a subgroup.

Note that because $H \lhd N_G(H)$, we know that $N_G(H) \curvearrowright H$ by conjugation. Or to put that in other words, there is a group homomorphism $c : N_G(H) \to \mathrm{Aut}(H)$ such that $c(g) = c_g : H \to H$ where $c_g(x) = gxg^{-1}$. In turn by the first isomorphism theorem, we know that there is an injective group homomorphism from the quotient group $N_G(H)/\ker(c)$ to $\mathrm{Aut}(H)$. So if we can show that $\ker(c) = C_G(H)$ then we will be done.

Fortunately, note that $c_g \in \ker(c)$ if and only if $c_g(x) = gxg^{-1} = x$ for all $x \in H$. Or in other words, $c(g) \in \ker(c)$ if and only if $gx = xg$ for all $x \in H$. This is the same as saying that $g \in C_G(H)$. $\blacksquare$

10/11/2025

Tonight I thought I'd go on a bit of a complete tangent and look at appendix D of John Lee's book *Introduction to Smooth Manifolds*. I'll be back to covering my actual classes tomorrow.

Theorem D.2 (ODE Comparison Theorem): Let $J \subseteq \mathbb{R}$ be an open interval containing 0 and suppose the differentiable function $u : J \to \mathbb{R}^n$ satisfies for all $t \in J$ that $\|u'(t)\|_2 \leq f(\|u(t)\|_2)$ where $f : [0,\infty) \to [0,\infty)$ is Lipschitz continuous. If we have that $v : [0,\infty) \to [0,\infty)$ is a differentiable real-valued function satisfying the initial-value problem $v'(t) = f(v(t))$ and $v(0) = \|u(0)\|_2$, then the inequality holds for all $t \in J$ that $\|u(t)\|_2 \leq v(|t|)$.

Proof:
Let $J^+ \coloneqq \{t \in J : t \geq 0\}$. Then we shall first show that $\|u(t)\|_2 \leq v(|t|) = v(t)$ for all $t \in J^+$.

Note that for all $t \in J^+$ where $\|u(t)\|_2 > 0$, we have that:

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}\|u(t)\|_2 &= \tfrac{\mathrm{d}}{\mathrm{d}t}(u(t)\cdot u(t))^{1/2} \\ &= \tfrac{1}{2}(u(t)\cdot u(t))^{-1/2}(2u(t)\cdot u'(t)) \\ &= (\|u(t)\|_2)^{-1}(u(t)\cdot u'(t)) \leq (\|u(t)\|_2)^{-1}(\|u(t)\|_2\|u'(t)\|_2) \\ &= \|u'(t)\|_2 \leq f(\|u(t)\|_2)\end{aligned}$$

Let A be a Lipschitz constant for f (i.e. $|f(t) - f(s)| \leq A|t-s|$ for all $s, t \in [0,\infty)$) and then consider the continuous function $w : J^+ \to \mathbb{R}$ defined by:

$w(t) = e^{-At}(\|u(t)\|_2 - v(t))$. Now $w(t_0) = w(0) = 0$.

For any $t \in J^+$ we have that $\|u(t)\|_2 \leq v(|t|) = v(t)$ iff $w(t) \leq 0$. But note that if $t \in J^+$ and $w(t) > 0$ (meaning $\|u(t)\|_2 > v(t) \geq 0$), then w is differentiable with:

$$\begin{aligned}w'(t) &= -Ae^{-At}(\|u(t)\|_2 - v(t)) + e^{-At}\tfrac{\mathrm{d}}{\mathrm{d}t}(\|u(t)\|_2 - v(t)) \\ &\leq -Ae^{-At}(\|u(t)\|_2 - v(t)) + e^{-At}(f(\|u(t)\|_2) - f(v(t))) \\ &\leq -Ae^{-At}(\|u(t)\|_2 - v(t)) + Ae^{-At}(\|u(t)\|_2 - v(t)) = 0\end{aligned}$$

Therefore, suppose there is some $t_1 \in J^+$ such that $w(t_1) > 0$ and let:

$$\tau := \sup\{t \in [0, t_1] : w(t) \leq 0\}.$$

By continuity we know that $w(\tau) = 0$ and $w(t) > 0$ for $t \in (\tau, t_1]$. And since w is continuous on $[\tau, t_1]$ and differentiable on (τ, t_1), the mean value theorem implies that there exists $t \in (\tau, t_1)$ such that $w(t) > 0$ and $w'(t) > 0$. But this contradicts our prior calculation that $w'(t) \leq 0$ whenever $w(t) > 0$. Thus, we must have that $w(t) \leq 0$ for all $t \in J^+$.

Meanwhile, to prove this theorem for all $t \in J^- := \{t \in J : t \leq 0\}$ we define $U(t) := u(-t)$. Then clearly $U(0) = u(0)$ so $v(0) = \|U(0)\|_2$. Also:

$$\|\tfrac{\mathrm{d}}{\mathrm{d}t}U(t)\|_2 = |-1|\|\tfrac{\mathrm{d}}{\mathrm{d}t}u(-t)\|_2 \leq f(\|u(-t)\|_2) = f(\|U(t)\|_2).$$

So, we can repeat all the prior reasoning to $U(t)$ to finish our proof. ■

<u>Generalization of Theorem D.2:</u> Let $J \subseteq \mathbb{R}$ be an open interval and suppose the differentiable function $u : J \to \mathbb{R}^n$ satisfies for all $t \in J$ that $\|u'(t)\|_2 \leq f(\|u(t)\|_2)$ where $f : [0,\infty) \to [0,\infty)$ is Lipschitz continuous. If for some $t_0 \in J$ we have that $v : [0,\infty) \to [0,\infty)$ is a differentiable real-valued function satisfying the initial-value problem $v'(t) = f(v(t))$ and $v(0) = \|u(t_0)\|_2$, then the inequality holds for all $t \in J$ that $\|u(t)\|_2 \leq v(|t - t_0|)$.

Proof:

We can apply the theorem we already proved to the function $\widetilde{u}(t) := u(t + t_0)$ defined on the interval $\widetilde{J} := \{t : t + t_0 \in J\}$ to show that $\|\widetilde{u}(t)\|_2 = \|u(t+t_0)\|_2 \leq v(|t|)$. Or equivalently, this means that $\|u(t)\|_2 \leq v(|t - t_0|)$ for all $t \in J$. ■

Recap: In broadstrokes, what we've essentially proven is that if f is Lipschitz, $\|u'\|_2 \leq f(\|u\|_2)$, $v' = f(v)$, and $v(0) = \|u(t_0)\|_2$, then $\|u(t)\|_2 \leq v(|t - t_0|)$.

Now before proving the existence and uniqueness theorem for solutions of differential equations, we need to quickly introduce our notation / conventions, as well as a few definitions.

Here is what an initial value problem is:

Suppose $V_1, \ldots, V_n$ are continuous real-valued functions defined on an open subset $W \subseteq \mathbb{R}^{n+1}$. Then our goal is to find differentiable real-valued functions $y_1, \ldots, y_n$ satisfying that:

- $y_i' = V_i(t, y_1(t), \ldots, y_n(t))$ for all $i \in \{1, \ldots, n\}$;
- $y_i(t_0) = c_i$ for all $i \in \{1, \ldots, n\}$ where $(t_0, c_1, \ldots, c_n)$ is some arbitrary point in W that we fixed.

If none of the V_i depend on their first argument, we say we are trying to solve an autonomous system. Otherwise, we say we are trying to solve a nonautonomous system.

We shall focus on the autonomous case first since it is easier. Also note that when we are working in the autonomous case, we can without loss of generality just take W to be an open subset of $\mathbb{R}^n$ rather than $\mathbb{R}^{n+1}$.

One more note: we shall define the $\mathbb{R}^n$-valued functions $V := (V_1, \ldots, V_n)$ and $y := (y_1, \ldots, y_n)$ for ease of notation.

Suppose (M_1, d_1), (M_2, d_2) are metric spaces and $F : M_1 \to M_2$ is a map. Then F is locally Lipschitz continuous if every point $x \in M_1$ has a neighborhood on which F is Lipschitz continuous.

Proposition C.29: Let $U \subseteq \mathbb{R}^n$ be open and suppose $F : U \to \mathbb{R}^m$ is a C^1 function. Then F is Lipschitz continuous on every compact convex subset $K \subseteq U$.

Specifically, you can take the Lipschitz constant to be $\sup_{x \in K} \|\mathrm{D}F(x)\|_F$ where $\|\cdot\|_F$ is the Frobenius norm of the matrix associated with $\mathrm{D}F(x)$ (when we equip $\mathbb{R}^n$ and $\mathbb{R}^m$ with the standard basis vectors)...

Proof:
$\|\mathrm{D}F(x)\|_F$ is a continuous function of x. So, it is bounded on the compact set K and we can let $M := \sup_{x \in K} \|\mathrm{D}F(x)\|_F$. Now for any $a, b \in K$ we have that $a + t(b-a) \in K$ for all $t \in [0,1]$. And so by the fundemental theorem of calculus we have that:

$$F(b) - F(a) = \int_0^1 \tfrac{\mathrm{d}}{\mathrm{d}t} F(a + t(b-a))\mathrm{d}t = \int_0^1 \mathrm{D}F(a + t(b-a))(b-a)\mathrm{d}t$$

In turn, by Minkowski's inequality we have that:

$$\begin{aligned}\|F(b) - F(a)\|_2 &\leq \| \textstyle\int_0^1 \mathrm{D}F(a + t(b-a))(b-a)\mathrm{d}t\|_2 \\ &\leq \textstyle\int_0^1 \|\mathrm{D}F(a + t(b-a))(b-a)\|_2 \mathrm{d}t \\ &\leq \textstyle\int_0^1 \|\mathrm{D}F(a + t(b-a))\|_{\mathrm{op}} \cdot \|b-a\|_2 \mathrm{d}t \leq M\|b-a\|_2\end{aligned}$$

Corollary C.30: If $U \subseteq \mathbb{R}^n$ is open and $F : U \to \mathbb{R}^m$ is a C^1 map, then F is locally Lipschitz continuous.

Theorem D.3 (Existence of ODE solutions): Let $U \subseteq \mathbb{R}^n$ be an open subset and suppose $V = (V_1, \ldots, V_n) : U \to \mathbb{R}^n$ is locally Lipschitz continuous. Let $(t_0, x_0) \in \mathbb{R} \times U$ be given. Then there exists an interval $J_0 \subseteq \mathbb{R}$ containing t_0; an open subset $U_0 \subseteq U$ containing x_0; and for each $c = (c_1, \ldots, c_n) \in U_0$, a C^1 map $y = (y_1, \ldots, y_n) : J_0 \to U$ satisfying the initial value problem:

- $y_i'(t) = V_i(y_1(t), \ldots, y_n(t))$ for all i;
- $y_i(t_0) = c_i$ for all i.

$(*)$

Proof:
To start off, by shrinking U we can assume without loss of generality that V is Lipschitz continuous on U. Or in other words, there exists $C > 0$ such that

$$\|V(x) - V(\widetilde{x})\|_2 \leq C\|y - \widetilde{y}\|_2 \text{ for all } x, \widetilde{x} \in U.$$

Next, given $t_0 \in \mathbb{R}$ and $x_0 \in U$, choose $r > 0$ small enough that $\overline{B_r(x_0)} \subseteq U$ (where $B_r(x_0)$ denotes the Euclidean ball of radius r about x_0 in $\mathbb{R}^n$). Then because $\overline{B_r(x_0)}$ is compact, we know that $M := \sup_{x \in B_r(x_0)} \|V(x)\|$ exists. So, pick $\delta > 0$ and $\varepsilon > 0$ small enough so that $\delta < \frac{r}{2}$ and $\varepsilon < \min(\frac{r}{2M}, \frac{1}{C})$. And finally, set $J_0 := (t_0 - \varepsilon, t_0 + \varepsilon)$ and $U_0 := B_\delta(x_0)$. These will be the restricted sets we work with.

Now, we show $(*)$ is equivalent to a certain integral equation.

$(\Longrightarrow)$
Suppose y is any solution to $(*)$ on some interval J_0 containing t_0. Then if we integrate $y_i'(t) = V_i(y_1(t), \ldots, y_n(t))$, since V is continuous we can apply the fundamental theorem of calculus to get that:

$$y_i(t) = c_i + \int_{t_0}^t V_i(y(s))\mathrm{d}s$$

$(\Longleftarrow)$
Suppose $y = (y_1, \ldots, y_n) : J_0 \to U$ is a continuous map such that for all i we have that $y_i(t) = c_i + \int_{t_0}^t V_i(y(s))\mathrm{d}s$. Then the fundamental theorem of calculus says that $y_i'(t) = V_i(y(t))$. Also it is easy to evaluate that $y_i(t_0) = c_i$ for all i.

This motivates the following definition. Given any fixed $c \in U_0$, let I be an operator such that $Iy : J_0 \to \mathbb{R}^n$ is given by $Iy(t) = c + \int_{t_0}^t V(y(s))\mathrm{d}s$ for all $y \in C(J_0, U)$.

Note: because both y and V are continuous and $[t_0, t]$ (or $[t, t_0]$) is compact, we know that $\int_{t_0}^t V(y(s))\mathrm{d}s$ exists for all $t \in J_0$.

Importantly, any fixed point of I is a solution to $(*)$. So, if we could apply the Banach fixed point theorem somehow (see page 26 of my math 140c notes), then we'd be done.

Thus, let $\mathcal{M}_c$ be the collection of all continuous maps $y : J_0 \to \overline{B_r(x)}$ satisfying that $y(t_0) = c$. We know $\mathcal{M}_c$ is nonempty since we can just consider the constant map $y = c$. Also, it is easy to see that $\mathcal{M}_c$ is complete with respect to the uniform norm.

We claim I maps $\mathcal{M}_c$ into itself. After all, if $y \in \mathcal{M}_c$ then for any $t \in J_0$ we have that:

$$\begin{aligned}\|Iy(t) - x_0\|_2 &= \|c + \textstyle\int_{t_0}^t V(y(s))\mathrm{d}s - x_0\|_2 \\ &\leq \|c - x_0\|_2 + \textstyle\int_{t_0}^t \|V(y(x))\|_2 \mathrm{d}s \quad \text{(hehe, used Minkowski's inequality)} \\ &< \delta + M\varepsilon < r\end{aligned}$$

Finally, we claim I is a contraction on $\mathcal{M}_c$. After all, if $y, \widetilde{y} \in \mathcal{M}_c$ then:

$$\begin{aligned}\|Iy - I\widetilde{y}\|_u &= \sup_{t\in J_0} \|c + \textstyle\int_{t_0}^t V(y(s))\mathrm{d}s - (c + \int_{t_0}^t V(\widetilde{y}(s))\mathrm{d}s)\|_2 \\ &\leq \sup_{t \in J_0} \textstyle\int_{t_0}^t \|V(y(s)) - V(\widetilde{y}(s))\|_2 \mathrm{d}s \\ &\leq \sup_{t\in J_0} \textstyle\int_{t_0}^t C\|y(s) - \widetilde{y}(s)\|_2 \mathrm{d}s \leq C\varepsilon \|y - \widetilde{y}\|_u\end{aligned}$$

And because $\varepsilon < \frac{1}{C}$, this proves that I is a contraction. So, we can apply the Banach fixed point theorem to get a solution to $(*)$. ■

Theorem D.4: (Uniqueness of ODE Solutions): Let $U \subseteq \mathbb{R}^n$ be an open subset and suppose $V : U \to \mathbb{R}^n$ is locally Lipschitz continuous. For any $t_0 \in \mathbb{R}$ and $c \in U$, any two differentiable solutions to the initial value problem $y'(t) = V(y(t))$ and $y(t_0) = c$ $(**)$ are equal on their common domain.

Proof:

Suppose J_1 and J_2 are both open intervals of $\mathbb{R}$ containing t_0 and $y : J_1 \to U$ and $\widetilde{y} : J_2 \to U$ are differentiable functions satisfying that $y'(t) = V(y(t))$, $\widetilde{y}'(t) = V(\widetilde{y}(t))$, and $y(t_0) = c = \widetilde{y}(t_0)$. By setting $J_0 := J_1 \cap J_2$ we can now without loss of generality assume y and $\widetilde{y}$ have the same domain.

Let J_0' be any bounded open interval containing t_0 such that $\overline{J_0'} \subseteq J_0$. Then the union of $y(\overline{J_0'})$ and $\widetilde{y}(\overline{J_0'})$ is a compact subset of U. And since V is locally Lipschitz continuous on U, we can find a constant C such that:

$$\|\tfrac{\mathrm{d}}{\mathrm{d}t}(\widetilde{y}(t) - y(t))\|_2 = \|V(\widetilde{y}(t)) - V(y(t))\|_2 \leq C\|\widetilde{y}(t) - y(t)\|_2.$$

Why? Suppose (M_1, d_1), (M_2, d_2) are metric spaces and $F : M_1 \to M_2$ is a locally Lipschitz continuous map. Then if $K \subseteq M_1$ is compact, we know $F|_K$ is Lipschitz continuous.

Proof:

We know $F(K)$ is compact and thus bounded. So let $D := \mathrm{diam}(F(K))$. Next for each $x \in K$ pick $\delta(x) > 0$ such that F is Lipschitz continuous on $B_{2\delta(x)}(x)$ with Lipschitz constant $C(x)$. Since K is compact, there are finitely many points $x_1, \ldots, x_n \in K$ such that $K \subseteq B_{\delta(x_1)}(x_1) \cup \ldots \cup B_{\delta(x_n)}(x_n)$. So, set $C := \max(C(x_1), \ldots, C(x_n))$ and $\delta = \min(\delta(x_1), \ldots, \delta(x_n))$.

Now it is easy to see that $d_2(F(x), F(y)) \leq \max(C, {}^D/_\delta) d_1(x, y)$.

- If $d_1(x, y) < \delta$, then $d_2(F(x), F(y)) \leq C d_1(x, y)$.
- If $d_1(x, y) \geq \delta$, then $d_2(F(x), F(y)) \leq D < {}^D/_\delta d_1(x, y)$.

This shows that F is Lipschitz on K.

Now by applying the the *ODE comparison theorem* with $u(t) = \widetilde{y}(t) - y(t)$, $f(v) = Cv$, and $v(t) = 0$, we can conclude that:

$$\|\widetilde{y}(t) - y(t)\|_2 \leq 0 \text{ for all } t \in J_0'.$$

So $\widetilde{y}|_{J_0'} = y|_{J_0'}$. And since every point of J_0 can be contained in some such interval J_0', we know that $\widetilde{y} = y$ everywhere on J_0. ■

The obvious corollary to this is that given our function $V : U \to \mathbb{R}^n$ from before, if $y : J_1 \to U$ and $\widetilde{y} : J_2 \to U$ satisfy that $y'(t) = V(y(t))$ and $\widetilde{y}'(t) = V(\widetilde{y}(t))$ and there is some $t_0 \in J_1 \cap J_2$ such that $y(t_0) = \widetilde{y}(t_0)$, then you can get a well-defined function by taking the union of y and $\widetilde{y}$ (considered as subsets of $\mathbb{R} \times U$). And this union will also satisfy our differential equation on it's domain.

The big application of the last theorem is that if we were to find solutions for initial value problems like $(**)$ but on much larger domains than just $(t_0 - \varepsilon, t_0 + \varepsilon)$, then we'd be able to conclude that our solution is unique on that larger domain as well.

I am not done working through appendix D of *Introduction to Smooth Manifolds*. That said, I have other things I need to focus on for now. So I'll continue working on this later on *page* ___.

Math 220a (Lecture 7):

To start off, while these theorems were covered in a previous class I took, I never actually took notes on them. So, I thought I might as well prove them right now.

- Theorem: Suppose $\sum_{n=0}^{\infty} a_n(z-\alpha)^n$ is a power series with radius of convergence $R > 0$. (As a reminder, $R = (\limsup_{n\to\infty} |a_n|^{1/n})^{-1}$). If $0 < r < R$, then the series converges uniformly on $\{z : |z - \alpha| \leq r\}$.

  Proof:
  Pick $\rho \in (r, R)$ and let N be such that $|a_n| < \rho^{-n}$ for all $n \geq N$. Then for all z with $|z - \alpha| \leq r$ we have that $|a_n(z-\alpha)^n| \leq (\frac{r}{\rho})^n$ for all $n \geq N$ and $\frac{r}{\rho} < 1$.

  As for the $n < N$, since $a_n(z-\alpha)^n$ is continuous and $\{z : |z - \alpha| \leq r\}$ is compact, we know by the extreme value theorem that for each $n < N$ there is a finite number C_n with $a_n(z-\alpha)^n \leq C_n$. By combining everything above, we can conclude by the Weierstrass M-test that the power series converges uniformly on $\{z : |z - \alpha| \leq r\}$. ■

- Corollary: If $f(z) = \sum_{n=0}^{\infty} a_n(z-\alpha)^n$ has radius of convergence R, then f is continuous at all z with $|z - \alpha| < R$.

- Theorem (Baby Rudin 8.2): Suppose $\sum_{n=0}^{\infty} c_n$ converges and put $f(x) = \sum_{n=0}^{\infty} c_n z^n$ for $|z| < 1$. Then $\lim_{z\to 1} f(z) = \sum_{n=0}^{\infty} c_n$.

Proof:
Let $s_{-1} = 0$ and $s_n = c_0 + \ldots + c_n$ for all n. Then:

$$\sum_{n=0}^{m} c_n z^n = \sum_{n=0}^{m}(s_n - s_{n-1})z^n = s_m z^m + (1-z)\sum_{n=0}^{m-1} s_n z^n$$

Next let $s = \lim_{n\to\infty} s_n$ (this exists because $\sum_{n=0}^{\infty} c_n$ converges). Now when $|z| < 1$, we may take $m \to \infty$ to get that:

$$f(z) = (1-z)\sum_{n=0}^{\infty} s_n z^n + \lim_{m\to\infty} s_m z^m$$
$$= (1-z)\sum_{n=0}^{\infty} s_n z^n + (s \cdot 0) = (1-z)\sum_{n=0}^{\infty} s_n z^n$$

Now for any $\varepsilon > 0$ pick $N \in \mathbb{N}$ such that $|s_n - s| < \varepsilon/2$ for all $n \geq N$. Then note that because $(1-z)\sum_{n=0}^{\infty} z^n = 1$ whenever $|z| < 1$, we have that:

$$|f(z) - s| = |(1-z)\sum_{n=0}^{\infty} s_n z^n - ((1-z)\sum_{n=0}^{\infty} z^n)\, s|$$
$$= |(1-z)\sum_{n=0}^{\infty}(s_n - s)z^n|$$
$$\leq |1-z|\sum_{n=0}^{\infty} |s_n - s||z|^n$$
$$= |1-z|\sum_{n=0}^{N-1} |s_n - s||z|^n + |1-z|\sum_{n=N}^{\infty} |s_n - s||z|^n$$
$$\leq |1-z|\sum_{n=0}^{N-1} |s_n - s| + \tfrac{\varepsilon}{2}|1-z|\sum_{n=N}^{\infty} |z|^n$$
$$\leq |1-z|\sum_{n=0}^{N-1} |s_n - s| + \tfrac{\varepsilon}{2}(1)$$

And in turn, clearly there is some $\delta > 0$ such that whenever $|1-z| < \delta$, then $|f(z) - s| < \varepsilon$. So, $\lim_{z\to 1} f(z) = s$. ■

- Theorem (Baby Rudin 3.51): If $\sum_{n=0}^{\infty} a_n = A$, $\sum_{n=0}^{\infty} b_n = B$, $\sum_{n=0}^{\infty} c_n = C$, and $c_n = a_0 b_n + a_1 b_{n-1} + \ldots + a_n b_0$ for all n, then $AB = C$.

  Proof:
  Define $f(x) = \sum_{n=0}^{\infty} a_n x^n$, $g(x) = \sum_{n=0}^{\infty} b_n x^n$, and $h(x) = \sum_{n=0}^{\infty} c_n x^n$. Then f, g, and h all have radii of convergence at least 1. So, f, g, and h converge absolutely when $|x| < 1$. And by Merten's theorem (see my math 140a notes), we can therefore say that $f(x)g(x) = h(x)$ for all x with $|x| < 1$. And by taking the limit as $x \to 1$, we get that $AB = C$. ■

Ok. Now that I've gotten that out of my system, I'm now going to return to the book that math 220 is actually using.

Proposition III.1.5: Let $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ be two absolutely convergent series and put $c_n = \sum_{k=0}^{n} a_k b_{n-k}$ for all $n \in \mathbb{N}$. Then $\sum_{n=0}^{\infty} c_n$ is absolutely convergent.

Proof:
Let $d_n = \sum_{k=0}^{n} |a_k||b_{n-k}|$ for all $n \in \mathbb{N}$. Then clearly $|c_n| \leq d_n$. Also, by Merten's theorem (see my math 140a notes) we know that $\sum_{n=0}^{\infty} d_n = (\sum_{n=0}^{\infty} |a_n|)(\sum_{n=0}^{\infty} |b_n|) < \infty$. So, we know by comparison test that $\sum_{n=0}^{\infty} c_n$ is absolutely convergent. ■

We say $G \subseteq \mathbb{C}$ is a region if it is open and connected.

Now suppose G is a region and $f : G \to \mathbb{C}$ is a function. We say f is $\mathbb{C}$-differentiable at $a \in G$ if $f'(a) := \lim_{z \to a} \frac{f(z)-f(a)}{z-a}$ exists. And when $f'(a)$ exists we call it the derivative of f.

We say $f \in H(G)$ and call f holomorphic on G if f' exists at every point in G and is continuous.

> This is a bit different from my complex analysis notes I took last Spring in which I (and Rudin) didn't require f' to be continuous in order to say that f is holomorphic. And sure enough, it turns out that Conway proves (much) later that the continuity assumption is unnecessary. So...

What is the difference between $\mathbb{C}$-differentiability and $\mathbb{R}^2$-differentiability?

> Let $f : G \to \mathbb{C}$ be given by $f = u + iv$ and write $z = x + iy$. If we consider $\mathbb{C} \cong \mathbb{R}^2$, then we can express f as a real-valued function $f(x, y) = (u(x, y), v(x, y))$.
>
> Suppose f is $\mathbb{R}^2$-differentiable at (x_0, y_0). Then there exists a 2x2 matrix A such that:
> $$\lim_{(h,k)\to(0,0)} \frac{\|f(x_0+h, y_0+k) - f(x_0, y_0) - A(h,k)\|_2}{\|(h,k)\|_2} = 0$$
>
> In turn, the partial derivatives u_x, u_y, v_x, v_y exists at (x_0, y_0) and $A = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}$.
>
> Meanwhile, suppose f is $\mathbb{C}$-differentiable at $x_0 + iy_0$ and that $f'(x_0 + iy_0) = r + is$. Now $f'(x_0 + iy_0)(h + ik) = (r + is)(h + ik) = rh - sk + i(sh + rk) \cong \begin{pmatrix} r & -s \\ s & r \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix}$.
>
> Also note that $\lim\limits_{h+ik\to 0} \frac{|f(x_0+h+i(y_0+k)) - f(x_0+iy_0) - f'(x_0+iy_0)(h+ik)|}{|h+ik|} = 0.$
>
> This leads to the following theorem:
>
> Theorem: If G is a region of $\mathbb{C} \cong \mathbb{R}^2$, then $f \in H(G)$ if and only if $f \in C^1(G)$ with $u_x = v_y$ and $v_x = -u_y$.

You can prove sum rule, product rule, quotient rule, and chain rule for holomorphic functions using identical proofs to how we proved those rules for real-valued functions. Also, polynomials differentiate identically to how you would expect and complex differentiability implies continuity.

There was one more theorem partially covered in the lecture about power series being infinitely $\mathbb{C}$-differentiable within their radii of convergence. However, I'll just leave that proof to my paper notes from last Spring.

I'll continue with this class on *page 309*.

Math 200a (lecture 6)

Sylow's 2nd. Theorem: Suppose $P_0 \in \mathrm{Syl}_p(G)$ and Q is a p-subgroup of G. Then there exists $x \in G$ such that $Q \subseteq xP_0x^{-1}$.

Proof:
$Q \curvearrowright G/P_0$ by left-translation. Then by the theorem in the middle of *page 271*, we have that $|G/P_0| \equiv |(G/P_0)^Q| \pmod{p}$. But since $P_0 \in \mathrm{Syl}_p(G)$, we have that $|G/P_0| \not\equiv 0 \pmod{p}$. Hence, there must exist some $gP_0 \in (G/P_0)^Q$.

In turn, $xgP = gP$ for all $x \in Q$. Or in other words, $g \in gPg^{-1}$ for all $x \in Q$. Hence, $Q \subseteq gPg^{-1}$. ■

Corollary: If $P_1, P_2 \in \mathrm{Syl}_p(G)$ then there exists $g \in G$ such that $gP_1g^{-1} = P_2$.

Proof:
By Sylow's 2nd theorem we know there exists $g \in G$ such that $P_2 \subseteq gP_1g^{-1}$. And since $|P_2| = |P_1|$ we deduce $P_2 = gP_1g^{-1}$. ■

Note the following observations:

- If $\theta \in \mathrm{Aut}(G)$ and $P \in \mathrm{Syl}_p(G)$ then $\theta(P) \in \mathrm{Syl}_p(G)$.
- $G \curvearrowright \mathrm{Syl}_p(G)$ by conjugation and this actions is transitive (by the last corollary).
- A subgroup $H < G$ is called a characteristic subgroup if $\forall\theta \in \mathrm{Aut}(G)$ we have that $\theta(H) = H$. By the last two observations, if $\mathrm{Syl}_p(G) = \{P\}$, then P is a characteristic subgroup of G (which automatically means P is normal since conjugation is an automorphism of G).

Corollary: If $P \lhd G$ and $P \in \mathrm{Syl}_p(G)$, then P is a characteristic subgroup of G.

Proof:
Since $P \lhd G$, $P \in \mathrm{Syl}_p(G)$, and $G \curvearrowright \mathrm{Syl}_p(G)$ transitively via conjugation, we must have that $\mathrm{Syl}_p(G) = \{P\}$. Hence P is a characteristic subgroup of G. ■

Lemma: If $P \in \mathrm{Syl}_p(G)$, then $\mathrm{Syl}_p(N_G(P)) = \{P\}$.

Proof:
We know $|P| = p^{\nu_p(|G|)}$. Also, $P < N_G(P) < G$ means that $|P|$ divides $|N_G(P)|$ and $|N_G(P)|$ divides $|G|$. Thus $\nu_p(|G|) = \nu_p(|N_G(P)|)$ and so $P \in \mathrm{Syl}_p(N_G(P))$. Finally, since $P \lhd N_G(P)$, we know from the last corollary that $\mathrm{Syl}_p(N_G(P)) = \{P\}$. ■

Lemma: If $P_0 \in \mathrm{Syl}_p(G)$ and we consider $P_0 \curvearrowright \mathrm{Syl}_p(G)$ by conjugation, then $(\mathrm{Syl}_p(G))^{P_0} = \{P_0\}$.

Proof:
$P \in (\mathrm{Syl}_p(G))^{P_0}$ if and only if for all $x \in P_0$, $xPx^{-1} = P$. That's to say, iff $P_0 \subseteq N_G(P)$. But that would mean $P_0 \in \mathrm{Syl}_p(N_G(P)) = \{P\}$. So $(\mathrm{Syl}_p(G))^{P_0} = \{P_0\}$. ■

Sylow's 3rd. Theorem: If G is a finite group, $|\mathrm{Syl}_p(G)| \equiv 1 \pmod{p}$.

Proof:

Suppose $P_0 \in \mathrm{Syl}_p(G)$. Then $|\mathrm{Syl}_p(G)| \equiv |(\mathrm{Syl}_p(G))^{P_0}| \pmod{p}$. But from the prior lemma we know $|(\mathrm{Syl}_p(G))^{P_0}| = 1$. ■

So as a recap, suppose G is a finite group and p is a prime number dividing $|G|$. Then:

- Sylow's first theorem guarentees that $\mathrm{Syl}_p(G) \neq \emptyset$.
- Sylow's third theorem guarentees that $|\mathrm{Syl}_p(G)| \equiv 1 \pmod{p}$.
- Sylow's second theorem guarentees that $|\mathrm{Syl}_p(G)|$ equals the number of conjugates of P_0 where $P_0 \in \mathrm{Syl}_p(G)$. Thus (see *page 271*), we have for any $P_0 \in \mathrm{Syl}_p(G)$ that $|\mathrm{Syl}_p(G)| = [G : N_G(P_0)]$. And in particular, since $P_0 < N_G(P_0) < G$, we have that $|\mathrm{Syl}_p(G)| = \frac{|G|}{|N_G(P_0)|} = \frac{[G:P_0]|P_0|}{[N_G(P_0):P_0]|P_0|} = \frac{[G:P_0]}{[N_G(P_0):P_0]}$. So $|\mathrm{Syl}_p(G)|$ divides $[G : P_0]$.

Proposition: $P \in \mathrm{Syl}_p(G) \Longrightarrow N_G(N_G(P)) = N_G(P)$.

Proof:

The $\supseteq$ inclusion is obvious. Meanwhile, $x \in N_G(N_G(P))$ implies that $xN_G(P)x^{-1} = N_G(P)$. But note that if $\theta \in \mathrm{Aut}(G)$ and $H < G$, then $\theta(N_G(H)) = N_G(\theta(H))$.

If $x \in N_G(H)$ then we know that $xHx^{-1} = H$. So:

$$\phi(x)\phi(H)\phi(x)^{-1} = \phi(xHx^{-1}) = \phi(H).$$

This shows that $\phi(x) \in N_G(\phi(H))$ and hence $\phi(N_G(H)) \subseteq N_G(\phi(H))$ whenever $\phi \in \mathrm{Aut}(G)$. Using this fact, now note that for any $\phi \in \mathrm{Aut}(G)$, we have that:

$$N_G(H) = \phi^{-1}(\phi(N_G(H))) \subseteq \phi^{-1}(N_G(\phi(H))) \subseteq N_G(\phi^{-1}(\phi(H))) = N_G(H)$$

So, $N_G(H) = \phi^{-1}(N_G(\phi(H)))$. And by composing ϕ we get that:

$$\phi(N_G(H)) = N_G(\phi(H)).$$

It follows that $N_G(xPx^{-1}) = xN_G(P)x^{-1} = N_G(P)$ whenever $x \in N_G(N_G(P))$. But in that case we have that $\mathrm{Syl}_p(N_G(xPx^{-1})) = \mathrm{Syl}_p(N_G(P))$. And as P and xPx^{-1} are both Sylow p-groups, we conclude $xPx^{-1} = P$. So $x \in N_G(P)$

I probably should have been taught this in math 100a but never was. So, I guess I'll just refresh myself now. The book I'm following along with is *Abstract Algebra* by Dummit and Foote.

Suppose G is a group and H, K are subgroups of G. Then we define:

$$HK := \{hk \in G : h \in H \text{ and } k \in K\}.$$

Proposition 3.2.13: If H and K are finite subgroups of a group, then $|HK| = \frac{|H||K|}{|H \cap K|}$.

Proof:

Note that $HK = \bigcup_{h \in H} hK$. Thus $|HK|$ equals $|K|$ times the number of distinct left cosets hK where $h \in H$. But note that for any $h_1, h_2 \in H$:

$$h_1K = h_2K \iff h_2^{-1}h_1 \in H \cap K \iff h_1(H \cap K) = h_2(H \cap K).$$

Hence $|HK| = |K| \cdot [H : H \cap K] = |K|\frac{|H|}{|H \cap K|}$ by Lagrange's theorem. ■

Proposition 3.2.14: If H and K are subgroups of G, then $HK < G$ iff $HK = KH$.

$(\Longrightarrow)$
Suppose $HK < G$. Since $K < HK$ and $H < HK$, we thus know that $KH \subseteq HK$. Meanwhile, suppose $h \in H$ and $k \in K$. Since HK is a group, we know $(hk)^{-1} \in HK$. So there exists $h' \in H$ and $k' \in K$ such that $(hk)^{-1} = h'k'$. But then $hk = (k')^{-1}(h')^{-1}$ which is in KH. So $HK \subseteq KH$.

$(\Longleftarrow)$
Assume $HK = KH$ and let $a, b \in HK$. Then there exists $h_1, h_2 \in H$ and $k_1, k_2 \in K$ such that $a = h_1k_1$ and $b = h_2k_2$. Now it's clear that $1_G \in HK$. So, if we can show that $ab^{-1} \in HK$, then we will know that HK is a group.

Fortunately, $ab^{-1} = h_1k_1k_2^{-1}h_2^{-1}$. And since $KH = HK$, we know there is $h_3 \in H$ and $k_3 \in K$ such that $(k_1k_2^{-1})h_2^{-1} = h_3k_3$. Thus, $ab^{-1} = (h_1h_3)(k_3) \in HK$. ■

Corollary 3.2.15: If H and K are subgroups of G and $H < N_G(K)$, then HK is a subgroup of G. In particular, if $K \lhd G$ then $HK < G$ for any $H < G$.

Proof:
Let $h \in H$ and $k \in K$. Then $hkh^{-1} \in K$. So $hk = (hkh^{-1})h \in KH$ and we've proven that $HK = KH$. ■

Second Isomorphism Theorem: Let G be a group, let A and B be subgroups of G, and assume $A < N_G(B)$. Then $AB < G$, $B \lhd AB$, $A \cap B \lhd A$, and $AB/B \cong A/(A \cap B)$.

Proof:
By the last corollary we know that $AB < G$. Also, since $A < N_G(B)$ and $B < N_G(B)$, it follows $AB < N_G(B)$. Hence $B \lhd AB$.

Now we know there is a well-defined group homomorphism $\phi : A \to AB/B$ given by $\phi(a) = aB$. Clearly ϕ is surjective. Meanwhile, it's easy to see that $\ker(\phi) = A \cap B$. So by the first isomorphism theorem, we have that $A \cap B \lhd A$ and that:

$$AB/B \cong A/(A \cap B). \blacksquare$$

Here is one more miscellaneous result before getting back to the lecture:

Lemma: If $N_1, N_2 \lhd G$, then $\forall x \in N_1$ and $\forall y \in N_2$ we have that $xyx^{-1}y^{-1} \in N_1 \cap N_2$.

Proof:
$(xyx^{-1}) \in N_2$ and $(yx^{-1}y^{-1}) \in N_1$ since both N_1 and N_2 are normal. Hence:

$$(xyx^{-1})y^{-1} = x(yx^{-1}y^{-1}) \in N_1 \cap N_2. \blacksquare$$

Corollary: If $N_1, N_2 \lhd G$ and $N_1 \cap N_2 = \{1\}$, then $xy = yx$ for all $x \in N_1$ and $y \in N_2$.

So here are some uses of Sylow's theorems:

- Suppose $p < q$ are distinct primes with $p \nmid q - 1$. If $|G| = pq$ then $G \cong C_{pq}$.

  Let s_q and s_p be shorthand for $|\mathrm{Syl}_q(G)|$ and $|\mathrm{Syl}_p(G)|$. Now we know by Sylow's third theorem that $s_q \equiv 1 \pmod q$.

Also, we know that $s_q \mid p$ by Sylow's second theorem. And since $p < q$, we must have that $s_q = 1$. Hence $\mathrm{Syl}_q(G) = \{Q\}$ for some $Q \lhd G$ such that $|Q| = q$ and Q is cylic of order q.

Next, note once again by Sylow's second theorem that $s_p \mid q$. Hence, we must have that either $s_p = 1$ or $s_p = q$. That said, we know $q - 1 \not\equiv 0 \pmod p$ by assumption and that $s_p \equiv 1 \pmod p$ by Sylow's third theorem. So, we must have that $s_p = 1$ and it follows that $\mathrm{Syl}_p(G) = \{P\}$ for some $P \lhd G$ such that $|P| = p$ and P is cyclic of order p.

Now $|P \cap Q| \mid \gcd(|P|, |Q|) = 1$. So $P \cap Q = \{1\}$. And by our prior corollary, this means that $xy = yx$ for all $x \in P$ and $y \in Q$.

Now consider the map $f : P \times Q \to G$ given by $(x, y) \mapsto xy$. We claim this is a group isomorphism.

- Note that:
$$\begin{aligned} f(x_1, y_1)f(x_2, y_2) = x_1y_1x_2y_2 &= x_1x_2y_1y_2 \\ &= f(x_1x_2, y_1y_2) = f((x_1, y_1)(x_2, y_2)). \end{aligned}$$

  Thus f is a group homomorphism.

- Suppose $f(x, y) = 1$. Then $xy = 1$ which means that $x = y^{-1}$. But now $x, y^{-1} \in P \cap Q = \{1\}$. So $(x, y) = (1, 1)$ and we've shown that f is injective.

- $|\mathrm{im}(f)| = |PQ| = \frac{|P||Q|}{|P \cap Q|} = \frac{pq}{1} = |G|$. So f is surjective.

It follows that $G \cong P \times Q \cong \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z} \cong \mathbb{Z}/pq\mathbb{Z}$. (The last equivalence is from Chinese remainder theorem...)

I am not fully caught up with this class yet. But, I'll stop here for now so that I can go back to taking functional analysis notes. For more math 200a notes go to *page 304*.

10/14/2025

Math 241a (lectures 3-5 continued):

If $\mathcal{X}$ is a topological vector space, then here is one more topology on $\mathcal{X}^*$ to be aware of.

> Let $\mathcal{A}$ be the collection of all (Von-Neumann) bounded sets in $\mathcal{X}$ and then for each $A \in \mathcal{A}$ define the seminorm $p_A(\lambda) = \sup_{x \in A} |\lambda(x)|$ on $\mathcal{X}^*$. Since every singleton is bounded, we know this defines a sufficient family. Also, the topology generated by that family is finer than the weak* topology. So, we call it the strong topology on $\mathcal{X}^*$.

(Definition 1.2.19:) If $\mathcal{X}$, $\mathcal{Y}$ are topological (K-)vector spaces and $T \in B(\mathcal{X}, \mathcal{Y})$, then T's adjoint is defined as the map $T^* : \mathcal{Y}^* \to \mathcal{X}^*$ given by $T^*(\lambda) = \lambda \circ T$.

Note that:

- T^* is a well-defined linear operator.

  To show that T^* is well defined, suppose $c_1, c_2 \in K$ and $x_1, x_2 \in \mathcal{X}$. Then:
  $$T^*(\lambda)(c_1x_1 + c_2x_2) = \lambda(T(c_1x_1 + c_2x_2)) = c_1\lambda(T(x_1)) + c_2\lambda(T(x_2)) = c_1T^*(\lambda)(x_1) + c_2T^*(\lambda)(x_2).$$

  Next, to show that T^* is linear, suppose $c_1, c_2 \in K$ and $\lambda_1, \lambda_2 \in \mathcal{Y}^*$. Then for any $x \in \mathcal{X}$ we have that:
  $$T^*(c_1\lambda_1 + c_2\lambda_2)(x) = (c_1\lambda_1 + c_2\lambda_2)(T(x)) = c_1\lambda_1(T(x)) + c_2\lambda_2(T(x)) = c_1T^*(\lambda_1)(x) + c_2T^*(\lambda_2)(x) = (c_1T^*(\lambda_1) + c_2T^*(\lambda_2))(x)$$

- T^* is continuous if $\mathcal{X}^*$ and $\mathcal{Y}^*$ are equipped with the weak* topologies.

  This is because if $\lambda_\beta(y) \to \lambda(y)$ for all $y \in \mathcal{Y}$ then $\lambda_\beta(Tx) \to \lambda(\beta)(Tx)$ for all $x \in \mathcal{X}$. Hence, we have for any weak*-ly convergent net $\langle \lambda_\beta \rangle$ that $\langle T^*(\lambda_\beta) \rangle$ is also weak*-ly convergent.

- T^* is also continuous if $\mathcal{X}^*$ and $\mathcal{Y}^*$ are equipped with the strong topologies.

  This is because if T is continuous, then T maps bounded sets to bounded sets.

  Proof:
  Suppose $A \subseteq \mathcal{X}$ is bounded and let N be any neighborhood of $0 \in \mathcal{Y}$. Because T is continuous, we know that $T^{-1}(N)$ is a neighborhood of $0 \in \mathcal{X}$. And since A is bounded, there is some $r > 0$ such that $A \subseteq sT^{-1}(N)$ for all $s \in K$ with $|s| \geq r$. In turn, $T(A) \subseteq T(sT^{-1}(N)) = sT(T^{-1}(N)) = sN$ whenever $|s| \geq r$. And this proves that $T(A) \subseteq \mathcal{Y}$ is bounded.

  Thus by similar logic to the last bullet point, if $\langle \lambda_\beta \rangle$ is a strongly convergent net then $\langle T^*(\lambda_\beta) \rangle$ is also a strongly convergent net.

  As a side note, technically only the third bullet point actually required the continuity of T.

Lemma 1.2.21: If $\mathcal{X}$ and $\mathcal{Y}$ are normed vector spaces and $T \in B(\mathcal{X}, \mathcal{Y})$, then $\|T^*\|_{\mathrm{op}} = \|T\|_{\mathrm{op}}$.

Proof:
For all $x \in \mathcal{X}$ and $\lambda \in \mathcal{Y}^*$, we have that $|T^*(\lambda)(x)| = |\lambda(Tx)| \leq \|\lambda\|\|T\|\|x\|$. So $\|T^*(\lambda)\| \leq \|\lambda\|\|T\|$ for all $\lambda \in \mathcal{Y}^*$. And this shows that $\|T^*\| \leq \|T\|$.

On the other hand, for all $\varepsilon > 0$ there exists $x \in E$ such that $\|x\| = 1$ and $\|Tx\| \geq \|T\| - \varepsilon$. Also, as a consequence of the Hahn Banach theorem (see Folland theorem 5.8 in my math 240b notes), there exists $\lambda \in \mathcal{Y}^*$ such that $\lambda(Tx) = \|Tx\|$ and $\|\lambda\| = 1$. So:
$$\|T^*\| \geq \|\lambda\|^{-1}\|T^*(\lambda)\| = 1 \cdot \|T^*(\lambda)\| \geq \|x\|^{-1}|T^*(\lambda)(x)| = 1 \cdot |T^*(\lambda)(x)| = |\lambda(Tx)| = \|Tx\| \geq \|T\| - \varepsilon. \blacksquare$$

Let $\mathcal{H}$ be a real of complex Hilbert space. Then recall that there is a norm-preserving conjugate linear bijection $i : \mathcal{H} \to \mathcal{H}^*$ where we identify every $x \in \mathcal{H}$ with the linear functional $i(x) := \langle \cdot, x \rangle$. Therefore, when working on Hilbert spaces it's often convenient to just identify $\mathcal{H} \cong \mathcal{H}^*$.

As an example of this, consider any $T \in B(\mathcal{H})$ and define $T' = i^{-1} \circ T^* \circ i$. Then $T' \in B(\mathcal{H})$ as well. Also, since $i \circ T' = T^* \circ i$, we have that:

$$\langle Tx, y \rangle = (i(y))(Tx) = (T^*(i(y)))(x) = (i(T'(y)))(x) = \langle x, T'y \rangle$$

Now by a typical abuse of notation, we just say $T' \cong T^*$.

Note: if $\{e_i\}_{i \in I}$ is an orthonormal basis for $\mathcal{H}$, then:

$$T^*_{i,j} := \langle T^* e_j, e_i \rangle = \overline{\langle e_i, T^* e_j \rangle} = \overline{\langle T e_i, e_j \rangle} =: \overline{T_{j,i}}$$

So, if we "expressed T^* and T as matrices", then T^* would be the conjugate transpose of T.

We say $T \in B(\mathcal{H})$ is self-adjoint if $T^* = T$.

We say $U \in B(\mathcal{H})$ is unitary if U is an isometric isomorphism. Also, we often denote $\mathrm{Iso}(\mathcal{H})$ as $U(\mathcal{H})$ when working on Hilbert spaces.

Proposition: $U \in B(\mathcal{H})$ is unitary if and only if U is surjective and $\langle Ux, Uy \rangle = \langle x, y \rangle$ for all x, y.

($\Longleftarrow$)
We know that U is an isometry since $\|Ux\| = \langle Ux, Ux \rangle = \langle x, x \rangle = \|x\|$ for all $x \in \mathcal{H}$. This also proves that U is injective and continuous. And when we then consider that U is also surjective, we know by the open map theorem that U^{-1} is continuous. Hence $U \in U(\mathcal{H})$.

($\Longrightarrow$)
Since U is an isomorphism, we automatically know U is surjective. Meanwhile, to see that U preserves inner products, note that:

$$\langle x, y \rangle = \tfrac{1}{4}(\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2)$$

To prove this, note that:

- $$\frac{\|x+y\|^2 - \|x-y\|^2}{4} = \frac{\|x\|^2 + \langle x,y \rangle + \langle y,x \rangle + \|y\|^2 - \|x\|^2 + \langle x,y \rangle + \langle y,x \rangle - \|y\|^2}{4}$$
$$= \frac{2\langle x,y \rangle + 2\langle y,x \rangle}{4} = \frac{\langle x,y \rangle + \langle y,x \rangle}{2} = \mathrm{Re}(\langle x, y \rangle)$$
- $$\frac{\|x+iy\|^2 - \|x-iy\|^2}{4} = \frac{\|x\|^2 + \langle x,iy \rangle + \langle iy,x \rangle + \|iy\|^2 - \|x\|^2 + \langle x,iy \rangle + \langle iy,x \rangle - \|iy\|^2}{4}$$
$$= \frac{-2i\langle x,y \rangle + 2i\langle y,x \rangle}{4} = -i\frac{\langle x,y \rangle - \langle y,x \rangle}{2} = \mathrm{Im}(\langle x, y \rangle)$$

And since $\langle x, y \rangle = \mathrm{Re}(\langle x, y \rangle) + \mathrm{Im}(\langle x, y \rangle)i$, our claimed identity falls out. And now it is clear that by preserving norms U also preserves inner products. ■

Proposition: $U \in B(\mathcal{H})$ is unitary iff $U^{-1} = U^*$.

($\Longrightarrow$)

Suppose U is unitary. Then for all $x, y \in \mathcal{H}$ we have that:

- $(U^*Ux)(y) = \langle y, U^*Ux\rangle = \langle Uy, Ux\rangle = \langle y, x\rangle = x(y)$,
- $(UU^*x)(y) = \langle y, UU^*x\rangle = \langle U^{-1}y, U^*x\rangle = \langle UU^{-1}y, x\rangle = \langle y, x\rangle = x(y)$

Thus $U^*U(x) = x = UU^*(x)$ for all $x \in \mathcal{H}$. And this proves that $U^{-1} = U^*$.

($\Longleftarrow$)

Since U has an inverse, we automatically know that U is surjective. Also note that for any $x, y \in \mathcal{H}$, since $U^*Uy = y$, we have that $\langle x, y\rangle = \langle x, U^*Uy\rangle = \langle Ux, Uy\rangle$. ■

Suppose X is a measure space and let $\mathcal{H} = L^2(X)$. Then recalling *example 1.2.1* on page 284, let $\varphi \in L^\infty(X)$ and consider the linear operator $M_\varphi \in B(L^2(X))$. Then note for all $f, g \in \mathcal{H}$ that:

$$\langle g, M_\varphi^* f\rangle = \langle M_\varphi g, f\rangle = \int M_\varphi g\overline{f} = \int \varphi g\overline{f} = \int g\overline{\overline{\varphi}f} = \langle g, \overline{\varphi}f\rangle = \langle g, M_{\overline{\varphi}}f\rangle$$

This implies that $M_\varphi^* = M_{\overline{\varphi}}$. So, we are able to say that M_φ is self-adjoint iff φ is real a.e. and M_φ is unitary iff $|\varphi| = 1$ a.e.

(Definition 1.3.1:) Let $\mathcal{X}$ and $\mathcal{Y}$ be normed vector spaces. Then we define the following topologies on $B(\mathcal{X}, \mathcal{Y})$.

(a) The norm topology on $B(\mathcal{X}, \mathcal{Y})$ is the topology defined by the operator norm on $B(E, F)$.

(b) For all $x \in \mathcal{X}$ define the seminorm $p_x(T) \coloneqq \|Tx\|$ for all $T \in B(\mathcal{X}, \mathcal{Y})$. This is a well-defined seminorm and the family of these seminorms is sufficient on $B(\mathcal{X}, \mathcal{Y})$. The topology generated by that family is the strong operator topology.

(Note that strong operator convergence on $B(\mathcal{X}, \mathcal{Y})$ is equivalent to pointwise convergence on $\mathcal{X}$…)

(c) For all $x \in \mathcal{X}$ and $\lambda \in \mathcal{Y}^*$ let $p_{x,\lambda}(T) \coloneqq |\lambda(Tx)|$. By the Hahn-Banach theorem, this defines a sufficient family of seminorms. And the topology generated by that family is the weak operator topology.

(Note that weak operator convergence on $B(\mathcal{X}, \mathcal{Y})$ is equivalent to weak pointwise convergence on $\mathcal{X}$…)

Note: if $\mathcal{Y} = K$, then both the strong operator topology and the weak operator topology are just the weak* topology on $\mathcal{X}^*$.

(Example 1.3.2:) Suppose $\mathcal{H}$ is a Hilbert space with an orthonormal basis $\{e_i\}_{i \in I}$, and let $T, T_n \in B(\mathcal{H})$ for all $n \in \mathbb{N}$ with $\|T_n\|, \|T\| \leq 1$.

- If $T_n \to T$ in operator norm, then $T_n e_i \to T e_i$ uniformly over the $i \in I$.

  This is just because for all $i \in I$ we have that $\|T_n e_i - T e_i\| \leq \|T_n - T\|$ for all n.

- $T_n \to T$ operator strongly if and only if $T_n e_i \to T e_i$ for all $i \in I$.

  The $(\Longrightarrow)$ direction is obvious. To prove the converse, we need to show that if $T_n e_i \to T e_i$ for all $i \in I$, then $T_n x \to T x$ for all $x \in \mathcal{H}$. Fortunately, note that there is some countable collection $\{i_k\}_{k\in\mathbb{N}}$ such that $x = \sum_{k\in\mathbb{N}} \langle x, e_{i_k}\rangle e_{i_k}$ and the latter sum converges absolutely.

  Since T and each T_n are continuous, we have that:

  $$T(x) = T(\textstyle\sum_{k\in\mathbb{N}} \langle x, e_{i_k}\rangle e_{i_k}) = \sum_{k\in\mathbb{N}} \langle x, e_{i_k}\rangle T(e_{i_k})$$

  And similarly we have $T_n(x) = \sum_{k\in\mathbb{N}} \langle x, e_{i_k}\rangle T_n(e_{i_k})$.

  Thus $\|T_n(x) - T(x)\| \leq \sum_{k\in\mathbb{N}} |\langle x, e_{i_k}\rangle| \cdot \|T_n(e_{i_k}) - T(e_{i_k})\|$. And because $\|T_n\|, \|T\|$ are at most 1, we know $\|T_n(e_{i_k}) - T(e_{i_k})\| \leq 2$. So by standard analysis arguments you can take finitely many terms of this sum to 0 and bound the rest.

- $T_n \to T$ operator weakly if and only if $\langle T_n e_i, e_j\rangle \to \langle T e_i, e_j\rangle$ for all $i, j \in I$.

  Once again the $(\Longrightarrow)$ direction is obvious. As for the other direction, we need to show that for any $x, y \in \mathcal{H}$ we have that $\langle T_n x, y\rangle \to \langle T x, y\rangle$. If I'm inspired, I'll prove this later on *page* ___. But I'm tired. Goodnight.

10/15/2025

I need to work on math 200a again so I will be taking a break from the math 241a notes. See *page 433* to skip ahead to more functional analysis notes.

Math 200a (lectures 7-8):

Examples of uses of Sylow's theorems continued:

- Suppose p is prime and $|G| = p(p-1)$. Then there exists $P \lhd G$ such that $|P| = p$.

  By Sylow's theorems, we know that $s_p \,|\, p-1$ and $s_p \equiv 1 \pmod{p}$. Together, that tells us that $s_p = 1$. Hence, G has a unique Sylow p-subgroup which we'll call P. Also $P \lhd G$ and $|P| = p$.

- Suppose p is prime and $|G| = p(p+1)$. Then G has a normal subgroup of order either p or $p+1$.

  We may assume that $s_p \neq 1$ since otherwise we'd know that G has a unique subgroup of order p which is automatically normal.

  Now by Sylow's theorems, we have that $s_p \,|\, p+1$ (which means that $s_p \leq p+1$) and that $s_p \equiv 1 \pmod{p}$ (which means that $s_p \in \{1, p+1, 2p+1, \ldots\}$). Since we already assumed $s_p \neq 1$, this means that $s_p = p+1$. Hence, we may say that $\mathrm{Syl}_p(G) = \{P_1, \ldots, P_{p+1}\}$.

Now note that each $P_i \cong C_p$ (i.e. each P_i is cyclic with order p). As a consequence, we have that $P_i \cap P_j = \{1\}$ if $i \neq j$. So, let $X := G - (\bigcup_{i=1}^{p+1} P_i - \{1\})$. Also note that $|X| = p(p+1) - (p+1)(p-1) = p^2 + p - p^2 + 1 = p + 1$.

Note: For every finite group G, $\bigcup_{P \in \mathrm{Syl}_p(G)} P = \{x \in G : o(x) \text{ is a power of } p\}$.

To see why, first note that if $x \in P \in \mathrm{Syl}_p(G)$, then $o(x) \mid |P| = p^k$ for some $k \in \mathbb{N}$. Hence, the $\subseteq$ inclusion is obvious. Meanwhile, the other inclusion is just a direct application of Sylow's second theorem.

Hence, $X = \{x \in G : o(x) \neq p\}$. And from that we also know $\mathrm{Cl}(x) \subseteq X - \{1\}$ for all $x \in X - \{1\}$.

(As a reminder, $\mathrm{Cl}(x) := \{gxg^{-1} : g \in G\}$...)

Now by Sylow's second theorem, $p+1 = s_p = [G : N_G(P_i)]$ for all P_i. But also note that $P_i \subseteq N_G(P_i)$ and $[G : P_i] = p+1$. Hence, $N_G(P_i) = P_i$ for all $P_i \in \mathrm{Syl}_p(G)$. But note that since P_i has prime order, if $y \in P_i - \{1\}$ then $\langle y \rangle = P_i$. Also, note that for any $y \in G$ and positive integer n we have that $C_G(y) \subseteq C_G(y^n)$.

This is because if $gy = yg$, then $gy^2 = ygy = y(yg) = y^2 g$. And continuing by induction, if $gy^n = y^n g$, then $gy^{n+1} = y^n gy = y^n(yg) = y^{n+1}g$.

It follows for any $y \in P_i - \{1\}$ that the elements of $C_G(y)$ must commute with all the elements of P_i. So, $C_G(y) \subseteq N_G(P_i) = P_i$. But also since P_i is abelian (since it's cyclic), we have that $P_i \subseteq C_G(y)$. So, $C_G(y) = P_i$ for all $y \in P_i - \{1\}$.

Now from that we also know $C_G(x) \subseteq X$ for all $x \in X - \{1\}$. After all, if $x, y \in G$ then we have that $x \in C_g(y) \iff y \in C_g(x)$.

This is because $x \in C_G(y) \implies xy = yx \iff x \in C_G(y)$.

But we know that any $x \in X - \{1\}$ isn't in $C_G(y)$ for any $y \in \bigcup_{i=1}^{p+1} P_i$. Hence, $C_G(x) \subseteq X = G - \bigcup_{i=1}^{p+1} P_i$.

Now since $\mathrm{Cl}(x) \subseteq X - \{1\}$ and $C_G(x) \subseteq X$ for all $x \in X - \{1\}$, we in turn know that $|\mathrm{Cl}(x)| \leq p$ and $|C_G(x)| \leq p+1$ for all $x \in X - \{1\}$.

By the orbit stabilizer theorem (when considering the *action $G \curvearrowright G$ by conjugation*), we know $|\mathrm{Cl}(x)| = [G : C_G(x)]$. Also, by Lagrange's theorem we have that $|C_G(x)|[G : C_G(x)] = |G| = p(p+1)$. So, $|C_G(x)| \cdot |\mathrm{Cl}(x)| = p(p+1)$. And this implies that $|C_G(x)| = p+1$ and $|\mathrm{Cl}(x)| = p$ for all $x \in X - \{1\}$. Hence, $X = C_G(x)$ and $X - \{1\} = \mathrm{Cl}(x)$ for all $x \in X - \{1\}$.

This proves that X is an abelian normal subgroup of G with order $p+1$. ■

As a side note, the case where $s_p = p+1$ is actually going to be really rare. To see this, note that if p is odd, then $2 \mid p+1$ and hence there exists $x_0 \in X$ such that $o(x_0) = 2$ (by *Cauchy's theorem*). But then since $\mathrm{Cl}(x_0) = X - \{1\}$ and conjugation preserves the order of elements, we must have that $o(x) = 2$ for all $x \in X - \{1\}$. And so, by another application of Cauchy's theorem we know that $|X|$ must have no prime factor other than 2. Or in other words, $|X| = 2^n$ for some $n \in \mathbb{N}$.

This shows that in the prior example, we can only have that $s_p = p + 1$ if p is a Mersenne prime (i.e. a prime number such that $p = 2^n - 1$ for some $n \in \mathbb{N}$).

An exact sequence is a commutative diagram:

$$G_1 \xrightarrow{f_1} G_2 \xrightarrow{f_2} \dots \xrightarrow{f_k} G_{k+1}$$

where the nodes of the diagram are groups, the arrows are group homomorphisms, and $\mathrm{im}(f_i) = \ker(f_{i+1})$ for all $i \in \{1, \dots, k-1\}$.

If the first and last groups in an exact sequence are trivial, then we call that exact sequence a short exact sequence (or S.E.S.).

If G is a group and $N \lhd G$, then the standard S.E.S. is:

$$\{1\} \longrightarrow N \overset{i}{\hookrightarrow} G \overset{\pi}{\twoheadrightarrow} G/N \longrightarrow \{1\}$$

where i is the inclusion map and π is the projection map $x \mapsto xN$.

Note: $\hookrightarrow$ denotes an injective (i.e. monomorphic) homomorphism and $\twoheadrightarrow$ denotes a surjective (i.e. epimorphic) homomorphism.

Given two S.E.Ss (which for now I'll just take to have length 5), we say a homomorphism between those S.E.Ss is a ordered collection $(\theta_1, \theta_2, \theta_3)$ of group homomorphisms $\theta_i : G_i \to G_i'$ such that the diagram below commutes:

$$\begin{tikzcd}
\{1\} \arrow[r] & G_1 \arrow[r, "f_1"] \arrow[d, "\theta_1"'] & G_2 \arrow[r, "f_2"] \arrow[d, "\theta_2"'] & G_3 \arrow[r] \arrow[d, "\theta_3"'] & \{1\} \\
\{1\} \arrow[r] & G_1' \arrow[r, "f_1'"'] & G_2' \arrow[r, "f_2'"'] & G_3' \arrow[r] & \{1\}
\end{tikzcd}$$

Short Five Lemma: Suppose $(\theta_1, \theta_2, \theta_3)$ is an S.E.S. homomorphism from one S.E.S to another as shown in the above commutative diagram.

(a) If θ_1, θ_3 are injective, then so is θ_2.

Proof:
It suffices to show that $\ker(\theta_2)$ is trivial. So suppose $x_2 \in \ker(\theta_2)$. Then $\theta_3(f_2(x_2)) = f_2'(\theta_2(x_2)) = 1$. And since θ_3 is injective, this implies that $f_2(x_2) = 1$. Hence, $x_2 \in \ker(f_2) = \mathrm{im}(f_1)$.

Now pick $x_1 \in G_1$ such that $x_2 = f_1(x_1)$. Notably, $f_1'(\theta_1(x)) = \theta_2(f_1(x_1)) = 1$. So, $\theta_1(x) \in \ker(f_1')$. And since $\ker(f_1') = \mathrm{im}(\{1\} \to G_1')$ is trivial, this means that $\theta_1(x_1) = 1$. In turn, since θ_1 is injective, $x_1 = 1$. So, $x_2 = f_1(x_1) = f_1(1) = 1$.

(b) If θ_1, θ_3 are surjective, then so is θ_2.

Proof:
Let $x_2' \in G_2'$. Then since θ_3 is surjective, there exists $x_3 \in G_3$ such that $\theta_3(x_3) = f_2'(x_2')$. Also, since $\mathrm{im}(f_2) = \ker(G_3 \to \{1\}) = G_3$, we know there exists $x_2 \in G_2$ such that $f(x_2) = x_3$. And now:

$$f_2'(\theta_2(x_2)) = \theta_3(f_2(x_2)) = \theta_3(x_3) = f_2'(x_2')$$

We thus know that $\theta_2(x_2^{-1})x_2' \in \ker(f_2') = \mathrm{im}(f_1')$. Hence, there exists $x_1' \in G_1'$ such that $f_1'(x_1') = \theta_2(x_2^{-1})x_2'$. Also, since θ_1 is surjective, we know there exists $x_1 \in G_1$ such that $\theta_1(x_1) = x_1'$. And now:

$$\theta_2(f_1(x_1)) = f_1'(\theta_1(x_1)) = f_1'(x_1') = \theta_2(x_2^{-1})x_2'.$$

So $x_2' = \theta_2(x_2)\theta_2(f_1(x_1)) = \theta_2(x_2 f_1(x_1))$. ■

Note that every length five S.E.S. is isomorphic to a standard S.E.S.

Note that $\ker(f_1) = \mathrm{im}(\{1\} \to G_1) = \{1\}$ and so f_1 is injective. It follows that $G_1 \cong \mathrm{im}(f_1) = \ker(f_2)$ by the map $x \mapsto f_1(x)$. So, just define $\bar{f}_1$ to be f_1 with it's codomain restricted.

Meanwhile, note that $\mathrm{im}(f_2) = \ker(G_3 \to \{1\}) = G_3$. So, by the first isomorphism theorem we have that $G_2/\ker(f_2) \cong G_3$ via the map $x\ker(f_2) \mapsto f_2(x)$. We'll call this map $\bar{f}_2$.

Now our claim is that the following diagrams commute:

$$\begin{tikzcd}
\{1\} \arrow[r] & G_1 \arrow[r, "f_1"] & G_2 \arrow[r, "f_2"] & G_3 \arrow[r] & \{1\} \\
\{1\} \arrow[r] & \ker(f_2) \arrow[u, "\bar{f}_1^{-1}"] \arrow[r, hook, "i"'] & G_2 \arrow[u, "\mathrm{Id}"'] \arrow[r, two heads, "\pi"'] & G_2/\ker(f_2) \arrow[u, "\bar{f}_2"] \arrow[r] & \{1\}
\end{tikzcd}$$

$$\begin{tikzcd}
\{1\} \arrow[r] & G_1 \arrow[r, "f_1"] \arrow[d, "\bar{f}_1"'] & G_2 \arrow[r, "f_2"] \arrow[d, "\mathrm{Id}"'] & G_3 \arrow[r] \arrow[d, "\bar{f}_2^{-1}"'] & \{1\} \\
\{1\} \arrow[r] & \ker(f_2) \arrow[r, hook, "i"'] & G_2 \arrow[r, two heads, "\pi"'] & G_2/\ker(f_2) \arrow[r] & \{1\}
\end{tikzcd}$$

To prove this, it suffices to show that each square commutes (I'll prove this by induction after I'm done with this). Fortunately though, it is easy to see at a glance that each square commutes.

Consider the following diagram:

$$\begin{tikzcd}
A_1 \arrow[r, "f_1"] \arrow[d, "h_1"] & A_2 \arrow[r, "f_2"] \arrow[d, "h_2"] & \cdots \arrow[r, "f_{n-1}"] & A_n \arrow[r, "f_n"] \arrow[d, "h_n"] & A_{n+1} \arrow[d, "h_{n+1}"] \\
B_1 \arrow[r, "g_1"'] & B_2 \arrow[r, "g_2"'] & \cdots \arrow[r, "g_{n-1}"'] & B_n \arrow[r, "g_n"'] & B_{n+1}
\end{tikzcd}$$

To express composing two arrows i and j together (where j starts at where i ends), we write ij. Also, we can identify certain compositions of arrows with each other. For example, we always say $(ij)k = i(jk)$, thus making arrow composition associative. And hence, it is well defined to just write of a composition ijk.

Now we can also identify other arrow compositions with each other. For example, we say a diagram commutes if given any two compositions of arrows $i_1 \cdots i_r$ and $j_1 \cdots j_s$ starting and ending at the same node of our diagram we have that $i_1 \cdots i_r = j_1 \cdots j_s$.

We claim that specifically for the diagram above, the arrows in this diagram commute iff $f_i h_{i+1} = h_i g_i$ for all $i \in \{1, \dots, n\}$.

Proof:

The $(\implies)$ implication is trivial. Meanwhile, to show the other implication we proceed by induction. For our base case, we have that the claim is trivial if $n = 1$. Meanwhile, suppose we've already proven our desired claim for a diagram of the form (where $k \leq n$):

$$\begin{tikzcd}
A'_1 \arrow[r, "f'_1"] \arrow[d, "h_1"] & A'_2 \arrow[r, "f'_2"] \arrow[d, "h_2"] & \cdots \arrow[r, "f'_{k-2}"] & A'_{k-1} \arrow[r, "f'_{k-1}"] \arrow[d, "h_{k-1}"] & A'_k \arrow[d, "h_k"] \\
B'_1 \arrow[r, "g'_1"'] & B'_2 \arrow[r, "g'_2"'] & \cdots \arrow[r, "g'_{k-2}"'] & B'_{k-1} \arrow[r, "g'_{k-1}"'] & B'_k
\end{tikzcd}$$

Then in the $n+1$ case, by overlaying that smaller diagram we can show that every path from A_{k_1} or B_{k_1} to A_{k_2} or B_{k_2} commutes so long as $k_2 - k_1 < n$. Hence, we just need to show that any two walks in the diagram from A_1 to A_{n+1} or from A_1 to B_{n+1} or from B_1 to B_{n+1} are considered equivalent. But since there is only one walk from A_1 to A_{n+1} and B_1 to B_{n+1}, the only actual nontrivial thing to prove is that all walks from A_1 to B_{n+1} are considered equivalent.

So consider any walk in our diagram going from A_1 to B_{n+1}. Then we know there exists $r \in \{1, \dots, n+1\}$ such that the walk is equal to $f_1 \cdots f_{r-1} h_r g_r \dots g_n$. And then if $r \leq n$, we can say that $f_1 \cdots f_{r-1}(h_r g_r) \dots g_n = (f_1 \cdots f_{r-1} f_r g_{r+1} g_{r+1} \dots g_n)$.

By another induction argument, you can thus show that every walk from A_1 to B_{n+1} is considered equivalent to $f_1 \cdots f_n h_{n+1}$. ■

We say the following S.E.S. splits if there exists a group homomorphism $f : G_3 \to G_2$ such that $f_2 \circ f = \mathrm{Id}_{G_3}$:

$$\begin{tikzcd} 1 \arrow[r] & G_1 \arrow[r, "f_1"] & G_2 \arrow[r, shift left, "f_2"] & G_3 \arrow[l, shift left, "f"] \arrow[r] & 1 \end{tikzcd}$$

Note that we don't necessarily have that $f \circ f_2 = \mathrm{Id}_{G_2}$. After all, f_2 is not necessarily injective so it may not have a left inverse. f_2 is always surjective though so the question of whether f exists can be summed up as: does f_2 have a right inverse that's also a group homomorphism.

For more 200a notes, go to *page 326*.

Math 220a (lecture 8):

Using power series we can define more interesting holomorphic functions. For example (and I'm only doing this because I didn't take notes on this in math 140b) let:

- $\exp(z) \coloneqq \sum_{n=0}^{\infty} \frac{1}{n!} z^n$,
- $\cos(z) \coloneqq \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n}$,
- $\sin(z) \coloneqq \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1}$.

These power series have infinite radii of convergence. To see that, note by a basic induction argument that $(n+k)! \geq k^n$ for all positive integers n, k. Therefore, we can say for all $k \in \mathbb{N}$ that:

$$\limsup_{n\to\infty} \sqrt[n]{\frac{1}{n!}} \leq \limsup_{n\to\infty} \sqrt[n]{\frac{1}{k^{n-k}}} = \limsup_{n\to\infty} \frac{\sqrt[n]{k^k}}{k} = \frac{1}{k}.$$

And by taking $k \to \infty$ we get that all three power series have radius of convergence $1/0 = \infty$.

Lemma: If $G \subseteq \mathbb{C}$ is a region, $f : G \to \mathbb{C}$ is differentiable, and $f' = 0$ on G, then f equals a constant c.

Proof:
By corollary 9.20 in my math 140c notes, we know that this is true when G is convex. As for the general case, consider that every point in G has a convex neighborhood on which f is constant. So if $A \subseteq G$ is the set of points where $f(z) = f(w)$ for some arbitrary $w \in G$, then we can easily show that A is open. At the same time though, if $f(z') \neq f(w)$, then there must be a neighborhood around z' where $f \neq f(w)$. Hence, A^{C} is open too.

Since G is connected, the only way this is possible is if $A^{\mathsf{C}} = \emptyset$. ■

Proposition: $\exp(z, w) = \exp(z)\exp(w)$ for all $z, w \in \mathbb{C}$.

Proof:

Fix any $\alpha \in \mathbb{C}$ and note by product and chain rule that:

$$\frac{\mathrm{d}}{\mathrm{d}z}(\exp(z)\exp(a-z)) = \exp(z)\exp(a-z) - \exp(z)\exp(a-z) = 0$$

Thus $\exp(z)\exp(a-z)$ equals some constant. And by plugging in $z = 0$ we can then calculate that this constant is $\exp(a)$. Hence, $\exp(z)\exp(a-z) = \exp(a)$ for all $z, a \in \mathbb{C}$.

To complete the proof, set $a = z + w$. ■

Since $\exp(1) = e$ and $\exp(0) = 1$, we thus typically just use the abuse of notation that $\exp(z) = e^z$.

Also note that $e^{iz} = \cos(z) + i\sin(z)$ for all $z \in \mathbb{C}$.

I'd love to take more notes on rigorously defining $\exp$, $\cos$, and $\sin$ since I never got around to taking notes on that back when I took undergrad real analysis. But unfortunately I don't have time right now. Maybe one day in the future I'll work through it finally.

Math 200a Homework:

Set 3 Problem 2: Suppose G is a finite group, $N \lhd G$, and $P \in \mathrm{Syl}_p(N)$. Then $G = N_G(P)N$.

Proof:

Note that for any $g \in G$, we have that $gPg^{-1} \subseteq N$ since $P \subseteq N$ and N is normal. Hence $gPg^{-1} \in \mathrm{Syl}_p(N)$. But also note that by Sylow's second theorem there exists $n \in N$ such that $gPg^{-1} = nPn^{-1}$. Hence, $P = n^{-1}gPg^{-1}n = (n^{-1}g)P(n^{-1}g)^{-1}$. And this proves that $n^{-1}g \in N_G(P)$.

It follows that $g = n(n^{-1}g) \in NN_G(P)$. And since g was allowed to be anything, we've proven that $NN_G(P) = G$. Finally, $|NN_G(P)| = \frac{|N||N_G(P)|}{|N \cap N_G(P)|} = |N_G(P)N|$. And since $N_G(P)N \subseteq G$, we know that $N_G(P)N = G$. ■

Set 3 Problem 6: Suppose p and q are prime numbers and G is a group with $|G| = p^2 q$. Prove that G is not simple.

Proof:

Let s_p and s_q denote $|\mathrm{Syl}_p(G)|$ and $|\mathrm{Syl}_q(G)|$ respectively. Now by Sylow's theorems, we have that $s_p \equiv 1 \pmod{p}$ and that $s_p \in \{1, q\}$. But if $s_p = 1$, then we are already done showing that G is not simple. Hence, we can without loss of generality assume that $s_p = q$ and therefore $p \mid q - 1$.

Next, $s_q \equiv 1 \pmod{q}$ and $s_1 \in \{1, p, p^2\}$ by Sylow's theorems. But also like before, if $s_q = 1$ then we're already done showing that G is not simple. Hence, we shall assume $s_q \neq 1$. In turn, this means that either $q \mid p - 1$ (if $s_q = p$) or $q \mid p^2 - 1 = (p-1)(p+1)$ (if $s_q = p^2$). Or equivalently, this means that $q \mid p - 1$ or $q \mid p + 1$.

But note that if $q \mid p-1$, then $q+1 \leq p$. Yet this contradicts that $p \leq q-1$ (which we know since $p \mid q-1$). Hence, we must instead have that $q \mid p+1$. Firstly, this guarentees that $s_q = p^2$. Secondly, by also considering the fact that $p \mid q-1$, we know that $p+1=q$. And since p and q are both prime numbers, this must mean that $p=2$ and $q=3$.

Finally though, we now have that $|G| = 12$ and that there are $s_q(q-1) = 4(3-1) = 8$ elements of G with order 3. This is a contradiction since there aren't enough elements leftover for s_p to be greater than 1 and we already assumed $s_p = q = 3$. ■

Set 3 Problem 1: Suppose $p < q < \ell$ are three primes, G is a group, and $|G| = pq\ell$. Then G has a normal Sylow ℓ-subgroup.

Proof:

By Sylow's second theorem, we know that $|\mathrm{Syl}_\ell(G)| =: s_\ell \in \{1, p, q, pq\}$. But we also know by Sylow's third theorem that $s_\ell \equiv 1 \pmod{\ell}$. Since $1 < p, q < \ell$, this means that the only actual options that s_ℓ could be are 1 and pq. In the former case that $s_\ell = 1$, we'd already be done since the unique $L \in \mathrm{Syl}_\ell(G)$ would automatically be normal. Hence, we'll instead assume for the sake of contradiction that $s_\ell = pq$.

Next note that for any two distinct $L, L' \in \mathrm{Syl}_\ell(G)$, since L and L' are cyclic with prime order, we must have that $L \cap L' = \{1\}$. It follows that if $X = G - \bigcup_{L \in \mathrm{Syl}_\ell(G)} (L - \{1\})$ then we have that $|X| = pq\ell - pq(\ell - 1) = pq$. But also since X contains precisely the elements of G with order not equal to ℓ, we know that any Sylow q-groups must be entirely contained in X.

We now consider $|\mathrm{Syl}_q(G)| =: s_q$. By Sylow's theorems, we have that $s_1 \equiv 1 \pmod{q}$ and that $s_q \in \{1, p, \ell, p\ell\}$. But since $1 < p < q$, we automatically can rule out that $s_q = p$. By a slightly more involved argument, we can also rule out that $s_q = \ell$ or $p\ell$.

> To see why, note that for any distinct $Q, Q' \in \mathrm{Syl}_q(G)$, since Q and Q' are cyclic with prime order, we must have that $Q \cap Q' = \{1\}$. Hence, if $Y - \bigcup_{Q \in \mathrm{Syl}_q(G)} Q$, then we must have that $|Y| = s_q(q-1) + 1$.
>
> But also note that $Y \subseteq X$ and therefore $|Y| \leq |X| = pq$. Hence, we must have that $pq \geq s_1(q-1) + 1 \geq s_1 p + 1$ (where the last inequality follows since $q > p$). And thus s_1 equaling ℓ or ℓp (which are both greater than q) would be a contradiction.

It follows that $s_q = 1$ and hence there is a unique group $Q \in \mathrm{Syl}_q(G)$ which is automatically normal. And to finish off our proof, we now consider the subgroups QL_1 and QL_2 of G where L_1 and L_2 are distinct groups in $\mathrm{Syl}_\ell(G)$. Note that QL_i is a group for both i since $Q \lhd G$. Also, once again since Q and L_i are distinct cyclic groups of prime order, we know that $Q \cap L_i = \{1\}$ for both i. Hence $|QL_1| = |QL_2| = ql$.

Since $(QL_1) \cap (QL_2)$ is a subgroup of QL_1, we know $|(QL_1) \cap (QL_2)| \in \{1, q, \ell, q\ell\}$. However, we also know that $(QL_1)(QL_2) \subseteq G$ and hence:

$$|(QL_1)(QL_2)| = \frac{q^2\ell^2}{|(QL_1) \cap (QL_2)|} \leq |G| = pql.$$

Since $q^2\ell^2$, $q\ell^2$, and $q^2\ell$ are all greater than $pq\ell$, it must be that $|(QL_1) \cap (QL_2)| = q\ell$. But that implies that $QL_1 = QL_2$, which in turn gives us a different contradiction. After all, since $QL_1 = QL_2$ is a group and $L_1, L_2 \subseteq QL_1 = QL_2$, we have that $L_1 L_2 \subseteq QL_1$. However, we already went over that $L_1 \cap L_2 = \{1\}$. Hence $|L_1 L_2| = \ell^2$ and we've shown that $\ell^2 = |L_1 L_2| \leq |QL_1| = q\ell$. But that contradicts that $q < \ell$. ■

Set 3 Problem 7:

(a) Suppose $N \lhd G$ and K is a characteristic subgroup of N. Then $K \lhd G$.

Since $N \lhd G$, we know that conjugation by x is an automorphism of N for all $x \in G$. And since K is a characteristic subgroup of N, this means that $xKx^{-1} = K$ for all $x \in G$. Hence, $K \lhd G$.

(b) We say a group is characteristically simple if the only characteristic subgroups of H are 1 and H. Suppose N is a *minimal* normal subgroup of G, meaning that if $M < N$ and $M \lhd G$ then $M = \{1\}$ or N. Then N is characteristically simple.

Let M be a characteristic subgroup of N. Then by part (a) we know that $M \lhd G$. And since N is minimally normal, then means that either $M = \{1\}$ or $M = N$. ■

Math 220a Homework:

Exercise II.5.7: Let G be an open subset of $\mathbb{C}$ and P be a polygon (recall the definition on *page 247* of my journal) in G going from a to b. Then show that there is a polygon $Q \subseteq G$ from a to b which is composed of line segments which are parallel to either the real or imaginary axes.

For now, we'll just focus on the case that P is a line segment $[a, b]$. Then note that $[a, b]$ is precisely the image of the map $f(t) = tb + (1-t)a$ from $[0, 1] \subseteq \mathbb{R}$.

Since f is continuous and $[0, 1]$ is compact, it follows that $[a, b]$ is compact as well and that f is actually uniformly continuous. So firstly, for every $z \in [a, b]$ consider picking $r_z > 0$ such that the open ball $B_{r_z}(z) \subseteq G$. Then let $\mathcal{U}$ be an open cover of $[a, b]$ consisting of smaller balls: $\{B_{\frac{r_z}{3}}(z) : z \in [a, b]\}$.

By the Lebesgue number lemma, we know there is some $\varepsilon > 0$ such that whenever $w_1, w_2 \in [a, b]$ satisfy that $|w_1 - w_2| < \varepsilon$, then w_1, w_2 are contained in a single ball $B_{\frac{r_z}{3}}(z)$. And importantly in that case, if $w_1 = x_1 + iy_1$ and $w_2 = x_2 + iy_2$, then the polygon $[x_1 + iy_1, x_2 + iy_1, x_2 + iy_2]$ going from w_1 to w_2 is contained in G and clearly consists of line segments parallel to the real and imaginary axes.

Why? Since $B_{r_z}(z)$ is convex, it suffices to show that $x_2 + iy_1 \in B_{r_z}(z)$. But luckily, note that:

$$\begin{aligned}|x_2 + iy_1 - z| &= |x_2 - x_1 + x_1 + iy_1 - z| \\ &\leq |\mathrm{Re}(w_2 - w_1)| + |w_1 - z| \\ &\leq |w_2 - w_1| + |w_1 - z| \leq |w_2 - z| + 2|w_1 - z| < 3\tfrac{r_z}{3} = r_z\end{aligned}$$

Next, using the uniform continuity of f, pick $\delta > 0$ such that $|f(t_2) - f(t_1)| < \varepsilon$ when $|t_2 - t_1| \leq \delta$. In particular, this means that if $n \in \mathbb{N}$ satisfies that $n\delta \leq 1$ but $(n+1)\delta > 1$, then we apply the above observation to the points $f(0) = a, f(\delta), f(2\delta), \ldots, f(n\delta)$, $f(1) = b$ to construct a polygon from a to b contained in G which consists of $2(n+1)$ line segments parallel to either the real or imaginary axes.

To generalize this to when the polygon P isn't a single line segment, just apply the prior reasoning to each line segment making up P. ■

Exercise II.6.1: Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of uniformly continuous functions from the metric space (X, d) to the metric space (Ω, p), and suppose $f_n \to f$ uniformly. Then f is uniformly continuous.

Proof:
For any $\varepsilon > 0$ pick $n \in \mathbb{N}$ such that $p(f(x), f_n(x)) < \varepsilon/3$ for all $x \in X$. Then since f_n is uniformly continuous, pick $\delta > 0$ such that $p(f_n(x), f_n(y)) < \varepsilon/3$ whenever $d(x,y) < \delta$. Then, we can see that $p(f(x), f(y)) < \varepsilon$ whenever $d(x,y) < \delta$. Hence f is uniformly continuous.

Furthermore, if each f_n is a Lipschitz function with constant M_n and $\sup M_n < \infty$, then f is a Lipschitz function.

Proof:
Pick $M \geq \sup M_n$ and then note for all $n \in \mathbb{N}$ that:

$$\begin{aligned} p(f(x), f(y)) &\leq p(f(x), f_n(x)) + p(f_n(x), f_n(y)) + p(f(y), f_n(y)) \\ &\leq p(f(x), f_n(x)) + Md(x,y) + p(f(y), f_n(y)) \end{aligned}$$

And by taking $n \to \infty$ we get that $p(f(x), f_n(x)) \to 0$ and $p(f(y), f_n(y)) \to 0$. Hence $p(f(x), f(y)) \leq Md(x,y)$.

Finally, if $\sup M_n = \infty$ then f can fail to be Lipschitz.

Proof:
Let $X = [0, \infty)$, $\Omega = \mathbb{R}$, and define $f_n(x) := \sqrt{x + \frac{1}{n}}$ for all $x \in X$ and $n \in \mathbb{N}$. Our first claim is that $f_n \to f$ uniformly where $f(x) = \sqrt{x}$.

To see this, note that for all $x \in X$ and $n \in \mathbb{N}$ that:

$$\left(\sqrt{x + \tfrac{1}{n}} - \sqrt{x}\right)^2 = 2x + \tfrac{1}{n} - 2\sqrt{x^2 + \tfrac{x}{n}} \leq \tfrac{1}{n}.$$

Hence $|f_n(x) - f(x)| < n^{-1/2}$ for all $n \in \mathbb{N}$ and $x \in X$.

Next, we claim that each f_n is Lipschitz on X with the the constant $\frac{\sqrt{n}}{2}$.

To see this, note that $f_n'(x) = \frac{1}{2\sqrt{x + \frac{1}{n}}}$ for all $x \in X$.

It follows that $f_n'(x)$ attains a maximum of $\frac{\sqrt{n}}{2}$ at $x = 0$. And by the mean value theorem it follows that $\frac{\sqrt{n}}{2}$ is a Lipschitz constant for f on X.

That said, $\frac{\sqrt{n}}{2} \to \infty$ as $n \to \infty$. Also note that f is not Lipschitz on X.

To see this, note that f is differentiable when $x \neq 0$ and that $f'(x) = \frac{1}{2}x^{-1/2}$. But now $f'(x) \to 0$ as $x \to 0$. Hence for any $M > 0$ there is some interval $[a,b] \subseteq X$ such that $f'(x) > M$ for all $x \in [a,b]$. And in turn, by the mean value theorem we have that $|f(b) - f(a)| > M|b - a|$. So, M cannot be a Lipschitz constant for f and this proves f isn't Lipschitz.

Exercise III.1.5: If $(a_n)_{n \in \mathbb{N}}$ is a convergent sequence in $\mathbb{R}$ and $a = \lim a_n$, show that $a = \liminf a_n = \limsup a_n$.

To start off, how the hell is this a grad level problem? Like god I know the professor said that he reviews everything cause "A lOt Of PeOpLe ArE rUsTy" or something. But it's not his job to unrust us! Literally, I would argue that since math 140c is a prerequisite for this class, the professor should be obligated to assume we all have a working proficiency at undergrad real-analysis. Otherwise, why not just make the class have zero prerequisites?

Anyways the definition of $\liminf$ and $\limsup$ which Conway gives is that:

$\liminf a_n = \lim_{n \to \infty} \inf_{k \geq n} a_k$ and $\limsup a_n = \lim_{n \to \infty} \sup_{k \geq n} a_k$.

Now let $\varepsilon > 0$ and note that because $a_n \to a$ as $n \to \infty$, we know there exists $N \in \mathbb{N}$ such that $a - \varepsilon < a_n < a + \varepsilon$ for all $n \geq N$. Hence $\inf\{a_n, a_{n+1}, \ldots\} \geq a - \varepsilon$ and $\sup\{a_n, a_{n+1}, \ldots\} \leq a + \varepsilon$ for all $n \geq N$.

This in turn means that $\liminf a_n \geq a - \varepsilon$ and $\limsup a_n \leq a + \varepsilon$ for any $\varepsilon > 0$. Taking $\varepsilon \to 0$ and noting that $\liminf a_n \leq \limsup a_n$ just by comparison test, we have that:

$$a \leq \liminf a_n \leq \limsup a_n \leq a. \blacksquare$$

Exercise III.1.7: Show that the radius of convergence of the power series $f(z) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n} z^{n(n+1)}$ is 1 and discuss convergence for $z = 1, -1$, and i.

Firstly, consider the power series $g(z) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n} z^n$ since it's simpler than f. Importantly, $|\frac{(-1)^n}{n}|^{1/n} = \frac{1}{\sqrt[n]{n}} \to 1$ as $n \to \infty$. Hence, the radius of convergence of g is $1^{-1} = 1$.

This in turn also tells us that the radius of convergence of f is 1. After all, if $|z| < 1$ then $|z^{n(n+1)}| \leq |z|^n$ and so we know by comparison test with $g(|z|)$ that $f(z)$ converges. So, the radius of convergence of f is at least 1. Meanwhile, if $|z| > 1$ then $|z^{n(n+1)}| \geq |z|^n$ and so by comparison test with $g(|z|)$ we know that $f(z)$ doesn't absolutely converge. Hence, the radius of convergence is at most $|z|$ for any $z \in \mathbb{C}$ with $|z| > 1$.

Next we examine the convergence of $f(1)$, $f(-1)$, and $f(i)$.

- $f(1)$ is the alternating harmonic series. So it converges but not absolutely to $ln(2)$.
- $f(-1) = \sum_{n=1}^{\infty} \frac{1}{n}(-1)^n(-1)^{n(n+1)} = \sum_{n=1}^{\infty} \frac{1}{n}(-1)^{n(n+2)}$. But no matter if n is even or odd, $n(n+2)$ is even. So $(-1)^{n(n+2)} = 1$ and thus $f(-1)$ is the harmonic series which diverges.
- $f(i) = \sum_{n=1}^{\infty} \frac{1}{n} i^{2n} i^{n(n+1)} = \sum_{n=1}^{\infty} \frac{1}{n} i^{n(n+3)} = \sum_{n=1}^{\infty} \frac{1}{n}(-1)^{\psi(n)}$ where $\psi(n) = 0$ if $n \equiv 0$ or $1 \pmod 4$ and $\psi(n) = 1$ otherwise. In other words:

$$f(i) = 1 - \tfrac{1}{2} - \tfrac{1}{3} + \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} - \cdots$$

Now note that the partial sums $\sum_{k=1}^{n} (-1)^{\psi(k)}$ are bounded between -1 and 1. Also, $(\frac{1}{n})_{n \to \infty}$ is a decreasing sequence of nonnegative numbers converging to 0. Therefore, by the result below from my math 140a notes we know that $f(i)$ converges (although again not absolutely).

Proposition 57: If the partial sums of Σa_n are bounded and we have a sequence $b_0 \geq b_1 \geq b_2 \geq \cdots$ such that $b_n \to 0$, then $\sum a_n b_n$ will converge.

Proof:

Set $A_n = \sum_{k=0}^{n} a_k$. Then pick $M > 0$ such that $\forall n, \;\; |A_n| < M$.

Given $\varepsilon > 0$, pick N with $b_N < \frac{\varepsilon}{2M}$. Then when $q \geq p \geq N$, we have:

$$\begin{aligned}\left|\sum_{n=p}^{q} a_n b_n\right| &= \left|\sum_{n=p}^{q-1} A_n(b_n - b_{n+1}) + A_q b_q - A_{p-1} b_p\right| \\ &\leq \sum_{n=p}^{q-1} |A_n|\,(b_n - b_{n+1}) + |A_q| b_q + |A_{p-1}| b_p \\ &\leq M(b_p - b_q + b_q + b_p) = 2M b_p \leq 2M b_N < \varepsilon\end{aligned}$$

Exercise III.2.1: Show that $f(z) = |z|^2$ is complex differentiable only at the origin.

Identify $\mathbb{C}$ with $\mathbb{R}$ and consider f as the function $f(x,y) = (x^2+y^2, 0)$ going from $\mathbb{R}^2$ to $\mathbb{R}^2$. Then $f \in C^\infty(\mathbb{R}^2)$ with a derivative matrix $\left(\begin{smallmatrix} 2x & 2y \\ 0 & 0 \end{smallmatrix}\right)$. Now for f to satisfy the Cauchy-Riemann equations (see the theorem on *page 296*) at a point (x,y), we must have that $2x = 0$ and $-2y = 0$. Hence the only point where f is complex differentiable is at $(0,0) = 0 + i0$.

Exercise III.2.3: Show that $\lim_{n\to\infty} n^{1/n} = 1$.

If you want to prove this theorem without using logarithms (because you hadn't defined logarithms yet when you first relied on this fact), then here is the proof from math 140a:

(C) $\sqrt[n]{n} \to 1$

Proof:
Let $x_n = \sqrt[n]{n} - 1$. Then $x_n \geq 0$ and by the binomial theorem:

$$\tfrac{n(n-1)}{2}(x_n)^2 \leq \sum_{k=0}^{n} \tbinom{n}{k}(x_n)^k = (x_n+1)^n = n$$

Then we have that $0 \leq x_n \leq \sqrt{\frac{2n}{n(n-1)}} = \sqrt{\frac{2}{n-1}}$ when $n \geq 2$.

Now, $\sqrt{\frac{2}{n-1}} \to 0$.

Proof: Let $\varepsilon > 0$. Then $\sqrt{\frac{2}{n-1}} < \varepsilon$ whenever $n > 1 + \frac{2}{\varepsilon^2}$.

Therefore, by proposition 43, we know that $x_n \to 0$. So finally, we conclude that:

$$\sqrt[n]{n} \to \lim_{n\to\infty}(x_n) + 1 = 0 + 1$$

If you are willing to rely on logarithms and calculus though, then here is a slicker proof:

Note that $\log(n^{1/n}) = \frac{1}{n}\log(n)$ for all n. Then by L'Hôpital's rule we have that $\lim_{x\to\infty} x^{-1}\log(x) = \lim_{x\to\infty}(1)^{-1}\frac{1}{x} = 0$. And hence $\log(n^{1/n}) \to 0$ as $n \to \infty$.

Finally, since $\exp$ is continuous, we have that $n^{1/n} = \exp(\log(n^{1/n})) \to \exp(0) = 1$ as $n \to \infty$.

Exercise III.2.19: Let G be a region and define $G^* = \{z : \overline{z} \in G\}$. If $f : G \to \mathbb{C}$ is holomorphic prove that $f^* : G^* \to \mathbb{C}$ defined by $f^*(z) = \overline{f(\overline{z})}$ is also holomorphic.

Once again identify $\mathbb{C}$ with $\mathbb{R}^2$ and write f as $f(x,y) = (u(x,y), v(x,y))$. Then we have that $f^*(x,y) = (u(x,-y), -v(x,-y))$. And since f is C^1, we can calculate that the derivative matrix of f^* is $\mathrm{D}(f^*) = \begin{pmatrix} u_x(x,-y) & -u_y(x,-y) \\ -v_x(x,-y) & v_y(x,y) \end{pmatrix}$.

Firstly, this shows that f^* is also C^1 since all the partial derivatives of f^* are continuous. Also, this shows that if f satisfies the Cauchy-Riemann ewquations, then so does f^*. Hence f being holomorphic on G implies f^* is as well.

10/17/2025

So for functional analysis I need some fixed point theorems that would normally have been taught in a topology class. But I've never taken an actual topology class. Hence, my goal for today is to prove those theorems. First, I will be taking notes from the paper *Brouwer's Fixed Point Theorem* by Jasmine Katz. (See the *17th entry* of the bibliography...)

Katz defines a convex body in $\mathbb{R}^n$ to be a set $X \subseteq \mathbb{R}^n$ that is compact, convex, and has a nonempty interior (relative to the Euclidean topology of all of $\mathbb{R}^n$).

Also, for $m \geq n \geq 1$ suppose we are given any $n+1$ points $p_0, p_1 \ldots, p_n \in \mathbb{R}^m$ in general linear position (see my math 190b notes). Then we say the convex hull of those points (i.e. $\mathrm{conv}(p_0, p_1 \ldots, p_n)$) is an n-simplex. Recall from the notes I took before I dropped math 190b last Spring that all n-simplices are compact and homeomorphic to each other.

Katz in particular denotes $\Delta_0^n := \mathrm{conv}(e_1, \ldots, e_n, -\sum_{i=1}^n e_i)$ where $e_1, \ldots, e_n$ are the standard basis vectors of $\mathbb{R}^n$. Also, Katz denotes Δ^n to be the standard n-simplex $\mathrm{conv}(u_1, \ldots, u_{n+1})$ where $u_1, \ldots, u_{n+1}$ are the standard basis vectors of $\mathbb{R}^{n+1}$.

Proposition 3.2: Δ_0^n is a convex body in $\mathbb{R}^n$.

Proof:

The only thing not trivial from the definition of Δ_0^n is that Δ_0^n has a nonempty interior. So to prove this, first define for each $k \in \mathbb{N}$ the $(k+1)$ by $(k+1)$ matrices:

$$A_k := \begin{pmatrix} \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} & \begin{matrix} -1 \\ \vdots \\ -1 \end{matrix} \\ \begin{matrix} 1 & \cdots & 1 \end{matrix} & 1 \end{pmatrix} \text{ and } B_k := \begin{pmatrix} \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} & \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} \\ 1 & \begin{matrix} 1 & \cdots & 1 \end{matrix} \end{pmatrix}$$

We can calculate the determinant of A_k as follows:

Clearly $\det(A_1) = \det(\left[\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right]) = 2$ and $\det(B_1) = \det(\left[\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right]) = -1$.

Meanwhile, by considering the Laplace expansions going along the top rows of A_k and B_k where $k > 1$, we get that:

$$\det(A_k) = \det(A_{k-1}) + (-1)^{k+3}\det(B_{k-1}) \text{ and } \det(B_k) = -\det(B_{k-1})$$

From there it is easy to see that $\det(B_k) = (-1)^k$. And hence:

$$\begin{aligned} \det(A_k) = \det(A_{k-1}) + (-1)^{(k+3)+(k-1)} &= \det(A_{k-1}) + (-1)^{2(k+1)} \\ &= \det(A_{k-1}) + 1 \\ &= \det(A_{k-2}) + 2 \\ &\vdots \\ &= \det(A_1) + (k-1) = k+1 \end{aligned}$$

In particular, we now know that A_n is invertible since it has nonzero determinant. And this is important because we can now say that $x = (x_1, \ldots, x_n) \in \Delta_0^n$ if and only if:

$$f(x_1, \ldots, x_n) := (A_n)^{-1} \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ 1 \end{bmatrix} \in [0, \infty)^{n+1}$$

But f is a continuous injection and $f(0) = (\frac{1}{n+1}, \dots, \frac{1}{n+1})$ since:

$$0 = \sum_{i=1}^{n} \frac{1}{n+1} e_i - \frac{1}{n+1}(\sum_{i=1}^{n} e_i).$$

Hence, there must be some $\delta > 0$ such that $\|f(x) - f(0)\|_2 < \frac{1}{n+1}$ when $\|x\|_2 < \delta$. And in turn, the open ball of radius δ about 0 is contained in Δ_0^n. ■

Katz defines a ray from $x_0 \in \mathbb{R}^n$ to be a set $\{x_0 + ty : t \geq 0\}$ where $y \in \mathbb{R}^n$ with $\|y\|_2 = 1$.

Here is a basic topology fact I somehow haven't proved before:

Proposition: Suppose X is a topological space, $E \subseteq X$ is connected, and $A \subseteq X$ satisfies that $E \cap A \neq \emptyset$ and $E \cap A^{\mathsf{C}} \neq \emptyset$. Then $E \cap \partial A \neq \emptyset$.

Proof:

Note that $\partial A = \overline{A} \cap \overline{A^{\mathsf{C}}}$. So if $E \cap \partial A = \emptyset$, then this implies that $\overline{A} \cap E$ and $\overline{A^{\mathsf{C}}} \cap E$ are two disjoint nonempty closed sets (in the subspace topology of E) whose union is all of E. But this contradicts that E is connected. ■

Lemma 3.6: Suppose $X \subseteq \mathbb{R}^n$ is a convex body with $0 \in X^\circ$. Then every ray from 0 intersects ∂X exactly once.

Proof:

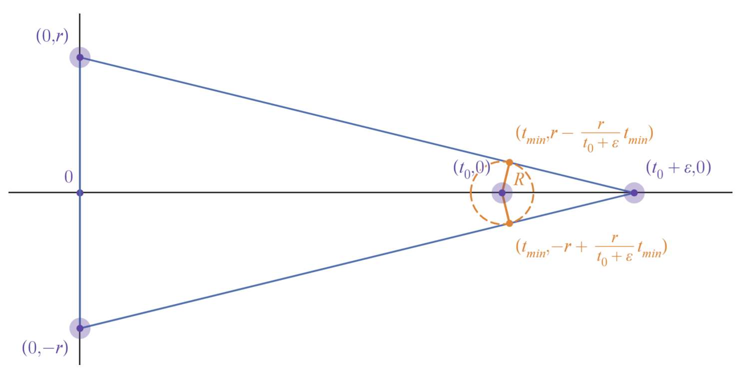
Fix $y \in \mathbb{R}^n$ such that $\|y\| = 1$ and let $f(t) = ty$. Then we want to show there is a unique $t_0 \geq 0$ such that $f(t_0) \in \partial X$. Fortunately, it's easy to see that $f(t)$ intercepts ∂X at least once. After all, $f(0) \in X$. But at the same time, since X is compact and thus bounded, we know that there is some $t' > 0$ such that $f(t') \in X^{\mathsf{C}}$. Since the ray traced by f is connected, it now follows that there exists some $t_0 > 0$ such that $f(t_0) \in \partial X$.

We now show that our t_0 is unique. Suppose for the sake of contradiction that there exists $\varepsilon > 0$ with $(t_0 + \varepsilon)y \in X$. Then pick $r > 0$ such that the ball $\overline{B_r(0)} \subseteq X$. Since X is convex, we know that any line segment from $(t_0 + \varepsilon)y$ to a point in $\overline{B_r(0)}$ is contained in X. And thus, while it is fucking awful to calculate, we have that the open ball of radius $R = \min(t_0, \sqrt{Q(t_{\min})})$ about $t_0 y$ is contained in X where:

$$Q(t) = (r - \frac{r}{t_0+\varepsilon}t)^2 + (t - t_0)^2 \text{ and } t_{\min} = \frac{\frac{r^2}{t_0+\varepsilon} + t_0}{\frac{r^2}{(t_0+\varepsilon)^2} + 1}$$

For a hint at how I got that fucked up radius R, observe the diagram to the right:

However, the fact we were able to find such an R contradicts that $t_0 y \in \partial X$. Hence, our supposed ε can't exist.



Since X is closed (since it's compact) and thus includes it's boundary, we've now proven that if $f(t_0) \in \partial X$ then $f(t) \notin \partial X$ for all $t > t_0$. And hence, our ray can intercept ∂X at most once. ■

As a side note: something even more fucked up is that the radius proposed by this paper doesn't work. It's too large. Anyways, because the paper handwaves this next bit, I'm going to deviate from the paper a bit.

Lemma: Let $\mathcal{X}$ be a topological K-vector space where $K = \mathbb{R}$ or $\mathbb{C}$. Also suppose $E \subseteq X$ and $c \in K$.

- If E is convex, then so is cE.

  Proof:
  If E is empty or $c = 0$, this is obvious. Otherwise, suppose $x, y \in cE$ and then note that $c^{-1}x, c^{-1}y \in E$. It follows that $c^{-1}(tx+(t-1)y) \in E$. So, $tx+(1-t)y \in cE$.

- $c(\overline{E}) = \overline{cE}$.

  Proof:
  If $c = 0$ then it must be the case that both sets are empty or are $\{0\}$. Meanwhile, note that as a general fact if $f : X \to Y$ is a continuous map and $A \subseteq X$, then $f^{-1}(\overline{f(A)})$ is a closed set containing A. Hence $\overline{A} \subseteq f^{-1}(\overline{f(A)})$ and this proves that $f(\overline{A}) \subseteq \overline{f(A)}$. Since scalar multiplication by c is a homeomorphism of $\mathcal{X}$ when $c \neq 0$, we thus have that $c\overline{E} = \overline{cE}$. ■

- $c(\partial E) = \partial(cE)$.

  Proof:
  By the last bullet point plus the fact that $f(A \cap B) = f(A) \cap f(B)$ for any function, we know that $c(\partial E) = c(\overline{E} \cap \overline{E^{\mathsf{C}}}) = \overline{cE} \cap \overline{c(E^{\mathsf{C}})}$. Now if $c = 0$, then we know that $c(\partial E)$ and $\partial(cE)$ are either both empty or both $\{0\}$. So, we may assume that $c \neq 0$. But in that case scalar multiplication is a homeomorphism on $\mathcal{X}$. So $c(E^{\mathsf{C}}) = (cE)^{\mathsf{C}}$ and we've shown that $c(\partial E) = \overline{cE} \cap \overline{(cE)^{\mathsf{C}}} = \partial(cE)$.

Lemma: Suppose $f : X \to Y$ is a homeomorphism and $A, B \subseteq X$. (This would include the case that $X = Y = \mathcal{X}$ and f is scalar multiplication by a nonzero scalar). Then $f(E^\circ) = (f(E))^\circ$.

Proof:
Since f is continuous and bijective, $f^{-1}(f(A)^\circ)$ is an open set contained in A. So, $f^{-1}(f(A)^\circ) \subseteq A^\circ$ and hence $f(A)^\circ \subseteq f(A^\circ)$. On the other hand, since f^{-1} is continuous, we know that $f(A^\circ)$ is an open subset of $f(A)$. So, $f(A^\circ) \subseteq f(A)^\circ$.

Theorem: Suppose $X \subseteq \mathbb{R}^n$ is a convex body and contains 0. Then X is homeomorphic to the closed unit ball $\overline{B_1(0)}$.

Proof:

Let p_{X° be the Minkowski functional associated with X°. In other words, $p_{X^\circ}(x) = \inf\{t \geq 0 : x \in tX\}$. Unfortunately, since X isn't a balanced set necessarily, we can't directly apply the proof on page 233 to get that $P_{X^\circ}(cx) = |c|x$ for all $c \in \mathbb{R}$ and $x \in \mathbb{R}^n$. That said, since X is convex and a neighborhood of 0, we are able to copy the reasoning that shows that p_{X° is well-defined and satisfies the triangle inequality.

Claim 1: Suppose y is any unit vector and t_0 is the unique nonnegative number satisfying that $t_0 y \in \partial X$. Then $p_{X^\circ}(ty) = {}^t\!/\!_{t_0}$.

Proof:

If $t = 0$, then it is obvious that $p_{X^\circ}(ty) = 0 = {}^t\!/\!_{t_0}$. As for when $t \neq 0$, first note that $t_0 y \in \partial X \Longleftrightarrow bt_0 y \in \partial(bX)$ for all $b \geq 0$. Also note that bX is easily checked to be convex body when $b \neq 0$ (and so we can apply theorem 3.6 to bX).

Suppose $b < {}^t\!/\!_{t_0}$. Then $bt_0 < t$ and thus $t \notin X$ since $bt_0 y \in \partial(bX)$. On the other hand, suppose $b > {}^t\!/\!_{t_0}$. Then $bt_0 > t$ and thus $t \in X - \partial X = X^\circ$ since again $bt_0 y \in \partial(bX)$. It follows that $p_{X^\circ}(ty) = {}^t\!/\!_{t_0}$.

Corollary: Suppose $x \in X$ and $t_0 > 0$ is the unique nonnegative number such that $t_0 \frac{x}{\|x\|_2} \in \partial X$. Then $p_{X^\circ}(x) = {}^{\|x\|_2}\!/\!_{t_0}$.

Claim 2: p_{X° is continuous.

Proof:

While we don't necessarily have reverse triangle inequality since $p_{X^\circ}(x) \neq p_{X^\circ}(-x)$, we can at least prove using just triangle inequality that:

$$|p_{X^\circ}(x) - p_{X^\circ}(y)| \leq \max(p_{X^\circ}(x-y), p_{X^\circ}(y-x))$$

Then, you just need to follow almost identical reasoning to that of page 233 to show that $p_{X^\circ}(x)$ is continuous.

At last we are ready to write our homeomorphism from X to B. Define $g : X \rightarrow \overline{B_1(0)}$ by $g(x) = p_{X^\circ}(x)\frac{x}{\|x\|_2}$ when $x \neq 0$ and $g(0) = 0$. And similarly, define $h : \overline{B_1(0)} \rightarrow X$ by $h(y) = ty$ such that $ty \in \partial(\|y\|X)$.

- g is continuous when $x \neq 0$ by virtue of being a scalar product of continuous functions. Also, given any sequence $(x_n)_{n \in \mathbb{N}}$ in X converging to 0 we have that $\|g(x_n)\|_2 = p_{X^\circ}(x_n) \cdot 1 \rightarrow 0$ as $n \rightarrow \infty$. So g is also continuous at 0.

- $h(g(0)) = 0$. Meanwhile, suppose $x \in X - \{0\}$. Then let t_0 be the unique nonnegative number satisfying that $t_0 \frac{x}{\|x\|_2} \in \partial X$. Now $g(x) = p_{X^\circ}(x)\frac{x}{\|x\|_2} = \frac{x}{t_0}$. Thus $h(g(x)) = t\frac{x}{t_0}$ where $t \geq 0$ satisfies that $t\frac{x}{t_0} \in \partial(\frac{\|x\|}{t_0}X)$. Or in other words, $t\frac{x}{\|x\|_2} \in \partial X$. It follows by theorem 3.6 that $t = t_0$ and so $h(g(x)) = x$.

- $g(h(0)) = g(0) = 0$. Meanwhile, if $y \in \overline{B_1(0)} - \{0\}$ then let t_0 be the unique nonnegative number such that $t_0 \frac{y}{\|y\|_2} \in \partial X$. Like before, we can show that $h(y) = t_0 y$. And now $p_{X^\circ}(t_0 y) = \frac{t_0\|y\|_2}{t_0} = \|y\|_2$ since $t_0 y = t_0\|y\|_2 \frac{y}{\|y\|_2}$. Thus $g(t_0 y) = g(t_0 y) = \|y\|_2 \frac{t_0 y}{\|t_0 y\|_2} = y$ and we've proven that $g(h(y)) = y$.

- Since $h = g^{-1}$, g is continuous, X is compact, and $\overline{B_1(0)}$ is Hausdorff, we have that h is continuous. ■

Corollary: All convex bodies in $\mathbb{R}^n$ are homeomorphic to the closed unit ball $\overline{B_1(0)}$.
Why? We can just translate them so that they contain 0 and then apply the last theorem.

Corollary 2: All n-simplices are homeomorphic to the closed unit ball $\overline{B_1(0)}$ in $\mathbb{R}^n$.

10/18/2025

I need to first do the rest of my math 200a homework. Then I think there is a different paper I want to try to following to prove the fixed point theorems from before.

Math 200a Homework:

Set 3 Problem 3: Suppose G is a finite group and $H < G$. Suppose also for all $x \in H - \{1\}$ that $C_G(x) \subseteq H$. Then $\gcd(|H|, [G:H]) = 1$.

Proof:
If $\gcd(|H|, [G : H]) \neq 1$, then we know there is some prime number p dividing both $|H|$ and $[G : H]$. And as a result, $\nu_p(|H|) < \nu_p(|G|)$. So, by choosing $P \in \mathrm{Syl}_p(H)$ and $Q \in \mathrm{Syl}_p(G)$, we know that $|P| < |Q|$. Also, by Sylow's second theorem, we can conjugate Q in order to say without loss of generality that $P \subseteq Q$. We'll need two observations:

- Since the order of Q doesn't divide H, we know that Q isn't a subgroup of H. Hence, $Q \cap (G - H) \neq \emptyset$.
- At the same time though, we know that $P < Q \cap H < H$. And since the only prime factors of $|Q \cap H|$ are p, this tells us by Lagrange's theorem that $|P| = p^{\nu_p(|H|)}$ divides $|Q \cap H|$ which itself divides $p^{\nu_p(|H|)}$. Hence, $|P| = |Q \cap H|$ and this proves that $Q \cap H = P$.

Now recall from the 2nd proposition on *page 272* that if $\{1\} \neq N \lhd P$ and P is a p-group, then $Z(P) \cap N \neq \{1\}$.

Side note: Whenever $H < G$, we define $Z(H) = \bigcap_{h \in H} C_H(h)$. Just wanted to make that's clear since I didn't know better before.

As a special case, if we set $N = P$ (where P is nontrivial), then this proposition says that the center of a p-group is always nontrivial. Hence, there exists $y \in Z(P) - \{1\}$.

Side note: How the hell hadn't I processed that consequence of the theorem we proved before?

Now note that $Z(Q) \subseteq C_Q(y) \subseteq C_G(y) \subseteq H$ where the last inclusion is by the assumption of the problem. Hence $Z(Q) \subseteq H \cap Q = P$. At the same time, we know that $Z(Q)$ is nontrivial. So, there exists $x \in Z(Q) - \{1\}$ which we know from before is in $P \subseteq H$. Finally, we now have that $Q \subseteq C_Q(x) \subseteq C_G(x) \subseteq H$. But this contradicts that $Q \cap (G - H) \neq \emptyset$. ■

Set 3 Problem 8: Suppose G is a finite group.

(a) Prove that a normal Sylow p-subgroup is a characteristic subgroup.

A Sylow p-subgroup P is normal if and only if it is the only Sylow p-subgroup. And since automorphisms preserve the order of subgroups, we have that if P is the only subgroup of G with order $p^{\nu_p(|G|)}$ then $\theta(P) = P$ for all $\theta \in \mathrm{Aut}(G)$. So, P is a characteristic subgroup.

(b) Suppose $H \lhd G$ and $\gcd(|H|, [G : H]) = 1$. Prove that H is a characteristic subgroup.

Suppose $\theta \in \mathrm{Aut}(G)$. Then since H is normal, we know that $\theta(H)H$ is a subgroup of G and $H \lhd \theta(H)H$. So by the correspondance theorem, we know that $|\frac{\theta(H)H}{H}|$ divides $|G/H| = [G : H]$. At the same time, $|\theta(H)H| = \frac{|H||\theta(H)|}{|H\cap\theta(H)|}$. So, $|\frac{\theta(H)H}{H}| = \frac{|\theta(H)|}{|H\cap\theta(H)|}$ divides $|\theta(H)| = |H|$.

This shows that $|\frac{\theta(H)H}{H}| \mid \gcd(|H|, [G : H]) = 1$. And hence, $\theta(H)H = H$. Since $|\theta(H)| = |H|$ and $\theta(H) \subseteq \theta(H)H$, we can thus conclude that $\theta(H) = H$. ■

Set 3 Problem 4: Suppose G is a finite group, $N \lhd G$, and p is a prime factor of $|N|$.

(a) Suppose $P \in \mathrm{Syl}_p(G)$ and $Q \in \mathrm{Syl}_p(N)$. Prove there exists $g \in G$ such that $Q = (gPg^{-1}) \cap N$.

Uhhhh, I think I already basically proved this fact when doing problem 3.

By Sylow's second theorem, we know there is some $g \in G$ such that $Q \subseteq gPg^{-1}$. Then since $(gPg^{-1}) \cap N$ is a subgroup of gPg^{-1} where the latter is a p-group, we know that $(gPg^{-1}) \cap N$ is also a p-group. Also, since $(gPg^{-1}) \cap N$ is a subgroup of N, we know that $|(gPg^{-1}) \cap N|$ divides $|N|$. It follows that $|(gPg^{-1}) \cap N|$ divides $p^{\nu_p(|N|)}$. At the same time, since $|Q| = p^{\nu_p(|N|)}$ and $Q \subseteq gPg^{-1} \cap N$, we have that $p^{\nu_p(|N|)} \leq |(gPg^{-1}) \cap N|$.

So, $Q \subseteq (gPg^{-1}) \cap N$ and $|Q| = p^{\nu_p(|N|)} = |(gPg^{-1}) \cap N|$. It follows that $Q = (gPg^{-1}) \cap N$.

Side note: none of the reasoning in this part requires having N be normal.

(b) Prove that the following is a well-defined surjective function:

$\Phi : \mathrm{Syl}_p(G) \to \mathrm{Syl}_p(N)$ where $\Phi(P) = P \cap N$.

Part (a) guarentees that the above map will be surjective provided that it is well-defined (since $gPg^{-1} \in \mathrm{Syl}_p(G)$ for any $P \in \mathrm{Syl}_p(G)$). Hence, all we need to show is that $P \cap N \in \mathrm{Syl}_p(N)$ whenever $P \in \mathrm{Syl}_p(G)$.

Fortunately since $P \cap N < P$, we know that $P \cap N$ is a p-group. But suppose for the sake contradiction that $|P \cap N| \neq p^{\nu_p(|N|)}$ (meaning $P \cap N \notin \mathrm{Syl}_p(G)$). Then, by Sylow's first and second theorems there exists a group $Q \in \mathrm{Syl}_p(N)$ such that $P \cap N \lneqq Q < N$. Also, by part (a) there exists $g \in G$ such that $Q = gPg^{-1} \cap N$. But since N is normal, $gPg^{-1} \cap N = gPg^{-1} \cap gNg^{-1} = g(P \cap N)g^{-1}$.

The last equality is obvious if you think about it. So I'd rather not write out a proof since I want to go on an outing soon.

So $P \cap N \lneqq Q = g(P \cap N)g^{-1}$. Since $|P \cap N| = |g(P \cap N)g^{-1}|$ this is a contradiction.

(c) For $P \in \mathrm{Syl}_p(G)$, prove that $N_G(P) \subseteq N_G(\Phi(P))$ and:
$$|\Phi^{-1}(\Phi(P))| = [N_G(\Phi(P)) : N_G(P)].$$

To start off, note that:
$$\begin{aligned}\Phi(xPx^{-1}) = xPx^{-1} \cap N &= xPx^{-1} \cap xNx^{-1} \\ &= x(P \cap N)x^{-1} = x\Phi(P)x^{-1}.\end{aligned}$$

Therefore, if $x \in N_G(P)$ then $\Phi(P) = \Phi(xPx^{-1}) = x\Phi(P)x^{-1}$. And this proves the first claim that $N_G(P) \subseteq N_G(\Phi(P))$.

Next, note by Sylow's second theorem that:
$$\Phi^{-1}(\Phi(P)) = \{gPg^{-1} : g \in G \text{ and } \Phi(P) \subseteq gPg^{-1}\}.$$

But $\Phi(P) \subseteq gPg^{-1}$ if and only if:
$$\begin{aligned}\Phi(P) = P \cap N \subseteq gPg^{-1} \cap N &= (gPg^{-1}) \cap (gNg^{-1}) \\ &= g(P \cap N)g^{-1} = g\Phi(P)g^{-1}.\end{aligned}$$

And since $|\Phi(P)| = |g\Phi(P)g^{-1}|$, this is the same as saying that $\Phi(P) = g\Phi(P)g^{-1}$. So, we really have that:
$$\Phi^{-1}(\Phi(P)) = \{gPg^{-1} : g \in N_G(\Phi(G))\}.$$

Hence, if we consider the action $N_G(\Phi(P)) \curvearrowright \mathrm{Syl}_p(G)$ by conjugation, then $\Phi^{-1}(\Phi(P))$ is precisely the orbit of P with respect to this action. Also, x is in the stabilizer of P with respect to this action precisely when $xPx^{-1} = P$. Or in other words, $x \in (N_G(\Phi(P)))_P$ when $x \in N_G(\Phi(P)) \cap N_G(P) = N_G(P)$. By the orbit-stabilizer theorem, we thus can conclude that:
$$|\Phi^{-1}(\Phi(P))| = [N_G(\Phi(P)) : N_G(P)].$$

(d) Prove that $|\mathrm{Syl}_p(N)|$ divides $|\mathrm{Syl}_p(G)|$.

We know that $\{\Phi^{-1}(Q) : Q \in \mathrm{Syl}_p(N)\}$ forms a partion of $\mathrm{Syl}_p(G)$. Hence, we now seek to prove that there exists a single integer $r \in \mathbb{N}$ such that $|\Phi^{-1}(Q)| = r$ for all $Q \in \mathrm{Syl}_p(N)$. After all, once we know that we will then be able to say that $|\mathrm{Syl}_p(N)| = r|\mathrm{Syl}_p(G)|$.

Luckily, note that since conjugation is an automorphism of G we have that:

- $|N_G(P)| = |g(N_G(P))g^{-1}| = |(N_G(gPg^{-1}))|$,
- $|N_G(\Phi(P))| = |g(N_G(\Phi(P)))g^{-1}| = |N_G(g\Phi(P)g^{-1})| = |N_G(\Phi(gPg^{-1}))|$.

It follows for all $g \in G$ and $P \in \mathrm{Syl}_p(G)$ that:
$$\begin{aligned}|\Phi^{-1}(\Phi(P))| &= [N_G(\Phi(P)) : N_G(P)] \\ &= \tfrac{|N_G(\Phi(P))|}{|N_G(P)|} = \tfrac{|N_G(\Phi(gPg^{-1}))|}{|N_G(gPg^{-1})|} = [N_G(\Phi(gPg^{-1})) : N_G(gPg^{-1})] \\ &= |\Phi^{-1}(\Phi(gPg^{-1}))|\end{aligned}$$

And since G acts transitively on $\mathrm{Syl}_p(G)$ by conjugation, this means that there exists $r \in \mathbb{N}$ such that $r = |\Phi^{-1}(\Phi(P))|$ for all $P \in \mathrm{Syl}_p(G)$. And since Φ is surjective, we have that $\Phi^{-1}(Q) = r$ for all $Q \in \mathrm{Syl}_p(G)$. ■

I'm now going to follow the paper: *The Milnor-Rogers Proof of the Brouwer Fixed Point Theorem*, by Ralph Howard (see bibliography *item 18*).

For ease of notation, I will denote $B^n := \{x \in \mathbb{R}^n : \|x\|_2 < 1\}$, and $S^{n-1} = \partial B^n$. Also, let $e_1, \ldots, e_n$ be the standard basis vectors for $\mathbb{R}^n$.

Lemma: Suppose $f : \overline{B^n} \to \mathbb{R}^m$ is C^1. As a reminder, this means there is an open set $\overline{B^n} \subseteq U \subseteq \mathbb{R}^n$ and a C^1 function $g : U \to \mathbb{R}^m$ such that $g|_{\overline{B^n}} = f$. We claim that if $x = (x_1, \ldots, x_n) \in S^{n-1}$, then $\frac{\partial}{\partial x_j} g(x)$ is entirely determined by f.

Proof:
To start off, note that $\|x + he_j\|_2^2 = \|x\|_2^2 + 2(x \cdot he_j) + h^2\|e_j\|_2^2 = 1 + 2hx_j + h^2$. Thus $\|x + he_j\|_2 \leq 1$ when $h(2x_j + h) \leq 0$. And so either $h \in (-2x_j, 0]$ or $h \in [0, -2x_j)$. It follows that:

- if $x_j > 0$, then $\frac{\partial}{\partial x_j} g(x) = \lim\limits_{h \to 0} \frac{g(x+he_j) - g(x)}{h} = \lim\limits_{h \to 0^-} \frac{f(x+he_j) - f(x)}{h}$;
- if $x_j < 0$, then $\frac{\partial}{\partial x_j} g(x) = \lim\limits_{h \to 0} \frac{g(x+he_j) - g(x)}{h} = \lim\limits_{h \to 0^+} \frac{f(x+he_j) - f(x)}{h}$.

This shows that so long as $x_j = (x \cdot e_j) \neq 0$, we have that $\frac{\partial}{\partial x_j} g(x)$ is entirely determined by g's restriction to $\overline{B^n}$ (i.e. f).

.
Meanwhile, suppose $x_j = 0$ and define $\alpha(t) = \cos(t)x + \sin(t)e_j$. Then α is smooth an so $\nabla g(\alpha(t)) \cdot \alpha'(t)$ will be a continuous function converging to $\frac{\partial}{\partial x_j} g(x)$ as $t \to 0$. But also note that for any $i \neq j$ we have that $\alpha'(t) \cdot e_i = -\sin(t)x_i$. So for any $t \in \mathbb{R}$, we have that:

$$\begin{aligned}\nabla g(\alpha(t)) \cdot \alpha'(t) &= \cos(t)\tfrac{\partial}{\partial x_j} g(\alpha(t)) - \sum_{\substack{1 \leq i \leq n \\ i \neq j}} \sin(t) x_i \tfrac{\partial}{\partial x_i} g(\alpha(t)) \\ &= \cos(t)\tfrac{\partial}{\partial x_j} g(\alpha(t)) - \sum_{\substack{1 \leq i \leq n \\ x_i \neq 0}} \sin(t) x_i \tfrac{\partial}{\partial x_i} g(\alpha(t))\end{aligned}$$

Next note that if $x_i \neq 0$ and $i \neq j$, then $\alpha(t) \cdot e_i = cos(t)x_i \neq 0$ for all $t \in (-\pi/2, \pi/2)$. Thus, by our prior work we know for all $t \in (-\pi/2, \pi/2)$ that $\frac{\partial}{\partial x_i} g(\alpha(t))$ is determined by f provided that $x_i \neq 0$. At the same time, for all $t \in (-\pi/2, \pi/2) - \{0\}$, we have that $\alpha(t) \cdot e_j = \sin(t) \neq 0$. Thus by our prior work, we know that $\frac{\partial}{\partial x_i} g(\alpha(t))$ is determined by f for all $t \in (-\pi/2, \pi/2) - \{0\}$.

And now we are done as the limit of $\nabla g(\alpha(t)) \cdot \alpha'(t)$ as $t \to 0$ is uniquely determined by f. ■

As a result, if $f \in C^r(\overline{B^n})$ (where $r \geq 1$) then we can without ambiguity refer to f as having partial derivatives on S^{n-1}. Also, by an application of *proposition C.29* on page 291, we have that if $f \in C^1(\overline{B^n})$ then f is Lipschitz.

With that out of the way, I'm now ready to get into Howard's paper.

Lemma 1: There is no C^1 map $f : \overline{B^n} \to S^{n-1}$ such that $f(x) = x$ for all $x \in S^{n-1}$.

Proof:

Assume such an f exists and for $t \in [0,1]$ let $f_t(x) \coloneqq (1-t)x + tf(x) = x + tg(x)$ (where $g(x) \coloneqq f(x) - x$). Then note that for all $x \in \overline{B^n}$:

$$\|f_t(x)\|_2 \leq (1-t)\|x\|_2 + t\|f(x)\|_2 \leq (1-t) + t = 1.$$

Hence, f_t maps $\overline{B^n}$ into $\overline{B^n}$. Also, f_t fixes S^{n-1} as:

$$f_t(x) = (1-t)x + tf(x) = (1-t)x + tx = x.$$

Now g is C^1 since f is. So, let $C > 0$ be a Lipschitz constant for g (i.e. for all $x_1, x_2 \in \overline{B^n}$ we have that $\|g(x_2) - g(x_1)\|_2 \leq C\|x_2 - x_1\|_2$). Using this fact, we claim that when t is small enough then f_t is injective. After all, suppose there eixsts distinct $x_1, x_2 \in \overline{B^n}$ with $f_t(x_1) = f_t(x_2)$. Then $x_2 - x_1 = t(g(x_1) - g(x_2))$. So:

$$\|x_2 - x_1\|_2 = t\|g(x_2) - g(x_1)\|_2 \leq tC\|x_2 - x_1\|_2$$

And since $\|x_2 - x_1\|_2 > 0$, this means that $tC \geq 1$. So if $t < 1/C$, then we must have that f_t is injective.

Next, for each $t \in [0,1]$ let $G_t \coloneqq f_t(B^n)$. We claim that if t is small enough then G_t is open. To see why, let $L(\mathbb{R}^n)$ be the space of linear maps from $\mathbb{R}^n$ to itself equipped with the standard operator norm. Then $A(x,t) = (f_t)'(x) = \mathrm{Id} + tg'(x)$ is a well-defined map from $\overline{B^n} \times [0,1]$ to $L(\mathbb{R}^n)$ that is continuous with respect to t and with respect to x (I'm not sure and don't really care to prove if it is continuous with respect to x and t jointly).

By the extreme value theorem, we know there exists $M = \max(\|g'(x)\|_{\mathrm{op}} : x \in \overline{B^n})$. And then since the set of invertible maps in $L(\mathbb{R}^n)$ is open, we know there is some $r > 0$ such that $A(x,t)$ is invertible when $\|A(x,t) - \mathrm{Id}\|_{\mathrm{op}} = t\|g'(x)\|_{\mathrm{op}} \leq r$. Thus, when $t \leq \frac{r}{M}$, we know that $A(x,t) = (f_t)'(x)$ is invertible for all x. And now it follows from the inverse function theorem (see my math 140c notes) that f_t restricted to B^n is an open map when $t \leq \frac{r}{M}$.

By letting $t_0 = \min(1/C, r/M) > 0$, we have thus shown that f_t is injective and that G_t is open for all $t \in [0, t_0]$. One more claim we'll make before we get to the main part of the proof is that $G_t = B^n$ when $t \in [0, t_0]$. To show this, note that since f_t is injective, we know that $G_t \subseteq \overline{B^n} - S^{n-1} = B^n$.

Now for the sake of contradiction suppose $G_t \neq B^n$. Then we know that $(\partial G_t) \cap B^n \neq \emptyset$. In particular, let $y^{(0)} \in (\partial G_t) \cap B^n$ and then consider a sequence $\{x_\ell\}_{\ell \in \mathbb{N}}$ in B^n such that $f_t(x_\ell) \to y_0$ as $\ell \to \infty$. By the compactness of $\overline{B^n}$, we can pass to a subsequence to say without loss of generality that $x_\ell \to x^{(0)}$ as $\ell \to \infty$ for some $x^{(0)} \in \overline{B^n}$.

By the continuity of f, we know that $f(x^{(0)}) = y^{(0)}$. But also note that since G_t is open, we know $y^(0) \notin G_t$. Hence $x^{(0)} \in \overline{B^n} - B_n = S^{n-1}$. And since f_t fixes S^{n-1}, we thus know that $y^{(0)} = x^{(0)} \in S^{n-1}$. This contradicts that we originally said $y^{(0)} \in B^n$. Hence, we conclude $G_t = B^n$.

Consequently, we know when $t \in [0, t_0]$ then f_t is a bijection from $\overline{B^n}$ to $\overline{B^n}$ and that f_t restricted to B^n is a C^1 diffeomorphism from B^n to itself.

Now define $F(t) = \int_{B^n} \det((f_t)'(x))\mathrm{d}x = \int_{B^n} \det(\mathrm{Id} + tg'(x))\mathrm{d}x$ for all $t \in [0, 1]$.

By the change of variables theorem, we know $F(t) = m(B^n)$ when $t \in [0, t_0]$ (where m is the Lebesgue measure). But at the same time, F is a polynomial in t. So, $F(t) = m(B^n) > 0$ for all $t \in [0, 1]$.

Yet at the same time, $f_1(x) = f(x) \in S^{n-1}$ for all $x \in \overline{B^n}$. Therefore, if $x \in B^n$ and $v \in \mathbb{R}^n$ is fixed and $s \in \mathbb{R}$ is small enough that $x + sv \in \overline{B^n}$, then $\|f_1(x + sv)\|_2 = 1$, Therefore for those small enough s we have that:

$$0 = \tfrac{\mathrm{d}}{\mathrm{d}s}(f_1(x + sv) \cdot f_1(x + sv)) = 2f_1'(x + sv)v \cdot f_1(x + sv)$$

In particular, by plugging in $s = 0$ we have that $f_1'(x)v \cdot f_1(x) = 0$ for all $v \in \mathbb{R}^n$ and $x \in B^n$. And this proves that f_1 is not full rank for all $x \in B^n$. Hence, we have a contradiction since:

$$F(1) = \int_{B^n} \det((f_t)'(x))\mathrm{d}x = \int_{B^n} 0\mathrm{d}x = 0 \neq m(B^n). \blacksquare$$

<u>Brouwer's Fixed Point Theorem:</u> Every continuous map $f : \overline{B^n} \to \overline{B^n}$ has a fixed point (i.e. there exists $x \in \overline{B^n}$ with $f(x) = x$).

Proof:

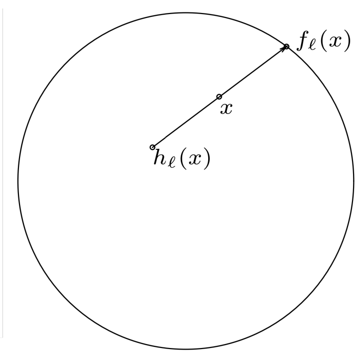
$C^\infty(\overline{B^n}, \mathbb{R})$ is a subalgebra of $C(\overline{B^n}, \mathbb{R})$ that separates points and vanishes nowhere. Thus, by applying the Stone-Weierstrass theorem to the component functions of f and then passing to a suitable subsequence, we can get a sequence $\{g_\ell\}_{\ell \in \mathbb{N}} \in C^\infty(\overline{B^n}, \mathbb{R}^n)$ such that $\|g_\ell(x) - f(x)\|_2 < 1/\ell$ for all $x \in \overline{B^n}$ and $\ell \in \mathbb{N}$.

Now $\|g_\ell(x)\|_2 \leq \|f(x)\|_2 + \|f(x) - g_\ell(x)\|_2 = 1 + 1/\ell$. Thus, if we set $h_\ell := (1 + 1/\ell)^{-1} g_\ell$, we get that h_ℓ smoothly maps $\overline{B^n}$ into $\overline{B^n}$ and that

$$\lim_{\ell \to \infty} \sup_{x \in \overline{B^n}} \|h_\ell(x) - f(x)\|_2 \to 0.$$

Now we claim each h_ℓ has a fixed point. After all, suppose not. Then for every $x \in B^n$ we know that $h_\ell(x) \neq x$. And so, it is well defined to let $f_\ell(x)$ equal the point where the ray starting at $h_\ell(x)$ and passing through x intercepts S^{n-1} (after going through x) for every $x \in \overline{B^n}$.

Howard includes the diagram to the right to describe f_ℓ and claims (without proof) that f_ℓ is a smooth function. To actually show f_ℓ is smooth, we need to find an actual formula for $f_\ell(x)$.



Note that $f_\ell(x) = h_\ell(x) + t(x - h_\ell(x))$ for some $t \geq 0$ and that $\|f_\ell(x)\|_2 = 1$. So, let's first try to find the $t \geq 0$ such that $\|h_\ell(x) + t(x - h_\ell(x))\|_2 = 1$.

Note $\|h_\ell(x) + t(x - h_\ell(x))\|_2^2 = \|h_\ell(x)\|_2^2 + 2t(h_\ell(x) \cdot (x - h_\ell(x))) + t^2\|x - h_\ell(x)\|_2^2$. Therefore, we need to solve the following quadratic equation:

$$(\|h_\ell(x)\|_2^2 - 1) + 2t(h_\ell(x) \cdot (x - h_\ell(x))) + t^2\|x - h_\ell(x)\|_2^2 = 0.$$

So, apply quadratic formula to get that:

$$t = \frac{-(h_\ell(x)\cdot(x-h_\ell(x))) \pm \sqrt{(h_\ell(x)\cdot(x-h_\ell(x)))^2 - (\|h_\ell(x)\|_2^2 - 1)\|x-h_\ell(x)\|_2^2}}{\|x-h_\ell(x)\|_2^2}$$

Now since $\|h_\ell(x)\|_2 \leq 1$, we have that:

$$(y\cdot(x-h_\ell(x)))^2 - (\|h_\ell(x)\|_2^2 - 1)\|x-h_\ell(x)\|_2^2 \geq (h_\ell(x)\cdot(x-h_\ell(x)))^2 \geq 0$$

Hence, any t solving our quadratic will be real valued. Also importantly, since the line passing through $h_\ell(x)$ and x intercepts $\overline{B^n}$ in more than one place (namely $h_\ell(x)$ and x), that line must intercept S^{n-1} in two places. Hence, two values of t must satisfy our quadratic. And this proves that the long expression under the radical never equals 0.

Finally, the root we want is specifically the greater of the two. So, we set:

$$T(x) := \frac{-(h_\ell(x)\cdot(x-h_\ell(x))) + \sqrt{(h_\ell(x)\cdot(x-h_\ell(x)))^2 - (\|h_\ell(x)\|_2^2 - 1)\|x-h_\ell(x)\|_2^2}}{\|x-h_\ell(x)\|_2^2}.$$

And importantly, $T(x)$ is a smooth real-valued function.

But now $f_\ell(x) = h_\ell(x) = T(x)(x - h_\ell(x))$ is smooth, maps all of $\overline{B^n}$ into S^{n-1}, and fixes S^{n-1}. This violates lemma 1. So, the premise that h_ℓ had no fixed point has to be false.

To finish off our proof, for each $\ell \in \mathbb{N}$ pick x_ℓ such that $h_\ell(x_\ell) = x_\ell$. Since $\overline{B^n}$ is compact, we know that there is a convergent subsequence $\{x_{\ell_k}\}_{k\in\mathbb{N}}$ converging to $x^{(0)} \in \overline{B^n}$. And now since $\sup_{x\in\overline{B^n}} \|h_\ell(x) - f(x)\|_2 \to 0$, we have that:

$$f(x^{(0)}) = \lim_{k\to\infty} h_{\ell_k}(x_{\ell_k}) = \lim_{k\to\infty} x_{\ell_k} = x^{(0)}. \blacksquare$$

While I'm at it, here is Howard's theorem 2:

Corollary: There is no continuous map $f : \overline{B^n} \to S^{n-1}$ with $f(x) = x$ for all $x \in S^{n-1}$.

Proof:

Assume such a map exists and let $g(x) = -f(x)$. Then g is also a map from $\overline{B^n}$ to $S^{n-1} \subseteq \overline{B^n}$. By the Brouwer fixed point theorem g must have a fixed point $y \in \overline{B^n}$. But if $y \in S^{n-1}$, then $g(y) = -f(y) = -y \neq y$. Meanwhile, $g(y) \notin B^n$ for all $y \in B^n$. So, g can't have a fixed point. This is a contradiction. $\blacksquare$

Math 200a (don't know what lecture)

Lemma: The standard S.E.S. (where $N \lhd G$):

$$\{1\} \longrightarrow N \xhookrightarrow{i} G \xrightarrow{\pi} G/N \longrightarrow \{1\}$$

splits iff there exists $H < G$ such that $HN = G$ and $H \cap N = \{1\}$.

$(\Longrightarrow)$

Assuming that it splits, we get a group homomorphism $f : G/N \to G$ such that $\forall x \in G,\ \ f(xN)N = xN$. If $xN \in \ker(f)$ then $xN = f(xN)N = N$. So, f is injective. And hence, by letting $H := \mathrm{im}(f)$ we get that $H < G$ and that $H \cong G/N$ (by the first isomorphism theorem).

- If $x \in H \cap N$, then we claim that $x = 1_G$.

Why?
If $x \in H$ then $x = f(x'N)$ for some $x' \in G$. And hence:
$$xN = f(x'N)N = x'N.$$

But if also $x \in N$, then $xN = N$. So $x'N = N$ and in turn:
$$x = f(x'N) = 1_G.$$

- For all $x \in G$ there exists $h \in H$ and $n \in N$ with $x = hn$.

  Why?
  We know for all $xN \in G/N$ that there exists $h = f(xN) \in H$ such that $hN = xN$. In turn, for any $x \in G$ we can find a coset hN containing x where $h \in H$. So, $G = HN$.

$(\Longleftarrow)$
By the second isomorphism theorem, we know that $(HN)/N \cong H/(H \cap N)$. Also, by our assumptions, we know that $G/N \cong (HN)/N$ and that $H/(H \cap N) \cong H$. So, let $\theta(hN) = h$ for all $h \in H$. (This is just the isomorphism going from G/N to H that would have been defined by our isomorphism theorem).

If you really don't trust that this is an isomorphism, note that since $HN = G$ we know for all $xN \in G/N$ there exists $h \in H$ with $xN = hN$. Also, if $h_1N = h_2N$, then $h_2^{-1}h_1 = n^{-1}$ for some $n \in N$. And since $H \cap N = \{1\}$, this means that $h_2^{-1}h_1 = 1$. So, $h_2 = h_1$. Hence, θ is a well-defined map from G/N to H.

Now the rest of the proof that θ is a group isomorphism is trivial...

Next, define $f = i \circ \theta$ so that the diagram below commutes.

$$\begin{tikzcd} G/N \arrow[r, "\theta"] \arrow[rr, bend right, "f"'] & H \arrow[r, hook, "i"] & G \end{tikzcd}$$

Clearly g is a group homomorphism and we claim that f splits our S.E.S.
Why? For any $xN \in G/N$ we have that $\theta(xN) \in H$. Thus $\theta(\theta(xN)N) = \theta(xN)$. And since θ is injective, we must have that $\theta(xN)N = xN$. ■

So now that we have a condition for when our standard S.E.S. splits in terms of the existence of a group $H < G$ with certain properties. Is there a way to give the set $N \times H$ a group structure such that the obvious bijection $\phi(n, h) = nh$ is a group isomorphism and also the below diagram commutes?

$$\begin{tikzcd} \{1\} \arrow[r] & N \arrow[r, hook, "i"] & G \arrow[r, shift left, "\pi"] & G/N \arrow[l, shift left, "f"] \arrow[r] & \{1\} \\ \{1\} \arrow[r] & N \arrow[r] \arrow[u, "\mathrm{Id}"] & N \times H \arrow[r] \arrow[u, "\phi"'] & H \arrow[r] \arrow[u, "\pi \circ i"'] & \{1\} \end{tikzcd}$$

Since our diagram needs to commute, we must have that have that:
$$\phi(n_1, h_1)\phi(n_2, h_2) = (n_1h_1)(n_2h_2) = nh_1n_2h_1^{-1}h_1h_2 = \phi(n_1h_1n_2h_1^{-1}, h_1h_2)$$

And this now motivates us to define the semidirect product of N and H.

Let H and N be groups (right now we won't assume they are subgroups of another group) and let $f : H \to \mathrm{Aut}(N)$ be a group homomorphism. Then define the operation:

$$(n_1, h_1) \cdot (n_2, h_2) = (n_1(f(h_1))(n_2), h_1 h_2)$$

This turns $(N \times H, \cdot)$ into a group which we denote as $N \rtimes_f H$ to distinguish it from the direct product of N and H (which has group product $(n_1, h_1) \cdot (n_2, h_2) = (n_1 n_2, h_1 h_2)$).

Proving that $N \rtimes_f H$ is a group would be boring and it's already in math 100a notes from last Fall.

Alireza (the professor teaching this class) likes to use the shorthand $\underline{n} := (n, 1)$ and $\underline{h} := (1, h)$. Note that $\theta_1 : N \to N \rtimes_f H$ given by $n \mapsto \underline{n}$ and $\theta : H \to N \rtimes_f H$ given by $h \mapsto \underline{h}$ are injective group homomorphisms. Furthermore, the map $\theta_2 : N \rtimes_f H \to H$ given by $(n, h) \mapsto h$ is a surjective group homomorphism.

I'm again skipping proving these since it's just a lot of symbol moving.

Corollary: The following is a splitting S.E.S. (which we call a standard splitting S.E.S.):

$$\begin{tikzcd} \{1\} \arrow[r] & N \arrow[r, "\theta_1"] & N \rtimes_f H \arrow[r, shift left, "\theta_2"] & H \arrow[l, shift left, "\theta"] \arrow[r] & \{1\} \end{tikzcd}$$

Theorem: Suppose $\begin{tikzcd} \{1\} \arrow[r] & G_1 \arrow[r, "\theta_1"] & G_2 \arrow[r, shift left, "\theta_2"] & G_3 \arrow[l, shift left, "\theta"] \arrow[r] & \{1\} \end{tikzcd}$ is a splitting S.E.S.

Then by letting $f : G_3 \to \mathrm{Aut}(G_1)$ be given by $(f(x_3))(x_1) := \theta_1^{-1}(\theta(x_3)\theta_1(x_1)\theta(x_3)^{-1})$ and $\phi : G_1 \rtimes_f G_3 \to G_2$ be given by $\phi(x_1, x_3) := \theta_1(x_1)\theta(x_3)$ we have that the following is an isomorphism of S.E.S.s.:

$$\begin{tikzcd} \{1\} \arrow[r] & G_1 \arrow[r, "\theta_1"] & G_2 \arrow[r, "\theta_2"] & G_3 \arrow[r] & \{1\} \\ \{1\} \arrow[r] & G_1 \arrow[u, "\mathrm{Id}"] \arrow[r, "g_1 \mapsto \underline{g_1}"'] & G_1 \rtimes_f G_3 \arrow[u, "\phi"'] \arrow[r, "(g_1, g_3) \mapsto g_3"'] & G_3 \arrow[u, "\mathrm{Id}"'] \arrow[r] & \{1\} \end{tikzcd}$$

Proof:
First we show $f(x_3) \in \mathrm{Aut}(G_1)$ for all $x_3 \in G_3$:

- Firstly, because $\mathrm{im}(\theta_1) = \ker(\theta_2)$ is a normal subgroup, we know that $\theta(x_3)\theta_1(x_1)\theta(x_3)^{-1}$ is also in $\mathrm{im}(\theta_1)$. Also, since θ_1 is injective, we know that there is a unique element in $\theta^{-1}(y)$ for any $y \in \mathrm{im}(\theta_1)$. So, $(f(x_3))$ is a well-defined function.

- Let $x_1, y_1 \in G_1$ and $x_3 \in G_3$. Then:
$$\begin{aligned} (f(x_3))(x_1) \cdot (f(x_3))(y_1) &= \theta_1^{-1}(\theta(x_3)\theta_1(x_1)\theta(x_3)^{-1})\theta_1^{-1}(\theta(x_3)\theta_1(y_1)\theta(x_3)^{-1}) \\ &= \theta_1^{-1}((\theta(x_3)\theta_1(x_1)\theta(x_3)^{-1})(\theta(x_3)\theta_1(y_1)\theta(x_3)^{-1})) \\ &= \theta_1^{-1}(\theta(x_3)\theta_1(x_1)\theta_1(y_1)\theta(x_3)^{-1}) \\ &= (f(x_3))(x_1 y_1) \end{aligned}$$

This shows that $(f(x_3))$ is a group homomorphism.

- Since θ_1 is injective, we know that if $\theta^{-1}(y) = 1$ (where $y \in \mathrm{im}(\theta_1)$), then $y = 1$. Hence if $(f(x_3))(x_1) = 1$ then $\theta(x_3)\theta_1(x_1)\theta(x_3)^{-1} = 1$. But then $\theta_1(x_1) = \theta(x_3)^{-1}\theta(x_3) = 1$. So by using the injectivity of θ_1 again have that $x_1 = 1$. And this proves that $(f(x_3))$ is injective.

- Finally, if $x_1 \in G_1$, then let $y_1 = \theta^{-1}(\theta(x_3)^{-1}\theta(x_1)\theta(x_3))$. Then it is easy to see that $\theta(y_1) = x_1$. So, $(f(x_3))$ is surjective.

Next we show that f is a group homomorphism:

Fix $x_1 \in G_1$ and let $x_3, y_3 \in G_3$. Then:

$$\begin{aligned}(f(y_3))((f(x_3))(x_1)) &= \theta_1^{-1}(\theta(y_3)\theta_1((f(y_3))(x_1))\theta(y_3)^{-1}) \\ &= \theta_1^{-1}(\theta(y_3)\theta_1(\theta_1^{-1}(\theta(x_3)\theta_1(x_1)\theta(x_3)^{-1}))\theta(y_3)^{-1}) \\ &= \theta_1^{-1}(\theta(y_3)\theta(x_3)\theta_1(x_1)\theta(x_3)^{-1}\theta(y_3)^{-1}) \\ &= \theta_1^{-1}(\theta(y_3x_3)\theta_1(x_1)\theta(y_3x_3)^{-1}) \\ &= (f(y_3x_3))(x_1)\end{aligned}$$

And this proves that $f(y_3) \circ f(x_3) = f(y_3x_3)$. So f is a group homomorphism.

With that, we now have a well-defined semidirect product $G_1 \rtimes_f G_3$. So, we next show that the function ϕ we defined before is a group homomorphism:

Note that for any $x_1, y_1 \in G_1$ and $x_3, y_3 \in G_3$:

$$\begin{aligned}\phi((x_1, x_3)\cdot(y_1, y_3)) &= \phi((x_1(f(x_3)(y_1), x_3y_3))) \\ &= \theta_1(x_1(f(x_3))(y_1))\theta(x_3y_3) \\ &= \theta_1(x_1)\theta_1((f(x_3))(y_1))\theta(x_3)\theta(y_3) \\ &= \theta_1(x_1)\theta(x_3)\theta_1(y_1)\theta(x_3)^{-1}\theta(x_3)\theta(y_3) \\ &= \theta_1(x_1)\theta(x_3)\theta_1(y_1)\theta(y_3) = \phi((x_1, x_3))\phi((y_1, y_3)).\end{aligned}$$

Finally, if our proposed diagram actually commutes, then we can apply the short five lemma to get that ϕ is a group isomorphism. And then we will be done.

- $\phi(\underline{x_1}) = \phi(x_1, 1) = \theta_1(x_1)\theta(1) = \theta_1(x_1)$. Thus the first square commutes.

- Since $\mathrm{im}(\theta_1) = \ker(\theta_2)$, we know that $\theta_2(\theta_1(x_1)) = 1$ for all $x_1 \in G_1$. Also, recall that $\theta_2(\theta(x_3)) = x_3$. Therefore:

$$\theta_2(\phi((x_1, x_3))) = \theta_2(\theta_1(x_1)\theta(x_3)) = \theta_2(\theta_1(x_1))\theta_2(\theta(x_3)) = x_3$$

And this proves the second square commutes. ■

I somehow just realized something really helpful that I want to write down:

Proposition: Suppose we have the following diagram:

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, shift left, "\theta^{-1}"] & B \arrow[d, shift left, "\phi^{-1}"] \\ C \arrow[r, "g"'] \arrow[u, shift left, "\theta"] & D \arrow[u, shift left, "\phi"] \end{tikzcd}$$

If $\theta f = g\phi$, then $f\phi^{-1} = \theta^{-1}g$. (i.e. if the vertical arrows are isomorphisms, then we only need to check if the diagram commutes going around one way).

Proof:
$f\phi^{-1} = (\theta^{-1}\theta)f\phi^{-1} = \theta^{-1}(\theta f)\phi^{-1} = \theta^{-1}(g\phi)\phi^{-1} = \theta^{-1}g(\phi\phi^{-1}) = \theta^{-1}g.$ ■

10/20/2025

Math 220a (I don't know which lecture):

If $G \subseteq \mathbb{C}$ is a region and $f : G \to \mathbb{C}$ is a continuous functions such that $z = \exp(f(z))$ for all $z \in G$, then f is a branch of the logarithm.

For the most important (or "principle") branch of the logarithm, let:
$$G = \mathbb{C} - \{z : \mathrm{Im}(z) = 0 \text{ and } \mathrm{Re}(z) \leq 0\}.$$

Clearly G is connected and open, and each $z \in G$ can be uniquely represented as $z = |z|e^{i\theta}$ for some $\theta \in (-\pi, +\pi)$. So, for all $re^{i\theta} \in G$ (where $0 < r$ and $\theta \in (-\pi, \pi)$) we define:
$$\mathrm{Log}(re^{i\theta}) = \log(r) + i\theta$$

(...where $\log$ refers to the normal real logarithm while Log refers to the principle branch of the complex logarithm).

Exercise III.2.9: Suppose $\{z_n\}_{n\in\mathbb{N}}$ is a sequence in G converging to z and $z_n = r_n e^{i\theta_n}$ and $z = re^{i\theta}$ where $r, r_n > 0$ and $\theta, \theta_n \in (-\pi, \pi)$. Then $r_n \to r$ and $\theta_n \to \theta$.

Proof:
$|\cdot|$ is continuous (as is seen by reverse triangle inequality). Therefore, since $z_n \to z$, we know that $r_n = |z_n| \to |z| = r$ as $n \to \infty$. As a consequence, since $r_n, r \neq 0$ for any r, we can say that $e^{i\theta_n} = \frac{z_n}{r_n} \to \frac{z}{r} = e^{i\theta}$ as $n \to \infty$.

Finally note that $e^{i\theta}$ is in the interior of either $\{z \in \mathbb{C} : \mathrm{Re}(z) > 0\}$, $\{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$, or $\{z \in \mathbb{C} : \mathrm{Im}(z) < 0\}$. Thus we must have that $e^{i\theta_n}$ is eventually in one of those open sets, which in turn means θ_n is eventually in $[-\pi/2, \pi/2]$, $[0, \pi]$, or $[-\pi, 0]$. In either of those cases, call the interval θ_n is eventually in K.

Now $f : K \to \mathrm{im}(f)$ given by f^{it} is a bijective function from a compact space to a metric space. Hence $f^{-1} : \mathrm{im}(f) \to K$ is also continuous. So, the fact that $f(\theta_n) \to f(\theta)$ means that $\theta_n = f^{-1}(f(\theta_n)) \to f^{-1}(f(\theta)) = \theta$. ■

Corollary: $\mathrm{Log} : G \to \mathbb{C}$ is continuous.

(Conway) Proposition 2.20: Let G and Ω be open subsets of $\mathbb{C}$ and suppose that $f : G \to \mathbb{C}$ and $g : \Omega \to \mathbb{C}$ are continuous functions such that $f(G) \subseteq \Omega$ and $g(f(z)) = z$ for all $z \in G$. If g is differentiable and $g' \neq 0$, then f is differentiable and $f'(z) = \frac{1}{g'(f(z))}$.

Proof:

Fix $a \in G$ and let $h \in \mathbb{C}$ such that $h \neq 0$ and $a + h \in G$. Then $g(f(a)) = a$ and $a + h = g(f(a+h))$ implies that $f(a) \neq f(a+h)$. So:

$$1 = \frac{a+h-a}{h} = \frac{g(f(a+h))-g(f(a))}{h} = \frac{g(f(a+h))-g(f(a))}{f(a+h)-f(a)} \cdot \frac{f(a+h)-f(a)}{h}$$

And now we have that $\lim_{h\to 0} \frac{f(a+h)-f(a)}{h} = \left(\lim_{h\to\infty} \frac{g(f(a+h))-g(f(a))}{f(a+h)-f(a)}\right)^{-1}$.

But since $|f(a+h)-f(a)| \to 0$ as $h \to 0$ (since f is continuous) and g is differentiable, we know that:

$$\lim_{h\to\infty} \frac{g(f(a+h))-g(f(a))}{f(a+h)-f(a)} = g'(f(a))$$

And this proves that $f'(a)$ exists and equals $\frac{1}{g'(f(a))}$.

Side note: In this case we also have that f' is as smooth as g' is.

(Conway) Corollary 2.21: A branch of the logarithm is holomorphic and it's derivative is z^{-1}.

Proof:

If $G \subseteq \mathbb{C}$ is a region and $f : G \to \mathbb{C}$ is a branch of the logarithm, then $f'(z) = (\exp'(f(z)))^{-1} = (\exp(f(z)))^{-1} = z^{-1}$. ■

A while ago I wrote down *a theorem* concerning the Cauchy-Rieman equations. But I never actually proved one of the proof directions. So I'd like to do that now.

Statement: If G is a region of $\mathbb{C} \cong \mathbb{R}^2$, and:

$$f(x+iy) = f((x,y)) = (u(x,y), v(x,y)) \in C^1(G)$$

with $u_x = v_y$ and $v_x = -u_y$, then f is holomorphic.

Proof:

Denote r to be the common value of u_x and v_y, and similarly define s to be the common value of v_x and $-u_y$. Then:

$$\begin{pmatrix} r & -s \\ s & r \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} \cong rh - sk + i(sh + rk) = (r+is)(h+ik)$$

Now since $f \in C^1(G)$, we know that:

$$0 = \lim_{h+ik\to 0} \frac{|f(x+iy+h+ik)-f(x+iy)-(r+is)(h+ik)|}{|h+ik|} = \lim_{h+ik\to 0} \left|\frac{f(x+iy+h+ik)-f(x+iy)-(r+is)(h+ik)}{h+ik}\right|$$
$$= \lim_{h+ik\to 0} \left|\frac{f(x+iy+h+ik)-f(x+iy)}{h+ik} - (r+is)\right|$$

And thus $f'(x+iy)$ is well defined with:

$$f'(x+iy) = \lim_{h+ik\to 0} \frac{f(x+iy+h+ik)-f(x+iy)}{h+ik} = r + is. \blacksquare$$

Suppose $f(x,y) = (u(x,y), v(x,y))$ has continuous second derivatives and satisfies the Cauchy-Riemann equations. Then by differentiating a second time we get that:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x \partial y} \text{ and } \frac{\partial^2 v}{\partial y \partial x} = -\frac{\partial^2 u}{\partial y^2}$$

Hence, we get another partial differential equation:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

Any function u with continuous second derivatives satisfying the above differential equation is said to be harmonic. Recalling that I proved last spring that complex differentiable functions on a region are infinitely differentiable, we thus know that if G is a region of $\mathbb{C}$ and $f : G \to \mathbb{C}$ is holomorphic, then $\mathrm{Re}(f)$ is harmonic.

On a related note, suppose $u : \mathbb{R}^2 \to \mathbb{R}$ is harmonic. Then can we find another function $v : \mathbb{R}^2 \to \mathbb{R}$ such that $f(x+iy) = u(x,y) + iv(x,y)$ is holomorphic? The answer is sometimes.

As a side note, that hypothetical other function v is called a harmonic conjugate of u.

(Conway) Theorem 2.30: Let G be either all of $\mathbb{C}$ or an open disk. If $u : G \to \mathbb{R}$ is a harmonic function (where we are identifying $\mathbb{C} \cong \mathbb{R}^2$), then u has a harmonic conjugate.

Proof:

Without loss of generality assume the disk is centered at the origin. Also denote the radius of the disk as R (where we allow $R = +\infty$). Now consider any function of the form:

$$v(x,y) = \int_0^y u_x(x,t)\mathrm{d}t + \psi(x)$$

We focus on functions of this form because it is immediately clear from the fundamental theorem of calculus that $v_y(x,y) = u_x(x,y) + 0$. Also, modifying ψ won't change what the yth partial derivative of v is. So, we now seek to find some function ψ such that $v_x = -u_y$.

Note that since $u_x : \mathbb{R}^2 \to \mathbb{R}$ is continuous (since it is differentiable), we know by the extreme value theorem that for any $0 < r < R$ there is a constant $C_r > 0$ such that $u_x(x,y) \leq C_r$ when $\|x,y\|_2 \leq r$. Thus, we can apply the theorem on *page 189* to show that for any $y < r$ and $x < \sqrt{r^2 - y^2}$ that:

$$\begin{aligned} v_x(x,y) &= \frac{\partial}{\partial x}\int_0^y u_x(x,t)\mathrm{d}t + \frac{\mathrm{d}}{\mathrm{d}x}\psi(x) \\ &= \int_0^y \frac{\partial}{\partial x}u_x(x,t)\mathrm{d}t + \frac{\mathrm{d}}{\mathrm{d}x}\psi(x) = -\int_0^y \frac{\mathrm{d}^2}{\mathrm{d}y^2}u(x,t)\mathrm{d}t + \frac{\mathrm{d}}{\mathrm{d}x}\psi(x) \\ &= -u_y(x,y) + u_y(x,0) + \frac{\mathrm{d}}{\mathrm{d}x}\psi(x) \end{aligned}$$

And by taking $r \to R$ we can say that $v_x(x,y) = -u_y(x,y) + u_y(x,0) + \frac{\mathrm{d}}{\mathrm{d}x}\psi(x)$ for all $(x,y) \in G$. Therefore, we have that $v_x(x,y) = -u_y(x,y)$ iff $\frac{\mathrm{d}}{\mathrm{d}x}\psi(x) = -u_y(x,0)$.

This clues us to fix ψ so that:

$$v(x,y) := \int_0^y u_x(x,t)\mathrm{d}t - \int_0^x u_y(s,0)\mathrm{d}s$$

And then u and v satisfy the Cauchy-Riemann equations. ■

A quick note that I thought of while talking about this with Hagan, we could replace our open disk with the half disk $\{z : |z| \leq R, \mathrm{Im}(z) \geq 0\}$.

10/22/2025

Today I want to first quickly work through Folland chapter 11 section 1 on topological groups. I specifically want to do this because it is relevant to math 241.

> One quick bit of notation: If $A \subseteq G$ (where A is not necessarily a subgroup), then I'll denote $A^{-1} = \{a^{-1} : a \in A\}$.

Proposition 11.1: Let G be a topological group.

(a) The topology of G is translation invariant: If U is open and $x \in G$, then Ux and xU are open.

> This is just a result of translation being a continuous homeomorphism on G.

(b) For every neighborhood U of 1 there is a symmetric neighborhood V of 1 with $V \subseteq U$. (recall from *page 238* what a symmetric subset of a group is).

> Proof:
> Since inversion is continuous, we know that U^{-1} also a neighborhood of 1. So, $U^{-1} \cap U$ is a neighborhood of 1.

(c) For every neighborhood U of 1 there is a neighborhood V of 1 with $VV \subseteq U$. (note that neighborhoods aren't necessarily subgroups but we still define $AB = \{ab : a \in A, b \in B\}$).

> Proof:
> Since $1 \cdot 1 = 1 \in U$ and the group product is continuous, we know there are open sets $V_1, V_2 \subseteq G$ containing 1 such that for all $a, b \in V_1 \times V_2$ we have that $ab \in U$. In particular, by taking $V = V_1 \cap V_2$ we have that $ab \in U$ for all $a, b \in V$ and that V is an open set containing 1.

(d) If H is a subgroup of G, so is $\overline{H}$.

> Proof:
> If $x, y \in \overline{H}$ then there are nets $\langle x_\alpha \rangle_{\alpha \in A}$ and $\langle y_\beta \rangle_{\beta \in B}$ contained in H which converge to x and y respectively. But now $x_\alpha y_\beta \in H$ for all $\alpha \in A$ and $\beta \in B$. Hence, we may consider net $\langle x_\alpha y_\beta \rangle_{(\alpha,\beta) \in A \times B}$ where $A \times B$ is given the obvious partial ordering.
>
> To see that this net converges to xy, consider any open neighborhood U about xy and note that since the group product is continuous, there are open sets N_1, N_2 around x and y respectively such that $x_\alpha y_\beta \in U$ whenever $x_\alpha \in N_1$ and $y_\beta \in N_2$. And since $x_\alpha \to x$ and $y_\beta \to y$, we can now easily find $(\alpha_0, \beta_0) \in A \times B$ such that $x_\alpha y_\beta \in U$ whenever $(\alpha, \beta) \gtrsim (\alpha_0, \beta_0)$.
>
> And this proves that $xy \in \overline{H}$.
>
> As for showing that $\overline{H}$ is closed under inversion, that is as simple as noting that inversion is a homeomorphism and thus preserves set closures (see the first lemma on *page 318*).

(e) Every open subgroup of G is also closed.

Proof:
If H is an open subgroup, then the cosets of H are also open and so $H^{\mathsf{C}} = \bigcup_{x \notin H} xH$ is open. This shows that H is clopen.

(f) If K_1, K_2 are compact subsets of G, so is K_1K_2.

Proof:
K_1K_2 is the image the compact set $K_1 \times K_2$ under the group composition map (which is continuous). ■

If f is a continuous function on a topological group G and $y \in G$, we denote $L_y f(x) = f(y^{-1}x)$ and $R_y f(x) = f(xy)$. This describes two group actions $G \curvearrowright \mathrm{Fun}(G, X)$ (where X is some other set).

f is called left (or right) uniformly continuous if for every $\varepsilon > 0$ there is a neighborhood V of 1_G such that $\|L_y f - f\|_u < \varepsilon$ (or $\|R_y f - f\|_u < \varepsilon$) for all $y \in V$.

Proposition 11.2: If $f \in C_c(G)$, then f is left and right uniformly continuous.

Proof:
Folland only proves right uniform continuity. Thus, since the two proofs are analogous, I will only prove left uniform continuity.

Let $K = \mathrm{supp}(f)$ and suppose $\varepsilon > 0$. For each $x \in K$ there is a neighborhood U_x of 1_G such that $|f(yx) - f(x)| < \varepsilon/2$ for all $y \in U_x$.

Why? We know that $|f(x') - f(x)| < \varepsilon/2$ when x' is in some neighborhood N of x. Then in turn, since right translation is continuous, we know that $U_x = \{y \in G : yx \in N\}$ is a neighborhood of 1_G.

By proposition 11.1, we can in turn find a symmetric neighborhood $V_x \subseteq U_x$ of 1_G such that $V_xV_x \subseteq U_x$. Hence, $V_x x$ is a neighborhoood of $x \in G$ with the property that if $g \in V_xV_x x$ then $|f(g) - f(x)| < \varepsilon/2$. (Note this also implies that if $g \in V_x x$ then $|f(g) - f(x)| < \varepsilon/2$).

Now since the $V_x x$ cover K as we range over all $x \in K$ and K is compact, we know that there are finitely many $x_1, \ldots, x_n$ such that $K \subseteq \bigcup_{j=1}^n V_{x_j} x_j$. So let $V = \bigcap_{j=1}^n V_{x_j}$ and note that V is a symmetric neighborhood of 1_G. We claim that $\|L_y f = f\|_u < \varepsilon$ for every $y \in V$.

- First suppose $x \in K$. Then $x \in V_{x_j} x_j$ for some j and consequently we have that $|f(x) - f(x_j)| < \varepsilon/2$. Also note that $xx_j^{-1} \in V_{x_j}$. So if $y \in V$ then $y^{-1}x = y^{-1}(xx_j^{-1})x_j \in V_{x_j}V_{x_j}x_j$ and we know that $|f(y^{-1}x) - f(x_j)| < \varepsilon/2$. Finally, by triangle inequality we have that:
$$|f(y^{-1}x) - f(x)| \leq |f(y^{-1}x) - f(x_j)| + |f(x_j) - f(x)| < \varepsilon.$$
- Meanwhile suppose $x \notin K$. Then $f(x) = 0$. And if $y \in V$ then either $y^{-1}x \in K$ or $y^{-1}x \notin K$. For the latter case, we then trivially have that $f(y^{-1}x) = 0$ and so $|f(y^{-1}x) - f(x)| = 0 < \varepsilon$. As for if $y^{-1}x \in K$, then we know that $y^{-1}x \in V_{x_j}x_j$ for some j. And this means that $|f(y^{-1}x) - f(x_j)| < \varepsilon/2$.

Also, $xx_j^{-1} = y(y^{-1}xx_j^{-1}) \in V_{x_j}V_{x_j} \subseteq U_{x_j}$. Hence:

$$|f(x_j)| = |f(x) - f(x_j)| = |f(xx_j^{-1}x_j) - f(x_j)| < \varepsilon/2.$$

So $|f(y^{-1}x) - f(x)| = |f(y^{-1}x)| \leq |f(y^{-1}x) - f(x_j)| + |f(x_j)| < \varepsilon$. ■

I'll continue with this on the weekend (see *page 349*).

More Math 220a Notes:

In this class we say a path γ in a region $G \subseteq \mathbb{C}$ is a continuous function from a closed interval $[a, b] \subseteq \mathbb{R}$ to G. We say a path γ is piecewise C^1 iff there exists $a = t_0 < t_1 < \ldots < t_n = b$ such that γ is C^1 on each subinterval $[t_{j-1}, t_j]$.

If γ_1, γ_2 are C^1 paths, $z_0 = \gamma_1(t_1) = \gamma_2(t_2)$, and $\gamma_1'(t_1) \neq 0$ and $\gamma_2'(t_2) \neq 0$, then we define the angle between the paths γ_1 and γ_2 at z_0 to be $\arg(\gamma_2'(t_2)) - \arg(\gamma_1'(t_1)) \pmod{2\pi}$.

Note that if $f : G \to \mathbb{C}$ is holomorphic and γ_1, γ_2 are C^1 paths in G, then $\sigma_1 = f \circ \gamma_1$ and $\sigma_2 = f \circ \gamma_2$ are also C^1 paths in $\mathbb{C}$. Now $\sigma_i'(t) = f'(\gamma_i(t))\gamma_i'(t)$ which means that $\arg(\sigma_i'(t)) = \arg(f'(\gamma_i(t))) + \arg(\gamma_i'(t))$ (if $f'(\gamma_i(t)) \neq 0$ and $\gamma_i'(t) \neq 0$). And consequently, if γ_1 intercepts γ_2 at some point (say $\gamma_1(t_1) = \gamma_2(t_2) = z_0$) and none of our derivatives are zero, then we have that $\sigma_1(t_1) = \sigma_2(t_2) = f(z_0)$ and:

$$\begin{aligned}\arg(\sigma_2'(t_2)) - \arg(\sigma_1'(t_1)) &= \arg(f'(z_0)) + \arg(\gamma_2'(t_2)) - \arg(f'(z_0)) - \arg(\gamma_1'(t_1)) \\ &= \arg(\gamma_2'(t_2)) - \arg(\gamma_1'(t_1))\end{aligned}$$

In conclusion, we have the following theorem:

> (Conway) Theorem III.3.4: If $f : G \to \mathbb{C}$ is holomorphic then f preserves angles at each point $z_0 \in G$ where $f'(z_0) \neq 0$.

A function $f : G \to \mathbb{C}$ which has the above described angle preserving property at $a \in G$ and which also satisfies that $\lim_{z \to a} \left|\frac{f(z)-f(a)}{z-a}\right|$ exists is called a conformal at a. If f is a bijection and f is conformal at every point of G, then f is called a conformal map.

A function $S(z) = \frac{az+b}{cz+d}$ (where $ad - bc \neq 0$) is called a Möbius transformation.

- Note that if $S(z)$ is a Möbius transformation, then $S^{-1}(z) = \frac{dz-b}{-cz+a}$ satisfies that $(da - (-b)(-c)) \neq 0$ and:

$$S(S^{-1}(z)) = \frac{a\frac{dz-b}{-cz+a}+b}{c\frac{dz-b}{-cz+a}+d} = \frac{a(dz-b)+b(a-cz)}{c(dz-b)+d(a-cz)} = \frac{adz-bcz}{-bc+ad} = z \text{ (if } z \neq a/c\text{)}$$

$$S^{-1}(S(z)) = \frac{d(\frac{az+b}{cz+d})-b}{-c(\frac{az+b}{cz+d})+a} = \frac{d(az+b)-b(cz+d)}{-c(az+b)+a(cz+d)} = \frac{adz-bcz}{-bc+ad} = z \text{ (if } z \neq -d/c\text{)}.$$

- Suppose $T(z) = \frac{a_1z+b_1}{c_1z+d_1}$ and $S(z) = \frac{a_2z+b_2}{c_2z+d_2}$, then:

$$(T \circ S)(z) = \frac{a_1 \frac{a_2 z + b_2}{c_2 z + d_2} + b_1}{c_1 \frac{a_2 z + b_2}{c_2 z + d_2} + d_1} = \frac{a_1 a_2 z + a_1 b_2 + b_1 c_2 z + b_1 d_2}{c_1 a_2 z + c_1 b_2 + d_1 c_2 z + d_1 d_2} = \frac{(a_1 a_2 + b_1 c_2) z + (a_1 b_2 + b_1 d_2)}{(c_1 a_2 + d_1 c_2) z + (c_1 b_2 + d_1 d_2)}$$

Clearly there is a connection to $2x2$ matrices here. I'll go more in depth into that in a second.

- If $S(z) := \frac{az+b}{cz+d}$, then S is defined on $\mathbb{C} - \{-d/c\}$. But note that $\lim_{|z| \to \infty} \frac{az+b}{cz+d} = \frac{a}{c}$. Also, since $ad - bc \neq 0$, we know that $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} z \\ 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} az+b \\ cz+d \end{smallmatrix}\right)$ can't have both components be zero at the same time. So:

$$\lim_{z \to -d/c} S(z) = \infty \in \mathbb{C}_\infty.$$

  Thus we can extend S to all of $\mathbb{C}_\infty$. Also, if $S^{-1}(z) = \frac{dz-b}{-cz+a}$, then by applying the same reasoning to S^{-1} we'd get that S^{-1} is the inverse function of S. Hence, Möbius transformations define a group of homeomorphisms on $\mathbb{C}_\infty$.

  I should also be careful to note that it suffices for all $z \in \mathbb{C}$ and $w \in \mathbb{C} - \{0\}$ to treat: $z + \infty = \infty$, $w \cdot \infty = \infty$, $\frac{z}{\infty} = 0$, $\frac{w}{0} = \infty$, and $\infty \cdot \infty = \infty$. If we use these conventions, then we don't need to have a separate edge case formula for when $c = 0$.

- If $S(z) := \frac{az+b}{cz+d}$ is a Möbius transformation, then S is holomorphic on $\mathbb{C} - \{\frac{-d}{c}\}$ with $S'(z) = \frac{a(cz+d) - c(az+b)}{(cz+d)^2} = \frac{ad-bc}{(cz+d)^2}$. Since $ad - bc \neq 0$, we know that S is a conformal map from $\mathbb{C} - \{\frac{-d}{c}\}$ to $\mathbb{C} - \{\frac{a}{c}\}$.

What is the connection of Möbius transformations to $2x2$ matrices?

To start off, consider the <u>complex projective plane</u> $\mathbb{CP}^1$. One way to think of $\mathbb{CP}^1$ is as the set of all one-dimensional vector subspaces of $\mathbb{C}^2$. Equivalently, we can think of $\mathbb{CP}^1$ as the set $(\mathbb{C}^2 - \{(0,0)\})/\sim$ where $(z_1, w_1) \sim (z_2, w_2)$ iff there exists $\lambda \neq 0$ such that $(z_1, w_1) = (\lambda z_2, \lambda w_2)$.

Given a line in $\mathbb{CP}^1$ passing through (z, w), we define $[z : w] := \frac{z}{w} \in \mathbb{C}_\infty$ to be the <u>homogeneous coordinates</u> identifying that line. Clearly this describes a bijective identification between $\mathbb{CP}^1$ and $\mathbb{C}_\infty$.

Next note that an invertible $2x2$ matrix with complex elements is a linear isomorphism of $\mathbb{C}^2$. In turn, it also defines a bijective map on $\mathbb{CP}^1$ as it sends lines in $\mathbb{C}^2$ to other lines in $\mathbb{C}^2$ in a bijective manner. In particular, suppose $\xi' = [z' : w']$, $\xi = [z : w]$, and $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} z \\ w \end{smallmatrix}\right) = \left(\begin{smallmatrix} z' \\ w' \end{smallmatrix}\right)$. Then:

$$\xi' = \frac{az+bw}{cz+dw} = \frac{a\xi+b}{c\xi+d} = S(\xi) \text{ where } S \text{ is a Möbius transformation.}$$

Thus, a Mobius transformation associated to a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ describes the transformation on $\mathbb{CP}^1$ induced by that matrix applied on $\mathbb{C}^2$.

Theorem: Given the triplets of distinct points $z_1, z_2, z_3 \in \mathbb{C}_\infty$ and $w_2, w_2, w_3 \in \mathbb{C}_\infty$, there exists a unique Möbius transformation S such that $S(z_i) = w_i$ for $i \in \{1, 2, 3\}$.

(Existence)
Without loss of generality we can assume $w_1 = 1$, $w_2 = 0$ and $w_3 = \infty$. This is because once we prove this case, we can by composing and inverting variations of this case prove every other case.

First suppose $z_1, z_2, z_3 \in \mathbb{C}$ and take $S(z) := \frac{z-z_2}{z-z_3}\left(\frac{z_1-z_2}{z_1-z_3}\right)^{-1}$. Then S is a Möbius transformation with $S(z_1) = 1$, $S(z_2) = 0$, and $S(z_3) = \infty$. Meanwhile:

- if $z_1 = \infty$, then define $S(z) := \frac{z-z_2}{z-z_3}$ to get that $S(\infty) = 1$, $S(z_2) = 0$, and $S(z_3) = \infty$;
- if $z_2 = \infty$, then define $S(z) := \frac{z_1-z_3}{z-z_3}$ to get that $S(z_1) = 1$, $S(\infty) = 0$, and $s(z_3) = \infty$;
- if $s_3 = \infty$, then define $S(z) := \frac{z-z_2}{z_1-z_2}$ to get that $S(z_1) = 1$, $S(z_2) = 0$, and $S(\infty) = \infty$.

(Uniqueness)
It suffices to show that if $z_1 = w_1 = 1$, $z_2 = w_2 = 0$, $z_3 = w_3 = \infty$, and S is a Möbius transformation sending z_i to w_i for $i \in \{1, 2, 3\}$, then $S = \mathrm{Id}$. After all, once we've proven this, we can then say that any two mobius transformations T_1, T_2 sending (z_1, z_2, z_3) to $(1, 0, \infty)$ must be equal (since $T_2^{-1} \circ T_1 = T_1^{-1} \circ T_2 = \mathrm{Id}$). And from there we can also say that if S maps (z_1, z_2, z_3) to (w_1, w_2, w_3) then we must have that $T' \circ S = T$ where T and T' are the unique maps sending (z_1, z_2, z_3) and (w_1, w_2, w_3) to $(1, 0, \infty)$. So, S is unique.

Now write $S(z) = \frac{az+b}{cz+d}$. If $S(\infty) = \infty$ we know that $c = 0$. Also, since $S(0) = 0$ we know that $\frac{b}{d} = 0$. And since $d \neq 0$ (since otherwise $ad - bc = 0$) we know that $b = 0$. This proves that $S(z) = \frac{az}{d}$ Finally, since $S(1) = 1$, we know $\frac{a}{d} = 1$. So $a = d$ and we have that $S(z) = z$. ■

Conway points out another succinct reason you would expect this result. Suppose $S(z) = \frac{az+b}{cz+d}$ is a Möbius transformation and $S(z) = z$ for some $z \in \mathbb{C}$. Then $cz^2 + (d-a)z - b = 0$. So, S can only have at most 2 fixed points unless $c = d - a = b = 0$, in which case S is the identity and every point is fixed.

So, we would expect that three points can uniquely determine a Möbius transformation since any transformation fixing three points has to be the identity.

Given $z_1, z_2, z_3, z_4 \in \mathbb{C}_\infty$ where z_2, z_3, z_4 are distinct, we denote the cross ratio (z_1, z_2, z_3, z_4) to be the value of $T(z_1)$ where T is the unique Möbius transformation such that $T(z_2) = 1$, $T(z_3) = 0$, and $T(z_4) = \infty$.

Proposition: If T is a Möbius transformation then:

$$(z_1, z_2, z_3, z_4) = (T(z_1), T(z_2), T(z_3), T(z_4)).$$

Proof:
Let S be the Möbius transformation taking $T(z_2)$ to 1, $T(z_3)$ to 0, and $T(z_4)$ to ∞. Then $S \circ T$ is the Möbius transformation taking z_2 to 1, z_3 to 0, and z_4 to ∞. So, we have that:

$$(T(z_1), T(z_2), T(z_3), T(z_4)) = S(T(z_1)) = (S \circ T)(z_1) = (z_1, z_2, z_3, z_4). \blacksquare$$

Math 220 Homework:

Exercise III.3.7: If $T(z) = \frac{az+b}{cz+d}$, then find z_2, z_3, z_4 (in terms of a, b, c, d) such that $Tz = (z, z_2, z_3, z_4)$.

It suffices to find values z_2, z_3, z_4 such that $T(z_2) = 1$, $T(z_3) = 0$, and $T(z_4) = \infty$. And, one easy way to find those values is to just invert T. So let $S = \frac{dz-b}{-cz+a}$. Then $S = T^{-1}$. So:

- $z_2 = S(1) = \frac{d-b}{-c+a}$,
- $z_3 = S(0) = \frac{-b}{a}$,
- $z_4 = S(\infty) = \frac{d}{-c}$.

In particular, if $-c + a = 0$, then we must have that $d - b \neq 0$ since $ad - bc \neq 0$. So in that case we'd have that $S(1) = \infty$. And by similar logic, if $a = 0$ or $c = 0$ then we'd have that $S(0)$ or $S(\infty)$ equals ∞ respectively.

Exercise III.2.2: Prove that if b_n, a_n are real and positive, $0 < b = \lim_{n\to\infty} b_n$, and $a = \limsup_{n\to\infty} a_n$, then $ab = \limsup_{n\to\infty}(a_n b_n)$.

Fix $\varepsilon > 0$ and without loss of generality assume ε is small enough that $b - \varepsilon > 0$. Also, pick $N \in \mathbb{N}$ such that $|b_n - b| < \varepsilon$.

Let us first deal with the case that $\limsup_{n\to\infty} a_n = \infty$. In that case, we know that $a_k b_k \geq a_k(b - \varepsilon)$ for all $k \geq N$. And then we know by comparison that:

$$\limsup_{n\to\infty} a_n b_n \geq (b - \varepsilon) \limsup_{n\to\infty} a_n = \infty = ab.$$

Next, the case that $a = 0$ is easy since in that case we know that $\lim_{n\to\infty} a_n = a$. And now $\limsup_{n\to\infty} a_n b_n = \lim_{n\to\infty} a_n b_n = ab$.

Finally, suppose $0 < a < \infty$. Then we may assume without loss of generality that ε is small enough that $a - \varepsilon > 0$. Also, by increasing N we can additionally assume that $|(\sup_{k\geq n} a_k) - a| < \varepsilon$ for all $n \geq N$. And now we clealy have that:

$$(a - \varepsilon)(b - \varepsilon) \leq \limsup_{n\to\infty} a_n b_n \leq (a + \varepsilon)(b + \varepsilon)$$

Taking $\varepsilon \to 0$ completes the proof.

Does this still hold if we drop the positivity requirement?

No. And an easy way to see this is to just take the positive sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ from before and negate them. Then we still have that $\limsup_{n\to\infty}(-a_n)(-b_n) = ab$. But, $a \neq \limsup_{n\to\infty}(-a_n)$ anymore. Instead, we have that $a = \liminf_{n\to\infty}(-a_n)$.

What if we let $\lim_{n\to\infty} b_n = b = 0$ but require all a_n and b_n to be positive and real.

If $\limsup_{n\to\infty} a_n = \infty$, then we can't know from just a and b what $\limsup_{n\to\infty} a_n b_n$ is. After all, let $\alpha > 0$ and define $b_n = \frac{1}{n}$ and $a_n = \alpha n$. Then $\limsup_{n\to\infty} a_n b_n = \alpha$ which can be anything. But $\lim_{n\to\infty} b_n = 0$ and $\limsup_{n\to\infty} a_n = \infty$.

Meanwhile, if $a < \infty$, then we can say that $\limsup_{n\to\infty} a_n b_n = 0 = ab$. To see this, note that since $\limsup_{n\to\infty} a_n < \infty$, we know that there exists $M > 0$ with $a_n < M$ for all $n \in \mathbb{N}$. And then $0 \leq \limsup_{n\to\infty} a_n b_n \leq M \lim_{n\to\infty} b_n = M(0) = 0$. ■

Exercise III.2.10: Let G and Ω be open in $\mathbb{C}$ and suppose f and h are functions defined on G with f continuous and h holomorphic and injective. Also suppose $g : \Omega \to \mathbb{C}$ is holomorphic with $g'(\omega) \neq 0$ for all $\omega \in \Omega$. Finally, suppose $f(G) \subseteq \Omega$ and $h(z) = g(f(z))$ for all $z \in G$. Then show that f is holomorphic and give a formula for $f'(z)$.

Fix $a \in G$ and consider any $k \in \mathbb{C}$ such that $a + k \in G$. Then since h is injective, we know that $g(f(a+k)) = h(a+k) \neq h(a) = g(f(a))$. And in turn, we know that $f(a+k) \neq f(a)$. So:

$$\frac{h(a+k)-h(a)}{k} = \frac{g(f(a+k))-g(f(a))}{f(a+k)-f(a)} \cdot \frac{f(a+k)-f(a)}{k}$$

Now $\lim_{k\to 0} \frac{h(a+k)-h(a)}{k} = h'(a)$. Also since f is continuous we get that:

$$\lim_{k\to 0} \frac{g(f(a+k))-g(f(a))}{f(a+k)-f(a)} = g'(f(a)).$$

And since $g' \neq 0$ anywhere, we in turn know that:

$$f'(a) = \lim_{k\to 0} \frac{f(a+k)-f(a)}{k} = \frac{h'(a)}{g'(f(a))}$$

Since h', g', and f are continuous and $g' \neq 0$, we also have that f' is continuous.

Exercise III.2.14: Suppose $f : G \to \mathbb{C}$ is holomorphic and that G is connected. If $f(z) \in \mathbb{R}$ for all $z \in G$ then f is constant.

By the Cauchy-Riemann equations, if $f(x+iy) = u(x,y) + iv(x,y)$ then $u_x = v_y$ and $u_y = -v_x$. But note that since $f(x+iy) \in \mathbb{R}$ for all $x+iy \in G$, we know that $v(x,y) = 0$ for all $x + iy \in G$. In particular, this means that $v_x = v_y = 0$. And so we also have that $u_x = u_y = 0$ and hence $f'(z) = 0$ for all $z \in G$. Now the fact that f is constant on G is just a result of Conway *proposition 2.10* (see the lemma on page 309). ■

We consider a circle in $\mathbb{C}_\infty$ to be either a circle (by the typical definition) in $\mathbb{C}$ or a straight line. For some justification on why, see *exercise I.6.4.*

Theorem: The image of $\mathbb{R}$ under a Möbius transformation S is a circle.

Proof:

By considering inverses, it is equivalent to show that $S^{-1}(\mathbb{R})$ is a circle for all Möbius transformations $S(z) := \frac{az+b}{cz+d}$. But note that $z \in S^{-1}(\mathbb{R})$ iff $S(z) = \overline{S(z)}$. Or in other words, $\frac{az+b}{cz+d} = \frac{\overline{az+b}}{\overline{cz+d}}$. This means that:

$$(az+b)(\overline{cz}+\overline{d}) - (cz+d)(\overline{az}+\overline{b}) = 0$$

Or in other words:

$$(a\overline{c} - c\overline{a})|z|^2 + (a\overline{d} - c\overline{b})z + (b\overline{c} - d\overline{a})\overline{z} + (b\overline{d} - d\overline{b}) = 0$$

Suppose $a\overline{c} - c\overline{a} = 0$. Then after setting $\alpha = a\overline{d} - \overline{b}c$ and $\beta = \operatorname{Im}(\overline{b}d)$ we have that $0 = \alpha z - \overline{\alpha z} + b\overline{d} - d\overline{b} = 2i\operatorname{Im}(\alpha z) - 2i\beta$. And in turn $\operatorname{Im}(\alpha z) = \beta$ for some $\alpha \in \mathbb{C}$ and $\beta \in \mathbb{R}$. This describes a line.

To see why this describes a line, write $z = x + iy$ and $\alpha = a_x + ia_y$. Then $\operatorname{Im}(\alpha z) = a_x y + a_y x = \beta$.

Meanwhile, suppose $a\overline{c} - c\overline{a} \neq 0$. Then we can say that:
$$|z|^2 + \frac{a\overline{d}-c\overline{b}}{a\overline{c}-c\overline{a}}z + \frac{b\overline{c}-d\overline{a}}{a\overline{c}-c\overline{a}}\overline{z} + \frac{b\overline{d}-d\overline{b}}{a\overline{c}-c\overline{a}} = 0$$

But note that $a\overline{c} - c\overline{a}$ is purely imaginary. So let us denote it's value by $i\gamma$. Also let $\alpha = a\overline{d} - c\overline{b}$ and $\beta = \mathrm{Im}(d\overline{b})$ be as in the case on the previous page. Then:
$$0 = |z|^2 + \frac{\alpha}{i\gamma}z - \frac{\overline{\alpha}}{i\gamma}\overline{z} - \frac{i2\beta}{i\gamma} = |z|^2 - i\frac{\alpha}{\gamma}z + i\frac{\overline{\alpha}}{\gamma}\overline{z} - \frac{2\beta}{\gamma}$$

Now suppose $\frac{\alpha}{\gamma} = x + iy$. Then:
$$-i\frac{\alpha}{\gamma} = -i(x+iy) = y - ix \text{ and } i\frac{\overline{\alpha}}{\gamma} = i(x-iy) = y + ix.$$

This all says that there is some $\eta \in \mathbb{C}$ and $\delta \in \mathbb{R}$ such that:
$$0 = |z|^2 + \overline{\eta}z + \eta\overline{z} + \delta$$

Thus $|z+\eta|^2 = |\eta|^2 - \delta$. And this is a formula for a circle with center $-\eta$ and radius $\sqrt{|\eta|^2 - \delta}$. ■

Corollary (Conway Proposition 3.10): Let z_1, z_2, z_3, z_4 be four distinct points in $\mathbb{C}_\infty$. Then $(z_1, z_2, z_3, z_4) \in \mathbb{R}$ iff all four points lie on a circle in $\mathbb{C}^\infty$.

Proof:

Let $S(z) = (z, z_2, z_3, z_4)$. We know from the last theorem that $S^{-1}(\mathbb{R}_\infty)$ is a circle in C_∞ which contains z_2, z_3, z_4 (where $\mathbb{R}_\infty = \mathbb{R} \cup \{\infty\}$). Thus $(z_1, z_2, z_3, z_4) \in \mathbb{R}$ if and only if z_1 lies in the same circle passing through z_2, z_3, z_4.

Side note: Given any three distinct points $z_1, z_2, z_3 \in \mathbb{C}_\infty$, there is a unique circle passing through all three points.

Proof:

To start off, if any $z_i = \infty$, then let our circle in $\mathbb{C}_\infty$ be the line passing through the other two points. Similarly, if all three of our points lie along a single line then let our circle in $\mathbb{C}_\infty$ be the line passing through all three points. No other line will pass through all three points and a circle (in the traditional sense) can only intercept a line at most twice. So this will be the unique "circle" in $\mathbb{C}_\infty$ passing through our points.

Next suppose z_1, z_2, z_3 are all in $\mathbb{C}$ and not collinear and express $z_i = x_i + iy_i$ for all $i \in \{1, 2, 3\}$. It's convenient to identify $\mathbb{C}$ with $\mathbb{R}^2$ here. Then any circle passing through (x_i, y_i) for each i will be described by an equation $x^2 + y^2 + ax + by + c = 0$ where $a, b, c \in \mathbb{R}$. This gives us the following matrix equation to solve:
$$\begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} -x_1^2 - y_1^2 \\ -x_2^2 - y_2^2 \\ -x_3^2 - y_3^2 \end{pmatrix}$$

Fortunately, we know $\det(\left(\begin{smallmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{smallmatrix}\right)) \neq 0$. After all, supposing our matrix didn't have a trivial kernel, we would know that there exists constants $a, b, c \in \mathbb{R}$ such that $ax_i + by_i = -c$ for all i. But that contradicts that the (x_i, y_i) aren't collinear. So, we know that there are unique coefficients satisfying our matrix equation which thus define a unique circle passing through z_1, z_2, z_3. ■

(Conway) Theorem 3.14: A Möbius transformation takes circles to circles.

Proof:

Let Γ be any circle in $\mathbb{C}_\infty$ and let S be any Möbius transformation. Then let z_2, z_3, z_4 be three distinct points on Γ and put $\omega_j = Sz_j$ for $j \in \{2,3,4\}$. Now $\omega_2, \omega_3, \omega_4$ uniquely determine a circle Γ', and we claim that $S(\Gamma) = \Gamma'$.

To see this, note that $(z, z_2, z_3, z_4) = (Sz, \omega_2, \omega_3, \omega_4)$ for all $z \in \mathbb{C}_\infty$ by a proposition a *few pages back*. In particular, by the last proposition we have that $z \in \Gamma$ iff both sides of the above expression are real iff $Sz \in \Gamma'$. In other words, $S(\Gamma) \subseteq \Gamma'$ and $S^{-1}(\Gamma') \subseteq \Gamma$. ■

Exercise III.3.9: If $T(z) = \frac{az+b}{cz+d}$, then find necessary and sufficient conditions for when $T(\Gamma) = \Gamma$ where $\Gamma = \{z : |z| = 1\}$.

(Finding Necessary Conditions):

Let $T(z) = \frac{az+b}{cz+d}$. If $T(\Gamma) = \Gamma$ we must have that $|az+b|^2 = |cz+d|^2$ whenever $|z| = 1$. But in that case we have that:

$$|a|^2 + a\overline{b}z + \overline{a}b\overline{z} + |b|^2 = |c|^2 + c\overline{d}z + \overline{c}d\overline{z} + |d|^2$$

Also, since $|z| = 1$ we have that $z \neq 0$ and $\overline{z} = z^{-1}$. Hence after some rearranging we get $(a\overline{b} - c\overline{d})z + (|a|^2 + |b|^2 - |c|^2 - |d|^2) + (\overline{a}b - \overline{c}d)\frac{1}{z} = 0$. And by multiplying by z we finally get a quadratic:

$$(a\overline{b} - c\overline{d})z^2 + (|a|^2 + |b|^2 - |c|^2 - |d|^2)z + (\overline{a}b - \overline{c}d) = 0$$

Now since we want $|T(z)| = 1$ for all $|z| = 1$ and there are clearly more than 2 elements in $\mathbb{C}$ with magnitude 1, we can conclude that a necessary condition for $T(\Gamma) = \Gamma$ is that:

$$a\overline{b} = c\overline{d} \text{ and } |a|^2 + |b|^2 = |c|^2 + |d|^2.$$

This condition is still clunkier than I'd like though. So first suppose $d = 0$. Then since $ad - bc \neq 0$ and $a\overline{b} = 0$, we must have that $a = 0$ and $b \neq 0$. Hence, we'd get a Mobius transformation of the form $T(z) = \frac{b}{cz}$. And then it is easy to check that $T(\Gamma) = \Gamma$ iff $|\frac{b}{c}| = 1$. Or in other words, $T(\Gamma) = \Gamma$ and $d = 0$ iff there exists $\theta \in [0, 2\pi)$ such that $T(z) = e^{i\theta}z^{-1}$.

Meanwhile, if $d \neq 0$ then we can scale our coefficients in order to without loss of generality say that $d = 1$. Then write $\lambda := c(\overline{b})^{-1}$ and note that $a = \lambda$ and $c = \lambda\overline{b}$. Consequently:

$$|\lambda|^2 + |b|^2 = |\lambda\overline{b}|^2 + 1 \Longrightarrow |\lambda|^2(1 - |b|^2) = 1 - |b|^2$$

It follows that $|\lambda| = 1$. And so, we can say that $T(z) = \frac{\lambda z + b}{\lambda\overline{b}z + 1} = \lambda\frac{z + \overline{\lambda}b}{\lambda\overline{b}z + 1}$.

By setting $z_0 = -\overline{\lambda}b$ and $\lambda = -e^{i\theta}$ we in turn get that $T(z) = -e^{i\theta}\frac{z - z_0}{-\overline{z_0}z + 1} = e^{i\theta}\frac{z - z_0}{\overline{z_0}z - 1}$.

And clearly $T(z_0) = 0$. Hence, the final necessary conditions I will settle on are that if $T(\Gamma) = \Gamma$ then either:

- $T(z) = e^{i\theta}\frac{1}{z}$ (where $\theta \in [0, 2\pi)$),
- $T(z) = e^{i\theta}\frac{z - z_0}{\overline{z_0}z - 1}$ (where $\theta \in [0, 2\pi)$ and $T(z_0) = 0$).

(Proving our prior conditions are sufficient):
I already showed that if $T(z) = e^{i\theta}\frac{1}{z}$ then $T(\Gamma) = \Gamma$. Meanwhile, suppose $T(z) = e^{i\theta}\frac{z-z_0}{\overline{z_0}z-1}$. Then if $|z| = 1$ we have that:

$$|T(z)|^2 = \frac{|z-z_0|^2}{|\overline{z_0}z-1|^2} = \frac{1-z\overline{z_0}-\overline{z}z_0+|z_0|^2}{|z_0|^2-z\overline{z_0}-\overline{z}z_0+1} = 1$$

So $T(\Gamma) = \Gamma$. ■

Math 200a Notes:

Note that for all $n \in N$ and $h \in H$, we have (in $N \rtimes_f H$) that:

$$\begin{aligned} \underline{h}\,\underline{n}\,\underline{h}^{-1} = (1,h)\cdot(n,1)\cdot(1,h^{-1}) &= ((f(h))(n),h)\cdot(1,h^{-1}) \\ &= ((f(h))(n)(f(h))(1), hh^{-1}) = ((f(h))(n),1) \\ &= \underline{(f(h))(n)} \end{aligned}$$

In particular, $f : H \to \mathrm{Aut}(N)$ is trivial iff $\underline{h}\,\underline{n}\,\underline{h}^{-1} = \underline{n}$ for all $h \in H$ and $n \in N$. Or in other, $\underline{h}\,\underline{n} = \underline{n}\,\underline{h}$ for all $h \in H$ and $n \in N$ iff f is trivial. But now note that if $\underline{h}\,\underline{n} = \underline{n}\,\underline{h}$ for all $h \in H$ and $n \in N$, then for any $h_1, h_2 \in H$ and $n_1, n_2 \in N$ we have that:

$$\begin{aligned} (n_1,h_1)\cdot(n_2,h_2) &= (n_1,1)\cdot(1,h_1)\cdot(n_2,1)\cdot(1,h_2) \\ &= (n_1,1)\cdot(n_2,1)\cdot(1,h_1)\cdot(1,h_2) = (n_1n_2, h_1h_2) \end{aligned}$$

Therefore $N \rtimes_f H = N \times H$ iff f is trivial.

Suppose G is a finite group with $|G| = pq$ where $p < q$ are primes. Then there is a unique Sylow q-subgroup Q. So consider the standard S.E.S.:

$$\{1\} \longrightarrow Q \longrightarrow G \longrightarrow G/Q \longrightarrow \{1\}$$

If P is a Sylow p-subgroup, we know that $P \cap Q = \{1\}$ and therefore that $PQ = G$. Hence this S.E.S. splits. (In particular, we can take $pQ \mapsto p$ (where $p \in P$) to be our backwards arrow).

Identifying Q with $\mathbb{Z}/q\mathbb{Z}$ and P with $\mathbb{Z}/p\mathbb{Z}$, we then have that $G \cong (\mathbb{Z}/q\mathbb{Z}) \rtimes_f (\mathbb{Z}/p\mathbb{Z})$ where $f : \mathbb{Z}/p\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/q\mathbb{Z})$ is a group homomorphism. But recall from problem 5 on homework set 1 (see *pages 244-245*) that $\mathrm{Aut}(\mathbb{Z}/q\mathbb{Z}) \cong (\mathbb{Z}/q\mathbb{Z})^\times \cong (\mathbb{Z}/(q-1)\mathbb{Z})$. Therefore, we've shown that there is a surjective map from $\mathrm{Hom}(\mathbb{Z}/p\mathbb{Z}, \mathbb{Z}/(q-1)\mathbb{Z})$ to the collection of all pq-groups.

Set 4 Problem 1: Suppose D_{2n} is a dihedral group. Prove that there exists a splitting S.E.S. of the form $\{1\} \to C_n \to D_{2n} \to C_2 \to \{1\}$ where C_k denotes the cyclic group of order k.

Let σ and τ be the elements of D_{2n} satisfying that $o(\sigma) = n$, $o(\tau) = 2$, $\tau\sigma\tau = \sigma^{-1}$, and $D_{2n} = \langle \sigma, \tau \rangle$. Then $C_n \cong \langle \sigma \rangle$ and $\langle \sigma \rangle \lhd D_{2n}$. It follows that $D_{2n}/\langle \sigma \rangle$ is a quotient group of order 2 (which means it is cyclic).

Hence, we have a well defined S.E.S. $\{1\} \to C_n \to D_{2n} \to C_2 \to \{1\}$ (which is isomorphic to the standard S.E.S. of D_{2n} and $\langle\sigma\rangle$). Finally, $\langle\sigma\rangle \cap \langle\tau\rangle = \{1\}$ and $\langle\sigma\rangle\langle\tau\rangle = D_{2n}$. Therefore, the S.E.S. splits.

Set 4 Problem 2: Suppose G is a group.

(a) Show that if N_1 and N_2 are normal subgroups of G and $N_1 \cap N_2 = \{1\}$, then for all $x_1 \in N_1$ and $x_2 \in N_2$, $x_1x_2 = x_2x_1$.

This was already proved earlier in my notes (see *page 299*). For the grader I guess I'll reprove it here.

It's equivalent to show that $x_1x_2x_1^{-1}x_2^{-1} = 1$ for all $x_1 \in N_1$ and $x_2 \in N_2$. Fortunately, since $N_2 \lhd G$, we know that $(x_1x_2x_1^{-1})x_2^{-1} \in N_2$. And similarly, since $N_1 \lhd G$, we know that $x_1(x_2x_1^{-1}x_2^{-1}) \in N_1$. So $x_1x_2x_1^{-1}x_2^{-1} \in N_1 \cap N_2 = \{1\}$.

(b) Suppose $N_1, \ldots, N_k$ are normal subgroups of G and $N_i \cap N_j = \{1\}$ for all $i \neq j$. Then prove that $f : \prod_{i=1}^k N_i \to N_1 \cdots N_k$ where $f(x_1, \ldots, x_k) = x_1 \cdots x_k$ is a group homomorphism.

Let $(x_1, \ldots, x_k)$ and $(y_1, \ldots, y_k)$ be elements of $\prod_{i=1}^k N_i$. Then we claim that:
$$f(x_1, \ldots, x_k)f(y_1, \ldots, y_k) = f(x_1y_1, \ldots, x_ky_k)$$

To prove this, we annoyingly have to do induction. And to make it so that we don't need to repeat a lot of reasoning in both our base case and our inductive step, we'll temporarily assume $x_1 = y_1 = 1$ and $N_1 = \{1\}$. That way we trivially have that $x_1x_2 \cdots x_ky_1y_2 \cdots y_k = (x_1y_1)x_2 \cdots x_ky_2 \cdots y_k$.

Now assume by induction we already know for $1 \leq j < k$ that:
$$x_1x_2 \cdots x_ky_1y_2 \cdots y_k = (x_1y_1) \cdots (x_jy_j)x_{j+1} \cdots x_ky_{j+1} \cdots y_k.$$

Then by part (a) we know $x_iy_{j+1} = y_{j+1}x_i$ for all $i > j+1$. And by applying this we get that $x_1x_2 \cdots x_ky_1y_2 \cdots y_k = (x_1y_1) \cdots (x_jy_j)x_{j+1}y_{j+1}x_{j+2} \cdots x_ky_{j+2} \cdots y_k$.
■

(c) Suppose $N_1, \ldots, N_k \lhd G$ and $N_i \cap (N_1 \cdots N_{i-1}N_{i+1} \cdots N_k) = \{1\}$ for all i. Then prove that our group homomorphism f from before is a group isomorphism.

Surjectivity is trivial from the definition of f and $N_1N_2 \cdots N_k$. So, we just need to prove injectivity. If G is finite, then we can employ the following counting argument (I don't want to remove it from my notes now that I've written it down):

Note that $(N_1 \cdots N_{j-1}) \cap N_j \subseteq (N_1 \cdots N_{j-1}N_{j+1} \cdots N_k) \cap N_j = \{1\}$. Therefore, for all $j > 1$ we have that:
$$|N_1 \cdots N_j| = \frac{|N_1 \cdots N_{j-1}||N_j|}{|(N_1 \cdots N_{j-1}) \cap N_j|} = |N_1 \cdots N_{j-1}||N_j|$$

And hence $|N_1 \cdots N_k| = |N_1||N_2| \cdots |N_k| = |\prod_{i=1}^k N_i|$. By the pigeonhole principle, we know that f is injective since f is surjective.

Meanwhile, here is an argument that works if G is infinite. Suppose $x_1x_2 \cdots x_k = 1$ where $x_i \in N_i$ for each i. Then we can prove by induction as follows that $x_i = 1$ for all i.

Let $j \in \{1, \ldots, n-1\}$ and suppose we've shown $x_i = 1$ for $i < j$ and that $x_j x_{j+1} \cdots x_k = 1$. Then:

$$N_j \cap (N_{j+1} \cdots N_k) \subseteq N_j \cap (N_1 \cdots N_{j-1} N_{j+1} \cdots N_k) = \{1\}.$$

Also $x_j^{-1} = x_{j+1} \cdots x_k \in N_j \cap (N_{j+1} \cdots N_k)$. Therefore $x_j = 1$ and $x_{j+1} \cdots x_k = 1$. ■

Set 4 Problem 3: Suppose G is a finite group and for every proper subgroup H we have that $H \lneqq N_G(H)$.

(a) Prove that all the Sylow subgroups of G are normal and therefore for all prime divisors p of $|G|$ there is a unique Sylow p-subgroup.

To start off, fix p as any prime divisor of $|G|$ and then consider any group $P \in \mathrm{Syl}_p(G)$. To show that P is normal, it suffices to prove that $N_G(P) = G$. But this is easy. After all, suppose for the sake of contradiction that $N_G(P) \lneqq G$. Then that would imply that $N_G(P) \lneqq N_G(N_G(P))$ by our problem's assumption. But since P is a Sylow p-subgroup, we know that $N_G(P) = N_G(N_G(P))$ by the second proposition on *page 298*. This is a contradiction.

(b) Prove that $G \cong \prod_{p \text{ is a prime factor of } |G|} P_p$ where P_p is the unique Sylow p-subgroup of G.

Let $p_1, \ldots, p_n$ be all the unique prime factors dividing $|G|$. Then it is clear that $P_{p_i} \cap P_{p_j} = \{1_G\}$ for all $i \neq j$. After all, P_{p_i} contains exactly all the elements of G whose order is some power of p_i and P_{p_j} contains exactly all the elements of G whose order is some power of p_j. But the only element whose order is both a power of p_i and p_j is 1_G.

Consequently, we can now apply part (b) of problem 2 on this homework set to say that if $x_i \in P_{p_i}$ for all i then $(x_1 x_2 \cdots x_n)^k = x_1^k x_2^k \cdots x_n^k$. Using this fact, we next claim that for all $j \in \{1, \ldots, n\}$ we have that $P_j \cap (P_{p_1} \cdots P_{p_{j-1}} P_{p_{j+1}} \cdots P_n) = \{1_G\}$.

Why?

Suppose $x = x_1 \cdots x_{j-1} x_{j+1} \cdots x_n \in P_{p_1} \cdots P_{p_{j-1}} P_{p_{j+1}} \cdots P_n$. Then since $x^k = x_1^k \cdots x_{i-1}^k x_{i+1}^k \cdots x_n^k$, we know that:

$$o(x) \mid \mathrm{lcm}(o(x_1), \ldots, o(x_{j-1}), o(x_{j+1}), \ldots, o(x_n))$$

But now it is easy to see that $o(x_j)$ divides $\frac{|G|}{(p_i)^{\nu_{p_i}(|G|)}}$ for all $j \neq i$.

This proves that the order of any $x \in P_{p_1} \cdots P_{p_{j-1}} P_{p_{j+1}} \cdots P_n$ will not have p_j as a prime factor. Yet at the same time, any $y \in P_j - \{1_G\}$ will have p_j as a prime factor. This proves $P_j \cap (P_{p_1} \cdots P_{p_{j-1}} P_{p_{j+1}} \cdots P_n) = \{1_G\}$.

By part (c) of problem 2 on this homework set, we now know that:

$$\prod_{p \text{ is a prime factor of } |G|} P_p \cong P_{p_1} \cdots P_{p_n}$$

Finally, to see that $P_{p_1} \cdots P_{p_n} = G$, note that $\left|\prod_{p \text{ is a prime factor of } |G|} P_p\right| = |G|$. And since $P_{p_1} \cdots P_{p_n} \subseteq G$ we are done. ■

This next theorem was partially proven throughout an entire lecture (a week ago AUGH) with the final length of the proof being assigned as homework. So here we go:

Before getting into the next theorem, here is a result from Dummit & Foote:

Third Isomorphism Theorem: Let G be a group and let H and K be normal subgroups of G with $H < K$. Then $K/H \lhd G/H$ and $(G/H)/(K/H) \cong G/K$.

Proof:

Define $\phi : G/H \mapsto G/K$ given by $gH \mapsto gK$. To see that ϕ is a well-defined map, note that if $g_1 H = g_2 H$, then because $H \subseteq K$, we know that $g_1 \in g_2 H \subseteq g_2 K$. So, $g_1 K = g_2 K$.

Next, since g can be any element of G, hopefully it's obvious that ϕ is surjective. Also note that $\phi(gH) = K$ iff $gK = K$, and that happens precisely when $g \in K$. So $\ker(\phi) = \{gH : g \in K\} = K/H$. Hence, by the first isomorphism theorem we have that $(G/H)/(K/H) \cong G/K$. ■

Schur-Zassenhaus Theorem: An S.E.S. $\{1\} \to G_1 \to G_2 \to G_3 \to \{1\}$ splits if $\gcd(|G_1|, |G_2|) = 1$.

Proof:

To start off, every S.E.S. is isomorphic to a standard S.E.S. So without loss of generality we may assume we are working with a standard S.E.S. $\{1\} \to N \to G \to G/N \to \{1\}$.

While I hadnt mentioned it before, hopefully it's obvious how by just following arrows you can show that an S.E.S. isomorphic to a splitting S.E.S. is also splitting.

So in other words, we want to show that if $N \lhd G$ and $\gcd(|N|, |G/N|) = 1$ then there exists $H < G$ with $H \cap N = \{1\}$ and $HN = G$.

Lemma: If $N \lhd G$, $\gcd(|N|, |G/N|) = 1$, and $H < G$ with $|H| = |G/N|$, then $H \cap N = \{1\}$ and $HN = G$.

Proof:

$|N \cap H|$ divides $\gcd(|N|, |H|) = \gcd(|N|, |G/N|) = 1$. So $N \cap H = \{1\}$.
And in turn we have that $|HN| = \frac{|H||N|}{|H \cap N|} = |G/N||N| = |G|$. ■

We say (G, N) is an SZ-pair if $N \lhd G$ and $\gcd(|N|, |G/N|) = 1$. By the prior lemma, it suffices to show that for any SZ-pair there exists $H < G$ with $|H| = |G/N|$ in order to prove our desired theorem.

We proceed by strong induction on $|G|$, noting that if $|G| = 1$ the theorem is obvious.

Claim 1: If N is not a minimal normal subgroup G, then there exists H.

Proof:

Suppose N is not a minimal normal subgroup and let $N' \lhd G$ be such that $\{1\} \subsetneqq N' \subsetneqq N$. Then we claim that $(G/N', N/N')$ is an SZ-pair.

Why?
Note that $N/N' \lhd G/N'$ and $|(G/N')/(N/N')| \cong |G/N|$ by the third isomorphism theorem. Also $|N/N'|$ divides $|N|$. And since $\gcd(|N|, |G/N|) = 1$ we know $\gcd(|N/N'|, |(G/N')/(N/N')|) = 1$.

Since $|G/N'| < |G|$, we can conclude by strong induction that there exists $\overline{H} < G/N'$ such that $|\overline{H}| = |(G/N')/(N/N')| = |G/N|$. And in turn, by the correspondance theorem there exists a subgroup $\widetilde{H} < G$ containing N' such that $\overline{H} = \widetilde{H}/N'$. Finally, since $N' \lneq N$ we know that $|\widetilde{H}| < |G|$. And since $(\widetilde{H}, N')$ is an SZ-pair (as $\gcd(|N'|, |\widetilde{H}/N'|) \mid \gcd(|N|, |G/N|) = 1$), we know by one more instance of strong induction that there exists $H < \widetilde{H}$ such that $|H| = |\widetilde{H}/N'| = |G/N|$.

Claim 2: If N is a minimal normal subgroup but not a p-group, then there exists H.
Proof:
Suppose N is not a p-group and pick a prime p dividing $|N|$. Then let P be a Sylow p-subgroup of N and note that $P \neq N$.

Recall problem 2 on the third problem set (see *page 310*) in which we proved that $G = N_G(P)N$ $(= NN_G(P))$. (As a side note, this is callled Frattini's Trick). Thus, consider that $G/N = (N_G(P)N)/N \cong N_G(P)/(N_G(P) \cap N)$ by the second isomorphism theorem.

It then follows that $(N_G(P), N_G(P) \cap N)$ is an SZ-pair since:
$$|N_G(P)/(N_G(P) \cap N)| = |G/N| \text{ and } N_G(P) \cap N \subseteq N.$$

Also, we know that $N_G(P) \neq G$ since the alternative would contradict that N is minimal. So by strong induction we know there is a subgroup $H < N_G(P) < G$ such that $|H| = |N_G(P)/((N_G(P) \cap N))| = |G/N|$.

Finally, we may now assume N is a minimal normal subgroup of G and a p-group. But note that this actually forces N to be abelian. To see why, first consider the following lemma:

Lemma: If $N \lhd G$ then $Z(N) \lhd G$.
Proof:
Suppose $z \in Z(N)$ and $x \in G$. Because $z \in N$ we know that $xzx^{-1} \in N$. And, if we can show that $(xzx^{-1})y = y(xzx^{-1})$ for any $y \in N$ then we will be done. Fortunately though, note that $x^{-1}yx \in N$ for all $y \in N$. Therefore:
$$(xzx^{-1})y = xz(x^{-1}yx)x^{-1} = x(x^{-1}yx)zx^{-1} = y(xzx^{-1}). \blacksquare$$

Now since N is a p-group, we know $Z(N)$ is not trivial. And since $Z(N) \lhd G$ and N is a minimal normal subgroup, we must have that $Z(N) = N$. Hence N is abelian. And the rest of the proof is now the subject of the homework problem below.

Set 4 Problem 4: Suppose G is a finite group and A is a normal abelian subgroup of G. Let us call any right inverse $s : G/A \to G$ of the projection map a section of the natural projection map. (This means if $gA \in G/A$ then $s(gA)$ returns an element of gA). Notably, the standard S.E.S. $\{1\} \to A \to G \to G/A \to \{1\}$ splits if s is a group homomorphism. So roughly speaking, our goal will be to start with any given section s and then modify it to get an actual homomorphism.

Let $H := G/A$ and define $c : H \times H \to A$ by $c(h_1, h_2) := s(h_1)s(h_2)s(h_1h_2)^{-1}$.

To see that c actually maps $H \times H$ into A, note that for any $h_1, h_2 \in H$ we have that:
$$s(h_1h_2)A = h_1h_2 = s(h_1)As(h_2)A = s(h_1)s(h_2)A.$$

It follows that $s(h_1)s(h_2)s(h_1h_2)^{-1}A = A$. So $s(h_1)s(h_2)s(h_1h_2)^{-1} \in A$.

You can think of c as measuring how much like a group homomorphism s is. After all, s is a group homomorphism precisely if $c(h_1, h_2) = 1_G$ for all $h_1, h_2 \in H$.

Next, notice that since A is normal and abelian, we have a well defined group action $H \curvearrowright A$ given by $h \cdot a = s(h)as(h)^{-1}$.

We know $h \cdot a \in A$ because A is normal. Also note that $s(1_H) \in A$ and therefore:
$$1_H \cdot a = s(1_H)as(1_H)^{-1} = as(1_H)s(1_H)^{-1} = a.$$

Meanwhile, suppose $h_1, h_2 \in H$ and then consider that:
$$\begin{aligned} h_1 \cdot (h_2 \cdot a) &= s(h_1)s(h_2)as(h_2)^{-1}s(h_1)^{-1} \\ &= c(h_1, h_2)s(h_1h_2)as(h_1h_2)^{-1}c(h_1, h_2)^{-1} \\ &= s(h_1h_2)as(h_1h_2)^{-1}c(h_1, h_2)c(h_1, h_2)^{-1} = s(h_1h_2)as(h_1h_2)^{-1} = (h_1h_2) \cdot a \end{aligned}$$

Now here is where the actual exercise part starts:

(a) Prove for all $h_1, h_2, h_3 \in H$ that $c(h_1, h_2)c(h_1h_2, h_3) = (h_1 \cdot c(h_2, h_3))c(h_1, h_2h_3)$.

Note that since A is abelian, this equation is more commonly written as:
$$c(h_1, h_2) + c(h_1h_2, h_3) = h_1 \cdot c(h_2, h_3) + c(h_1, h_2h_3).$$

Also, this equation is called the <u>2-cocycle relation</u>. And since c satisfies it, c is a <u>2-cocycle</u>.

To start off, note that:
$$\begin{aligned} c(h_1, h_2)c(h_1h_2, h_3) &= s(h_1)s(h_2)s(h_1h_2)^{-1}s(h_1h_2)s(h_3)s(h_1h_2h_3)^{-1} \\ &= s(h_1)s(h_2)s(h_3)s(h_1h_2h_3)^{-1} \end{aligned}$$

But also note that since $s(h)s(h') = c(h, h')s(hh')$ for all $h, h' \in H$, we can say that:
$$\begin{aligned} s(h_1)s(h_2)s(h_3) &= s(h_1)c(h_2, h_3)s(h_2h_3) \\ &= s(h_1)c(h_2, h_3)s(h_1)^{-1}s(h_1)s(h_2h_3) \\ &= (h_1 \cdot c(h_2, h_3))s(h_1)s(h_2h_3) = (h_1 \cdot c(h_2, h_3))c(h_1, h_2h_3)s(h_1h_2h_3) \end{aligned}$$

Therefore:
$$c(h_1, h_2)c(h_1h_2, h_3) = s(h_1)s(h_2)s(h_3)s(h_1h_2h_3)^{-1} = (h_1 \cdot c(h_2, h_3))c(h_1, h_2h_3).$$

(b) Prove that the standard S.E.S. $\{1\} \to A \to G \to H \to \{1\}$ splits if and only if there exists a function $b : H \to A$ satisfying that $c(h_1, h_2) = b(h_1)(h_1 \cdot b(h_2))b(h_1h_2)^{-1}$.

Like before it is customary to write this equation in additive notation:
$$c(h_1, h_2) = b(h_1) + h_1 \cdot b(h_2) - b(h_1h_2)$$

Also, if b satisfying this equation exists than we would call c satisfying this a 2-coboundary. With this in mind, part (b) is saying that if c is a 2-coboundary than our S.E.S. splits.

$(\Longrightarrow)$
Suppose $\psi : H \to G$ is a group homomorphism satisfying that $\psi(h)A = h$ for all $h \in H$. Then define $b : H \to A$ by $b(h) = s(h)\psi(h)^{-1}$.

To see that b is really mapping H into A, note that:

$$b(h)A = s(h)\psi(h)^{-1}A = s(h)h^{-1}.$$

But $s(h) \in h$ since s is a section. So $s(h)h^{-1} = 1_H = A$. And this proves that $b(h) \in A$.

But now note that $\psi(h) = b(h)^{-1}s(h)$. Therefore:

$$\begin{aligned} 1_G = \psi(h_1)\psi(h_2)\psi(h_1h_2)^{-1} &= b(h_1)^{-1}s(h_1)b(h_2)^{-1}s(h_2)\left(b(h_1h_2)^{-1}s(h_1h_2)\right)^{-1} \\ &= b(h_1)^{-1}s(h_1)b(h_2)^{-1}s(h_2)s(h_1h_2)^{-1}b(h_1h_2) \\ &= b(h_1)^{-1}s(h_1)b(h_2)^{-1}s(h_1)^{-1}c(h_1,h_2)b(h_1h_2) \\ &= b(h_1)^{-1}(s(h_1)b(h_2)s(h_1)^{-1})^{-1}c(h_1,h_2)b(h_1h_2) \\ &= b(h_1)^{-1}(h_1 \cdot b(h_2))^{-1}c(h_1,h_2)b(h_1h_2) \end{aligned}$$

This proves that $c(h_1, h_2) = (h_1 \cdot b(h_2))b(h_1)b(h_1h_2)^{-1}$. And finally, since A is abelian we can reorder the terms in the last expression to get the desired relation.

$(\Longleftarrow)$
Suppose there exists $b : H \to A$ such that $c(h_1, h_2) = b(h_1)(h \cdot b(h_2))b(h_1h_2)^{-1}$. Then by reversing the scratch work in the previous proof direction, we can show that $1_G = b(h_1)^{-1}s(h_1)b(h_2)^{-1}s(h_2)\left(b(h_1h_2)^{-1}s(h_1h_2)\right)^{-1}$ for any $h_1, h_2 \in H$. Hence, if we define $\psi : H \to G$ by $\psi(h) = b(h)^{-1}s(h)$ then we know:

$$\psi(h)\psi(h') = \psi(hh').$$

Also note that $\psi(1_H) = 1_G$.

Why:
We know $c(1_H, h) = b(h)b(1_H)b(h)^{-1} = b(1_H)$ for all h. Also, $c(1_H, h) = s(1_H)s(h)s(1_H h)^{-1} = s(1_H)$ for all h. Thus $b(1_H) = s(1_H)$ and in turn we have that $\psi(1_H) = b(1_H)^{-1}s(1_H) = 1_G$.

And now that we know that ψ is a group homomorphism, we finally claim that $\psi(h)A = h$.

To see this, note that $\psi(h)A = b(h)^{-1}s(h)A$. But because $b(h)^{-1} \in A$, we can equivalently say that $\psi(h)A = s(h)A$. And since s is a section, we know $s(h)A = h$.

And this proves that the S.E.S. $\{1\} \to A \to G \to H \to \{1\}$ splits.

(c) Now assume that $\gcd(|A|, |H|) = 1$ and prove that every 2-cocycle is a 2-coboundary. (By parts (a) and (b) this then proves that our short exact sequence splits).

The provided hints say that now I should switch over to using additive notation. So, I will do that.

Let $c' : H \times H \to A$ be a 2-cocycle, meaning that:

$$c'(h_1, h_2) + c'(h_1h_2, h_3) = h_1 \cdot c'(h_2, h_3) + c'(h_1, h_2h_3)$$

Since $\gcd(|A|, |H|) = 1$, we know that there is a unique $y \in A$ such that $a = |H|y = y + \cdots + y$ [$|H|$ times].

Why?

Define the map $f : A \to A$ by $f(y) = |H|y$. Then f is a group homomorphism because A is abelian. Also, f is injective. After all, if $y \in \ker(f)$ then we know that $o(y) \mid |H|$. But we also know by Lagrange's theorem that $o(y) \mid |A|$. Since $\gcd(|A|, |H|) = 1$, this means that $o(y) = 1$. So, $\ker(f) = \{0_A\}$ and we've shown f is injective. Finally, the fact that f is surjective is evident from the pigeonhole principle.

But now since $f : A \to A$ is a bijection, we know for any $a \in A$ that there is a unique $y \in A$ such that $f(y) = a$.

As a side note, since f is a group isomorphism, we have that if $f(x) = a$ and $f(y) = b$, then $f^{-1}(a + b) = x + y$.

Let us denote that y as $\frac{a}{|H|}$. Then define the map $b : H \to A$ by:

$$b(x) := \frac{\sum_{h \in H} c'(x,h)}{|H|} = \sum_{h \in H} \frac{c'(x,h)}{|H|}.$$

Note that because the action $H \curvearrowright A$ is defined using conjugation, if $h \in H$ and $a_1, a_2 \in A$ then $h \cdot (a_1 + a_2) = (h \cdot a_1) + (h \cdot a_2)$. Also, if $h_2 \in H$, then the map $h \mapsto h_2 h$ is a bijective map on H. Therefore, for any $h_1, h_2 \in G$ we have:

$$\begin{aligned}
|H|b(h_1) &= \sum_{h \in H} c'(h_1, h) \\
&= \sum_{h \in H} c'(h_1, h_2h) = \sum_{h \in H} (c'(h_1, h_2) + c'(h_1h_2, h) - h_1 \cdot c'(h_2, h)) \\
&= |H|c'(h_1, h_2) + \sum_{h \in H} c'(h_1h_2, h) - h_1 \cdot (\sum_{h \in H} c'(h_2, h)) \\
&= |H|c'(h_1, h_2) + |H|b(h_1h_2) - h_1 \cdot (|H|b(h_2)) \\
&= |H|c'(h_1, h_2) + |H|b(h_1h_2) - |H|(h_1 \cdot b(h_2))
\end{aligned}$$

And finally, by canceling out $|H|$ and moving terms over, we can see that c' is a 2-coboundary. ■

10/26/2025

Today I'm now going to continue reading through Folland chapter 11. Go to *page 335* to see where I left off. Also, for the time being, I'll use e to denote the identity element of a topological group since that's what Folland does.

Before getting to the next theorem, I need to learn some more about quotient topologies.

Lemma 1: If $f : X \to Y$ is a continuous map, then $f \times f : X \times X \to Y \times Y$ (given by $f \times f(x_1, x_2) = (f(x_1), f(x_2))$) is a continuous map.

If π_1, π_2 are the natural projection maps from $Y \times Y$ to Y, then we have that $f \times f$ is continuous iff $\pi_1 \circ (f \times f)$ and $\pi_2 \circ (f \times f)$ is continuous. But clearly if $U \subseteq Y$ is an open set then $(\pi_1 \circ (f \times f))^{-1} = f^{-1}(U) \times X$ is open in $X \times X$. And a similar statement holds for $\pi_2 \circ (f \times f)$. Hence $f \times f$ is continuous.

Lemma 2: If $p : X \to Y$ is an open continuous map, then $p \times p : X \times X \to Y \times Y$ is also an open continuous map.

We know from the last lemma that p is continuous. Meanwhile, suppose $U \subseteq X \times X$ is open. Then, we know there is a collection $\{V_\alpha \times W_\alpha\}_{\alpha \in A}$ of rectangles in $X \times X$ such that V_α, W_α are open in X and $U = \bigcup_{\alpha \in A}(V_\alpha \times W_\alpha)$. Then in turn:

$$p \times p(U) = p \times p(\bigcup_{\alpha \in A}(V_\alpha \times W_\alpha)) = \bigcup_{\alpha \in A} p \times p(V_\alpha \times W_\alpha) = \bigcup_{\alpha \in A}(p(V_\alpha) \times p(W_\alpha))$$

But since p is open, we know $p(V_\alpha) \times p(W_\alpha)$ is open in $Y \times Y$. So, $p \times p(U)$ is a union of open sets. ■

Corollary 3: If $p : X \to Y$ is an open quotient map then $p \times p : X \times X \to Y \times Y$ is also an open quotient map.

As for why this will be relevant, suppose G is a topological group and $H \lhd G$. Then if we consider the natural projection map $\pi : G \to G/H$ and equip G/H with the quotient topology (so that π is a quotient map), then we have that π is an open map.

Why?

Suppose $U \subseteq G$ is open. Then $\pi(U)$ is open in g/H if and only if $\pi^{-1}(\pi(U))$ is open in G. But note that $\pi^{-1}(\pi(U)) = \{x \in G : \exists y \in U \ s.t. \ x \in yH\} = UH = \bigcup_{h \in H} Uh$. Since each Uh is open (since G is a topological group), we thus have that $\pi^{-1}(\pi(U))$ is open. ■

Theorem 4: Suppose $X, Y, \overline{X}, \overline{Y}$ are topological spaces, $p_1 : X \to \overline{X}$ and $p_2 : Y \to \overline{Y}$ are quotient maps, and $f, \overline{f}$ are functions such that the following diagram commutes:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "p_1"'] & Y \arrow[d, "p_2"] \\ \overline{X} \arrow[r, "\overline{f}"] & \overline{Y} \end{tikzcd}$$

If f is continuous then so is $\overline{f}$.

Proof:

Let $U \subseteq \overline{Y}$ be open. Then $p_1^{-1}(\overline{f}^{-1}(U)) = (p_2 \circ f)^{-1}(U)$ is open since f and p_2 are continuous. And since p_1 is a quotient map, we thus know that $\overline{f}^{-1}(U)$ is open. This shows that $\overline{f}$ is continuous. ■

Also I'm dumb because this theorem can be simply restated as an application of the universal property of quotient maps.

Corollary 5: Suppose G is a topological group and $H \lhd G$. Then G/H is a topological group when equipped with the quotient topology (induced by the natural projection map $\pi : G \to G/H$).

By lemma 2 and corollary 3, we know that $\pi \times \pi : G \times G \to G/H \times G/H$ is a quotient map. Also, it is easy to verify that the following diagrams commute:

$$\begin{tikzcd} G \times G \arrow[r, "(x,y)\mapsto xy"'] \arrow[d, "\pi\times\pi"] & G \arrow[d, "\pi"'] \\ G/H \times G/H \arrow[r, "(xH,yH)\mapsto xyH"] & G/H \end{tikzcd} \qquad \begin{tikzcd} G \arrow[r, "x\mapsto x^{-1}"'] \arrow[d, "\pi"] & G \arrow[d, "\pi"'] \\ G/H \arrow[r, "xH\mapsto (xH)^{-1}"] & G/H \end{tikzcd}$$

Therefore, by theorem 4 we know that the group operations on G/H are continuous and so G/H is a topological group. ■

Proposition 11.3: Let G be a topological group.

(a) If G is T_1, then G is Hausdorff.

Proof:
If G is T_1 and $x \neq y \in G$, then we can find an open set U containing e but not containing xy^{-1}. And in turn, by *Proposition 11.1(c) and (d)* there is a symmetric neighborhood V of e such that $xy^{-1} \notin VV$.

Now we know Vx and Vy are neighborhoods of x and y. Also, we know $Vx \cap Vy = \emptyset$. After all, if this weren't true than there would exist $v_1, v_2 \in V$ with $v_1 x = v_2 y$. And in turn we'd have that $xy^{-1} = v_1^{-1} v_2$. But since V is symmetric and $xy^{-1} \notin VV$, this is a contradiction.

(b) If G is not T_1, then let H be the closure of $\{e\}$. Then H is a normal subgroup, and if G/H is given the quotient topology (i.e. $A \subseteq G/H$ is open if and only if $B = \{x \in G : xH \in A\}$ is open), then G/H is a Hausdorff topological group.

Proof:
H is a subgroup by Proposition 11.1(d). Also, to see why H is normal, suppose H' is a conjugate of H with $H' \neq H$. Because the group product is continuous, we know that H' is also closed. And since $e \in H'$, we know that $H' \cap H$ is a closed proper subset of H containing $\{e\}$. But this contradicts that H is the closure of $\{e\}$. Hence, we conclude H is normal.

Now that we know H is a normal subgroup, we can consider the quotient group G/H equipped with the quotient topology. I already proved that G/H is a topological group. Also note that if $\overline{e}$ is the identity of G/H then $\{\overline{e}\}$ is closed in G/H. (This is just a result of the definition of a quotient topology plus the fact that H is closed). In turn, we can see that every singleton in G/H is closed. So, G/H is T_1. And by part (a) we have that G/H is Hausdorff. ■

The prior theorem shows that it is not much of a restriction to assume our topological groups are Hausdorff. Hence, we now define that G is a locally compact group if G is an LCH topological group.

Suppose G is a locally compact group. Then a Borel measure μ on G is called left invariant (or right invariant) if $\mu(xE) = \mu(E)$ (or $\mu(Ex) = \mu(E)$) for all $x \in G$ and $E \in \mathcal{B}_G$. Similarly, a linear functional I on $C_c(G)$ is called left- or right-invariant if $I(L_x f) = I(f)$ or $I(R_x f) = I(f)$ for all $f \in C_c(G)$. Finally, a left (or right) Haar measure on G is a nonzero left-invariant (or right-invariant) Radon measure μ on G.

Let $C_c^+(G) := \{f \in C_c(G) : f \geq 0 \text{ and } \|f\|_u > 0\}$.

Proposition 11.4: Let G be a locally compact group.

(a) A Radon measure μ on G is a left Haar measure iff the measure $\widetilde{\mu}$ defined by $\widetilde{\mu}(E) = \mu(E^{-1})$ is a right Haar measure.

Proof:
If μ is a left Haar measure then:

$$\widetilde{\mu}(Ex) = \mu((Ex)^{-1}) = \mu(x^{-1}E^{-1}) = \mu(E^{-1}) = \widetilde{\mu}(E).$$

Meanwhile if $\widetilde{\mu}$ is a right Haar measure then:

$$\mu(xE) = \mu((E^{-1}x^{-1})^{-1}) = \widetilde{\mu}(E^{-1}x^{-1}) = \widetilde{\mu}(E^{-1}) = \mu((E^{-1})^{-1}) = \mu(E).$$

Also as a side note, you can see that $\widetilde{\mu}$ is a well-defined measure since it is merely the pushforward measure of μ by the inversion map. (Back on *page 193* I was calling this the image measure...)

(b) A nonzero Radon measure μ on G is a left Haar measure iff $\int f d\mu = \int L_y f d\mu$ for all $f \in C_c^+$ and $y \in G$.

$(\Longrightarrow)$
If μ is a left Haar measure then it is obvious that $\int f d\mu = \int L_y f d\mu$ whenever f is a simple function. And by the monotone convergence theorem we can extend this to all $f \in C_c^+$.

$(\Longleftarrow)$
Note that $\operatorname{supp}(L_y f) = y \cdot \operatorname{supp}(f)$ for all $y \in G$ and $f \in C_c(G)$.

After all, $L_y f(x) \neq 0$ iff $f(y^{-1}x) \neq 0$. So if $A = \{x : L_y f(x) \neq 0\}$ and $B = \{x : f(x) \neq 0\}$ then $x \in A$ iff $y^{-1}x \in B$ and that happens iff $x \in yB$. And finally, since translation by y is a homeomorphism on G, we have that:

$$\operatorname{supp}(L_y f) = \overline{A} = \overline{yB} = y\overline{B} = y \cdot \operatorname{supp}(f).$$

Thus, for any open set $U \subseteq G$ we know that if $f \in C_c(G)$ with $\operatorname{supp}(f) \subseteq U$ then $L_y f \in C_c(G)$ with $\operatorname{supp}(L_y f) \subseteq yU$. Similarly, if $f \in C_c(G)$ with $\operatorname{supp}(f) \subseteq yU$ then $L_{y^{-1}} f \in C_c(G)$ with $\operatorname{supp}(L_{y^{-1}} f) \subseteq U$. And when you consider for all open sets $V \subseteq G$ that $\mu(V) = \sup\{\int f d\mu : f \in C_c(G), 0 \leq f \leq 1, \operatorname{supp}(f) \subseteq V\}$, it becomes clear that $\mu(yU) = \mu(U)$ for all $y \in G$ and open sets $U \subseteq G$.

As for the case that E is a general Borel subset of G, we can just approximate E and xE using open sets.

(c) If μ is a left Haar measure on G, then $\mu(U) > 0$ for every nonempty open $U \subseteq G$ and $\int f d\mu > 0$ for all $f \in C_c^+$.

Proof:
Since $\mu \neq 0$, we can show by the regularity properties of μ that there exists a compact set $K \subseteq G$ with $\mu(K) > 0$. Then, for any open nonempty set $U \subseteq G$ we have that K can be covered by finitely many left translates of U. Hence, $\mu(U) > 0$.

Next, if $f \in C_c^+$ then let $U := \{x : f(x) > \frac{1}{2}\|f\|_u\}$. U is open since f is continuous. Therefore, since $\mu(U) > 0$ we have that $\int f d\mu \geq \frac{1}{2}\|f\|_u \mu(U) > 0$.

(d) If μ is a left Haar measure on G then $\mu(G) < \infty$ iff G is compact.

Proof:
The $(\Longleftarrow)$ direction is obvious from the definition of a Radon measure. Meanwhile, suppose G is not compact and let V be a compact neighborhood of e. Then we know that G cannot be covered by finitely many translates of V (lest G be a finite union of compact sets). So, we may find a sequence $\{x_n\}_{n\in\mathbb{N}}$ such that $x_n \in \bigcup_{j=1}^{n-1} x_j V$ for all n.

Next, by proposition 11.1 we can find a symmetric neighborhood U of e with $UU \subseteq V$. Importantly, if $m > n$ and $x_n U \cap x_m U \neq \emptyset$ then we would have that $x_m \in x_n UU \subseteq x_n V$. But that contradicts how we picked our x_n. Hence, we know $\{x_n U\}_{n\in\mathbb{N}}$ is a disjoint sequence of sets. And since $\mu(x_n U) = \mu(U) > 0$, we know that $\mu(G) \geq \mu(\bigcup_{n\in\mathbb{N}} x_n U) = \sum_{n\in\mathbb{N}} \mu(x_n U) = \infty$. ■

I'll continue with actually constructing a Haar measure later on *page 361*.

Math 200a Notes:

If $\sigma \in S_n$ we define $\mathrm{supp}(\sigma) := \{i \in \{1, \ldots, n\} : \sigma(i) \neq i\}$. Note that if we consider the obvious injection $S_n \hookrightarrow S_{n+1} \hookrightarrow \cdots$ then $\mathrm{supp}(\sigma)$ doesn't change. Also, we let $\mathrm{Fix}(\sigma) = \{1, \ldots, n\} - \mathrm{supp}(\sigma)$.

Note that this definition of $\mathrm{Fix}(\sigma)$ is equivalent to the set of fixed points of σ with respect to the obvious group action $S_n \curvearrowright \{1, \ldots, n\}$.

We say $\sigma_1, \sigma_2 \in S_n$ are disjoint if $\mathrm{supp}(\sigma_1) \cap \mathrm{supp}(\sigma_2) = \emptyset$.

Lemma: If σ_1, σ_2 are disjoint then $\sigma_1 \circ \sigma_2 = \sigma_2 \circ \sigma_1$ and $o(\sigma_1 \circ \sigma_2) = \mathrm{lcm}(o(\sigma_1), o(\sigma_2))$.

Proof:
For all $i \in \{1, \ldots, n\}$, if $i \in \mathrm{Fix}(\sigma) \cap \mathrm{Fix}(\sigma_2)$ then $\sigma_1 \circ \sigma_2(i) = i = \sigma_2 \circ \sigma_1(i)$.

Meanwhile, if $i \in \text{Fix}(\sigma_1)$ and $i \notin \text{Fix}(\sigma_2)$, then $\sigma_2 \circ \sigma_1(i) = \sigma_2(i)$. But now note that:

$$\begin{aligned} i \notin \text{Fix}(\sigma_2) \Longrightarrow i \in \text{supp}(\sigma_2) &\Longrightarrow \sigma_2(i) \in \text{supp}(\sigma_2) \\ &\Longrightarrow \sigma_2(i) \notin \text{supp}(\sigma_1) \\ &\Longrightarrow \sigma_2(i) \in \text{Fix}(\sigma_1) \Longrightarrow \sigma_1(\sigma_2(i)) = \sigma_2(i). \end{aligned}$$

So, $\sigma_1 \circ \sigma_2(i) = \sigma_2 \circ \sigma_1(i)$.

The case where $i \notin \text{Fix}(\sigma_1)$ and $i \in \text{Fix}(\sigma_2)$ is similar. And since σ_1 and σ_2 are disjoint, we never have the fourth case. This proves $\sigma_1 \circ \sigma_2 = \sigma_2 \circ \sigma_1$.

Next, let $o(\sigma_i) = d_i$ for both i and let $\ell = \text{lcm}(d_1, d_2)$. Then we know from before that $(\sigma_1 \circ \sigma_2)^\ell = \sigma_1^\ell \sigma_2^\ell = \text{Id}$. Hence, we must have that $o(\sigma_1\sigma_2)$ divides ℓ. On the other hand, suppose $(\sigma_1 \circ \sigma_2)^k = 1$. Then by considering the obvious action $\langle \sigma_2 \rangle \curvearrowright \{1, \ldots, n\}$, we know for all $j \in \text{supp}(\sigma_2)$ that $\langle \sigma_2 \rangle \cdot j \subseteq \text{supp}(\sigma_2)$ and in turn $\langle \sigma_2 \rangle \cdot j \subseteq \text{Fix}(\sigma_1)$. It follows that $j = \sigma_1^k \circ \sigma_2^k(j) = \sigma_2^k(j)$ for all $j \in \text{supp}(\sigma_2)$. Hence $\sigma_2^k = \text{Id}$ and we've shown that $d_2 \mid k$.

By analogous reasoning we can show that $d_1 \mid k$. So, $\ell \mid k$. ■

For the sake of convenience I'm just going to refer to $\sigma_2 \circ \sigma_1$ as $\sigma_2\sigma_1$ from now on.

If $k \geq 2$, we say $\sigma \in S_n$ is a k-cycle if there are distinct elements $a_1, \ldots, a_k \in \{1, \ldots, n\}$ such that $\text{supp}(\sigma) = \{a_1, \ldots, a_k\}$, $\sigma(a_i) = \sigma(a_{i+1})$ for $i < k$, and $\sigma(a_k) = a_1$. We typically denote a k-cycle as $(a_1 \;\; a_2 \;\; \cdots \;\; a_k)$.

A one-cycle is trivial since it is just the identity permutation. So we typically don't count them as "cycles".

Theorem: For all $\sigma \in S_n$ there are disjoint nontrivial cycles $\sigma_1, \ldots, \sigma_m$ such that $\sigma = \sigma_1 \cdots \sigma_m$. Also, this decomposition is unique up to permuting the terms. We call this decomposition a cycle decomposition of σ.

Proof:
Given any permutation $\sigma \in S_n$, we can say that $\langle \sigma \rangle \curvearrowright \{1, \ldots, n\}$ via the action $\sigma \cdot i = \sigma(i)$. Using this fact we can easily show the existence of a cycle decomposition of σ.

Given any $a \in \{1, \ldots, n\}$ we know there is a unique smallest integer k such that $\sigma^k(a) = a$. So, define $\sigma_{(a)} = (a \;\; \sigma(a) \;\; \sigma^2(a) \;\; \cdots \;\; \sigma^{k-1}(a))$. By doing this for all $a \in \{1, \ldots, n\}$, we can see that $\text{supp}(\sigma_{(a)}) = \langle \sigma \rangle \cdot a$ (if a has a non-singleton orbit). Also, it is easy to see that $\text{supp}(\sigma_{(a)}) = \text{supp}(\sigma_{(b)})$ implies that $\sigma_{(a)} = \sigma_{(b)}$. Hence $\Sigma := \{\sigma_{(a)} : a \in \{1, \ldots, n\} \text{ and } \sigma(a) \neq a\}$ is a collection of disjoint cycles. And it is easy to see to that σ equals the product of the elements of Σ in any order.

As for proving the uniqueness of this cycle decomposition, note that if $\sigma = \tau_1\tau_2 \cdots \tau_\ell$ where each τ_i is a disjoint cycle, then $a' \in \text{supp}(\tau_i)$ implies that $\tau_i = \sigma_{(a')}$ from before.

To see this, first note that if $a' \in \text{supp}(\tau_i)$ then we can show from the disjointness of the τ_i that $\sigma^k(a') = \tau_i^k(a')$ for all integers k. But then this alligns with how we defined the permutation $\sigma_{(a')} \in \Sigma$.

This says that each τ_i is equal to some $\sigma_{(a)} \in \Sigma$. Because each of the τ_i is disjoint, we know this correspondance is injective. Also, it's surjective since otherwise the products of the τ_i and the $\sigma_{(a)}$ wouldn't be the common permutation σ. ■

I realize I handwaved this entire proof. But at least I thought about it and wrote down something. Professor Alireza meanwhile skipped proving this.

The cycle type of σ is $(\ell_1 \geq \ell_2 \geq \cdots \geq \ell_m)$ where $\{\ell_1, \ldots, \ell_m\}$ is the set of sizes of the orbits of $\langle\sigma\rangle \curvearrowright \{1, \ldots, n\}$.

Lemma: $o(\sigma) = \operatorname{lcm}(\ell_1, \ldots, \ell_m)$ where $(\ell_1 \geq \cdots \geq \ell_m)$ is the cycle type of σ.

Proof:
Take a cycle decomposition of σ and then inductively apply the lemma at the beginning of this section.

Lemma: Let $a_1, \ldots, a_m, a_{m+1}, \ldots, a_{m+n}$ be distinct elements.

(a) $(a_1 \ \cdots \ a_m)(a_m \ \cdots \ a_{m+n}) = (a_1 \ \cdots \ a_{m+n})$.

(b) For any $\sigma \in S_n$, $\sigma(a_1 \ a_2 \ \cdots \ a_m)\sigma^{-1} = (\sigma(a_1) \ \sigma(a_2) \ \cdots \ \sigma(a_m))$.

Proof:
Showing part (a) is as simple as just going through and showing the permutations on either side of the identity map things to the same places.

To show part (b), recall from page 257 that $\operatorname{Fix}(\sigma\tau\sigma^{-1}) = \sigma \cdot \operatorname{Fix}(\tau)$. (where $A \subseteq \{1, \ldots, n\}$ implies that $\sigma \cdot A := \{\sigma(a) : a \in A\}$). In turn:

$$\operatorname{supp}(\sigma\tau\sigma^{-1}) = \sigma \cdot \operatorname{supp}(\tau)$$

And in particular, $\operatorname{supp}(\sigma(a_1 \ \cdots \ a_m)\sigma^{-1}) = \{\sigma(a_1), \ldots, \sigma(a_m)\}$. So we know that the permutations on both sides of our proposed equation have the same support.

Next note for any $k < m$ that $\sigma(a_1 \ \ldots \ a_m)\sigma^{-1}(\sigma(a_k)) = \sigma(a_{k+1})$ (and similarly plugging in $\sigma(a_m)$ gives $\sigma(a_1)$). So the permutations agree on their supports and thus everywhere. ■

Proposition: $\sigma_1, \sigma_2 \in S_n$ are conjugates if and only if they have the same cycle types.

$(\Longrightarrow)$
Suppose $\sigma = \tau_1 \cdots \tau_m$ is a cycle decomposition such that τ_i has cycle length ℓ_i and $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_m$. Then the cycle type of σ_1 is $(\ell_1 \geq \cdots \geq \ell_m \geq 1 \geq \cdots \geq 1)$ (with $n = \sum_{i=1}^{m} \ell_i$ many 1s at the end). Also, if $\sigma_2 = \sigma\sigma_1\sigma^{-1}$ then:

$$\sigma_2 = \sigma\tau_1 \cdots \tau_m\sigma^{-1} = (\sigma\tau_1\sigma^{-1})(\sigma\tau_2\sigma^{-1}) \cdots (\sigma\tau_m\sigma^{-1})$$

Additionally, it's not hard to see that each $(\sigma\tau_i\sigma^{-1})$ is a disjoint cycle of the same length as τ_i. Hence $(\sigma\tau_1\sigma^{-1})(\sigma\tau_2\sigma^{-1}) \cdots (\sigma\tau_m\sigma^{-1})$ is a cycle decomposition of σ_2 and the order type of σ_2 is also $(\ell_1 \geq \cdots \geq \ell_m \geq 1 \cdots \geq 1)$.

$(\Longleftarrow)$
Suppose the common cycle type is $(\ell_1 \geq \ell_2 \geq \cdots \geq \ell_m)$. Then there exists $n < m$ and cycle decompositions:

- $\sigma_1 = (a_1^{(1)} \ \cdots \ a_{\ell_1}^{(1)})(a_1^{(2)} \ \cdots \ a_{\ell_2}^{(2)}) \cdots (a_1^{(n)} \ \cdots \ a_{\ell_n}^{(n)})$
- $\sigma_2 = (b_1^{(1)} \ \cdots \ b_{\ell_1}^{(1)})(b_1^{(2)} \ \cdots \ b_{\ell_2}^{(2)}) \cdots (b_1^{(n)} \ \cdots \ b_{\ell_n}^{(n)})$

(Note all $a_i^{(j)}$ are distinct and similarly all $b_i^{(j)}$ are distinct.)

For every i such that $\ell_i = 1$ we pick a different $a_1^{(i)}$ among the fixed points of σ_1. Similarly, we pick a different $b_1^{(i)}$ among the fixed points of σ_2 for each i with $\ell_i = 1$. Then finally, we define the permutation $\tau \in S_n$ by $\tau(a_j^{(i)}) = b_j^{(i)}$. Then:

$$\begin{aligned}\tau\sigma_1\tau^{-1} &= (\tau(a_1^{(1)}) \ \cdots \ \tau(a_{\ell_1}^{(1)})) \cdots (\tau(a_1^{(m)}) \ \cdots \ \tau(a_{\ell_m}^{(m)})) \\ &= (b_1^{(1)} \ \cdots \ b_{\ell_1}^{(1)}) \cdots (b_1^{(m)} \ \cdots \ b_{\ell_m}^{(m)}) = \sigma_2. \ \blacksquare\end{aligned}$$

Corollary: The number of conjugate classes of S_n is equal to the number of integer partitions of n (see my paper math 188 notes).

By noting that $(a_1 \ a_2 \ \cdots \ a_n) = (a_1 \ a_2)(a_2 \ a_3) \cdots (a_{n-1} \ a_n)$ we can see that every permutation can be written as a product of transpositions (i.e. 2-cycles). That said, except in trivial cases there are many different ways to write a permutation as a product of transpositions. For example, $(1 \ 2)(2 \ 3)(1 \ 2) = (1 \ 3)$. So, is it still possible to characterize permutations somehow by how they are expressed as products of transpositions?

Note that $S_n \curvearrowright \mathbb{Z}[x_1, \ldots, x_n]$ by $(\sigma \cdot f)(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. Also, this group action importantly has the properties that $\sigma \cdot (f + g) = (\sigma \cdot f) + (\sigma \cdot g)$ and $\sigma \cdot (fg) = (\sigma \cdot f)(\sigma \cdot g)$.

If we let $\Delta(x_1, \ldots, x_n) \coloneqq \prod_{i<j}(x_i - x_j)$, then we we have that

$$\Delta^2(x_1, \ldots, x_n) = (-1)^{\frac{n(n-1)}{2}} \prod_{i \neq j}(x_i - x_j).$$

Hence, $(\sigma \cdot \Delta^2) = \Delta^2$ for all $\sigma \in S_n$ (i.e. Δ^2 is a symmetric polynomial). And since $(\sigma \cdot \Delta)^2 = (\sigma \cdot \Delta^2) = \Delta^2$, we must have that $\sigma \cdot \Delta = \pm\Delta$ for each $\sigma \in S_n$. So, there exists a map $\varepsilon : S_n \to \{-1, 1\}$ such that $\sigma \cdot \Delta = \varepsilon(\sigma)\Delta$.

Note that $\sigma \cdot (\sigma' \cdot \Delta) = (\sigma\sigma') \cdot \Delta = \varepsilon(\sigma\sigma')\Delta$. But we also have that:

$$\sigma \cdot (\sigma' \cdot \Delta) = \sigma \cdot (\varepsilon(\sigma')\Delta) = \varepsilon(\sigma')(\sigma \cdot \Delta) = \varepsilon(\sigma')\varepsilon(\sigma)\Delta.$$

Therefore $\varepsilon(\sigma\sigma') = \varepsilon(\sigma)\varepsilon(\sigma')$. And when you also consider that $\varepsilon(\mathrm{Id}) = +1$, this proves that ε is a group homomorphism.

One more note is that if τ is a transposition then $\varepsilon(\tau) = -1$. To see this, just note that if $\tau = (i \ \ j)$ where $i < j$, then $\tau \cdot \Delta = (-1)^{2(j-i-1)+1}\Delta = -\Delta$.

Thus, if $\sigma \in S_n$ is decomposed into a product of transpositions, we must have that the number of transpositions in that product is even if $\varepsilon(\sigma) = +1$ and odd if $\varepsilon(\sigma) = -1$. In other words, even and odd permutations are well-defined.

We define $A_n := \ker(\varepsilon) = \{\text{all even permutations}\}$. Note that $A_n \lhd S_n$ (since it is a kernel of a group homomorphism).

Theorem: If $n \geq 2$ then $[S_n : A_n] = 2$.

Proof:

By the first isomorphism theorem we have that:

$$S_n/A_n = S_n/\ker(\varepsilon) \cong \operatorname{im}(\varepsilon) = \{-1, +1\}.$$

Hence $[S_n : A_n] = 2$. ■

The last theorem that can be tested on the midterm tomorrow is that A_n is a simple group (if $n \geq 5$). So I'll try to prove this and then do some exercises.

Lemma: A_n is generated by the set of 3-cycles (if $n \geq 3$).

Proof:

It is enough to show that the product of 2 transpositions is in the subgroup generated by 3 cycles. Fortunately, if a, b, c, d are distinct elements in $\{1, \ldots, n\}$ then:

- $(a\ \ b)(a\ \ b) = \mathrm{Id}$
- $(a\ \ b)(b\ \ c) = (a\ \ b\ \ c)$
- $(a\ \ b)(c\ \ d) = (a\ \ b)(b\ \ c)(b\ \ c)(c\ \ d) = (a\ \ b\ \ c)(b\ \ c\ \ d)$. ■

Lemma: If $N \lhd A_n$ and N has a 3-cycle then $N = A_n$.

Proof:

Suppose $(a_1\ \ a_2\ \ a_3) \in N$. Then for any distinct $b_1, b_2, b_3 \in \{1, \ldots, n\}$ let $\sigma \in S_n$ be a permutation such that $\sigma(a_1) = b_1$, $\sigma(a_2) = b_2$, and $\sigma(a_3) = b_3$. Since N is normal, we have that $(b_1\ \ b_2\ \ b_3) = \sigma(a_1\ \ a_2\ \ a_3)\sigma^{-1} \in N$. So, N contains all 3-cycles. And by the last lemma this means that $A_n < N$. ■

Theorem: If $n \geq 5$ then A_n is simple.

Proof:

We shall first show that A_5 is simple. Note that $|A_5| = \frac{5!}{2} = 60 = 5 \cdot 2^2 \cdot 3$. So suppose for the sake of contradiction that there exists a subgroup $N \lhd A_5$ with $1 \lneqq N \lneqq A_5$.

We can't have that 3 divides $|N|$.

If 3 divides $|N|$ then we would know by Cauchy's theorem that N contains an element of order 3, say σ. So, let $(\ell_1 \geq \cdots \geq \ell_m)$ be the cycle type of σ. We must have that $\ell_1 + \ldots + \ell_m = 5$ and $\operatorname{lcm}(\ell_1, \ldots, \ell_m) = 3$. But the only cycle type satisfying those requirements is $(3 \geq 1 \geq 1)$. Hence, σ is a 3-cycle. That in turn implies by the last lemma that $N = A_n$ (which is a contradiction).

Similarly, we can't have that 5 divides $|N|$.

Let $P \in \mathrm{Syl}_5(N)$. If 5 divides N then we also have that $P \in \mathrm{Syl}_5(A_5)$. And since there exists $x \in A_5$ such that $P' = xPx^{-1} \subseteq N$ for all $P' \in \mathrm{Syl}_p(A_5)$, we can in turn say that N contains every element of A_n with order dividing 5.

But now note that the only elements of S_5 with order 5 are 5 cycles. Also, all 5 cycles are easily seen to be even permutations. So, $\{\sigma \in S_5 : \sigma$ is a 5-cycle$\}$ is a subset of N. Also, it is an easy counting exercise to see that the number of such permutations is $^{5!}/_{5} = 24$. Hence, we've proven that $24 < |N|$.

Since $|N|$ divides 60 and $N \subsetneqq A_5$, this forces $|N| = 30$. But we already showed that 3 can't divide $|N|$. So, we have a contradiction.

This narrows us down to the case that $|N| = 2$ or $|N| = 4$. To address this case, we bring up the following lemma:

<u>Lemma:</u> If $|G| < \infty$, $N \lhd G$, and N is a p-group, then $N \subseteq \bigcap_{P \in \mathrm{Syl}_p(G)} P =: O_p(G)$.

Proof:

By Sylow's 2nd theorem we know $N \subseteq P_0$ for some $P_0 \in \mathrm{Syl}_p(G)$. Then since N is normal, $N = \bigcap_{x \in G} xNx^{-1} \subseteq \bigcap_{x \in G} xP_0x^{-1} = \bigcap_{P \in \mathrm{Syl}_p(G)} P$. ■

Now pick $\sigma \in A_5$ with $o(\sigma) = 2$. The only cycle types of permutations in S_5 that yield permutations of order 2 are $(2 \geq 1 \geq 1 \geq 1)$ and $(2 \geq 2 \geq 1)$. However, permutations of the former cycle type are odd. Hence, we may assume $\sigma = (a\ \ b)(c\ \ d)$.

Next, note that $P_2 := \{\mathrm{Id}, (a\ \ b)(c\ \ d), (a\ \ c)(b\ \ d), (a\ \ d)(b\ \ c)\}$ is a Sylow 2-subgroup of A_5.

It is clear that P_2 is closed under inverses. We only need to verify that it really is closed under compositions.

- $(a\ \ b)(c\ \ d)(a\ \ c)(b\ \ d) = (a\ \ d)(b\ \ c)$ and $(a\ \ b)(c\ \ d)(a\ \ d)(b\ \ c) = (a\ \ c)(b\ \ d)$,
- $(a\ \ c)(b\ \ d)(a\ \ b)(c\ \ d) = (a\ \ d)(b\ \ c)$ and $(a\ \ c)(b\ \ d)(a\ \ d)(b\ \ c) = (a\ \ b)(c\ \ d)$,
- I'm bored and don't want to manually verify the last two relations.

Thus by our lemma, we know that $N \subseteq \bigcap_{P \in \mathrm{Syl}_2(G)} P = \bigcap_{x \in G} xP_2x^{-1}$.

But now note that:

$$((a\ \ b)(c\ \ e))P_2((a\ \ b)(c\ \ e))^{-1} = \{\mathrm{Id}, (b\ \ a)(e\ \ d), (b\ \ e)(a\ \ d), (b\ \ d)(a\ \ e)\}$$

Therefore $\bigcap_{x \in G} xP_2x^{-1}$ is trivial and we have a contradiction since $N = \{\mathrm{Id}\}$. This finishes the proof that A_5 is simple.

Now we proceed by induction on n. Suppose $1 \neq N \lhd A_n$ and for all $i \in \{1, \ldots, n\}$ let $G_i := \{\sigma \in A_n : \sigma(i) = i\}$. Then there is an obvious group isomorphism such that $G_i \cong A_{n-1}$ for all i. Also, $N \cap G_i \lhd G_i$. So, by induction we know for each i that either $N \cap G_i = \{1\}$ or $N \cap G_i = G_i$.

But if $N \cap G_i = G_i$ for any i then we are already done since that would imply that N contains a 3-cycle. So, without loss of generality we may now assume that $N \cap G_i = \{\mathrm{Id}\}$ for all i. As a consequence, if $\sigma, \sigma' \in N$ satisfy that $\sigma(i) = \sigma'(i)$ for any i then we must have that $\sigma = \sigma'$ since $\sigma(\sigma')^{-1} \in N \cap G_i$.

Suppose $\sigma \in N - \{\mathrm{Id}\}$. Then we have two cases:

- Suppose σ has a cycle of size ≥ 3. In other words, there exists distinct numbers $a, b, c \in \{1, \ldots, n\}$ such that $\sigma = (a \ b \ c \ \cdots)\cdots$. But now if $d, e \in \{1, \ldots, n\}$ are any other 2 distinct numbers, we have that $\sigma' := (c \ d \ e)\sigma(c \ d \ e)^{-1} \in N$ with $\sigma'(a) = b = \sigma(a)$ and $\sigma'(b) = d \neq c = \sigma(b)$. This is a contradiction.
- Meanwhile, suppose σ has only cycles of length 2. Since σ is even, we thus know there are distinct numbers $a, b, c, d \in \{1, \ldots, n\}$ such that $\sigma = (a \ b)(c \ d)\cdots$. But now if e, f are two other distinct elements of $\{1, \ldots, n\}$ (supposing $n \geq 6$), then consider $\sigma' := (c \ e \ f)\sigma(c \ e \ f)^{-1}$. Like before, $\sigma'(a) = b = \sigma(a)$. But $\sigma'(d) = e \neq c = \sigma(d)$. Hence we again have a contradiction. ■

Here is an application of the prior theorem. Suppose G is a finite group with $|G| = 2m$ where $2 \not| m$. Then there exists a characteristic subgroup $N < G$ such that $[G : N] = 2$.

Proof:
Consider the action of $G \curvearrowright G$ by left translations and let $\phi : G \to S_G$ be the induced group homomorphism. Note that we can always identify S_G with $S_{|G|}$ by just numbering the elements of G. Also note that for all $g \in G$ the cycle type of $\phi(g)$ is:

$(o(g) \geq o(g) \geq \cdots \geq o(g))$ (where there are $|G|/o(g)$ many orbits).

Next, by Cauchy's theorem there exists $g_0 \in G$ such that $o(g_0) = 2$. But now the cycle type of $\phi(g_0)$ is $(2 \geq 2 \geq \cdots \geq 2)$ (with $|G|/2 = m$ many orbits.) Hence, $\phi(g_0)$ is an odd permutation and we know that $\varepsilon \circ \phi : G \to \{\pm 1\}$ is a surjective group homomorphism.

Let $N := \ker(\varepsilon \circ \phi)$. Then $[G : N] = |\mathrm{im}(\varepsilon \circ \phi)| = 2$. Also, because of how we defined N we know that $g \in N$ if and only if the cycle type of $(o(g) \geq o(g) \geq \cdots \geq o(g))$ gives an even permutation.

But now note that for all $\theta \in \mathrm{Aut}(G)$ we have that $o(\theta(g)) = o(g)$. Thus $g \in N$ iff $\theta(g) \in N$ for all $\theta \in \mathrm{Aut}(G)$. And this implies that N is a characteristic subgroup of G.
■

(I didn't do these problems before they were due but I'm doing them now...)

Set 4 Problem 5: In this problem we show that $\mathrm{Inn}(S_6) \neq \mathrm{Aut}(S_6)$.

(a) Show that S_5 has 6 Sylow 5-subgroups and then use the action of S_5 on $\mathrm{Syl}_5(S_5)$ to show that S_6 has a subgroup H which is isomorphic to S_5.

Since $s_5 := |\mathrm{Syl}_5(S_5)|$ equals $1 \pmod 5$ and divides $\frac{|S_5|}{5} = \frac{5!}{5} = 24$, that already restricts s_5 to equaling either 1 or 6. That said, as I will prove later on *page 377* in my lecture notes, S_5 does not have a normal subgroup of size 5. Therefore, this forces $s_5 = 6$.

Now consider the action $S_5 \curvearrowright \mathrm{Syl}_5(S_5)$ by conjugation and let:

$\phi : S_5 \to S_{\mathrm{Syl}_5(S_5)} \cong S_6$ be the induced homomorphism.

Note that $\sigma \in \ker(\phi)$ iff $\sigma P \sigma^{-1} = P$ for all $P \in \mathrm{Syl}_5(G)$. In particular, if we fix $P \in \mathrm{Syl}_5(G)$ then we know that $\ker(\phi) \subseteq N_G(P)$ and hence:

$$|\ker(\phi)| \leq \frac{S_5!}{s_5} = \frac{120}{6} = 20.$$

Thus $\ker(\phi) \lhd S_5$ and $[S_5 : \ker(\phi)] > 2$. By the aforementioned proof on *page 377*, this means that $\ker(\phi) = \{\mathrm{Id}\}$. Hence ϕ is an injective homomorphism. And by letting $H \coloneqq \mathrm{im}(\phi)$ we have that $S_5 \cong H < S_6$.

(b) Show for every $\sigma \in S_6$ that $\mathrm{Fix}(\sigma H \sigma^{-1}) = \emptyset$ (where S_6 is acting on $\{1, \ldots, 6\}$ by evaluation).

We know that $\mathrm{Fix}(\sigma H \sigma^{-1}) = \bigcap_{\tau \in H} \mathrm{Fix}(\sigma \tau \sigma^{-1}) = \bigcap_{\tau \in H}(\sigma \cdot \mathrm{Fix}(\tau))$. But now recall that the action $S_5 \curvearrowright \mathrm{Syl}_5(S_5)$ is transitive and hence there is some $\tau' \in H$ such that $\mathrm{Fix}(\tau') = \emptyset$. In turn, $\sigma \cdot \mathrm{Fix}(\tau') = \emptyset$ and so $\mathrm{Fix}(\sigma H \sigma^{-1}) = \emptyset$ for all $\sigma \in S_n$.

Another way of thinking of this is that conjugation preserves cycle type. So if $\tau \in S_6$ has no fixed points (i.e. 1-cycles), then $\sigma \tau \sigma^{-1}$ also has no 1-cycles.

(c) Consider the action $S_6 \curvearrowright S_6/H$ by left-translation. Show that this induces a group homomorphism $\theta : S_6 \to S_6$ and that $\mathrm{Fix}(\theta(H)) \neq \emptyset$ (again with respect to the action of S_6 on $\{1, \ldots, 6\}$ by evaluation).

Since $|H| = |S_5|$, we know $[S_6 : H] = \frac{|S_6|}{|S_5|} = \frac{6!}{5!} = 6$. Hence, by numbering the cosets of H we can say that $S_{S_6/H} \cong S_6$. For convenience, I'll assume the coset H corresponds to $1 \in \{1, \ldots, 6\}$.

Now consider the induced homomorphism θ described in the problem statement. For any $\tau \in S_6$ we have that $\theta(\tau)$ describes the map $\sigma H \mapsto \tau \sigma H$. But note that if $\tau \in H$ then $\theta(\tau)$ fixes H. After applying our correspondance $S_{S_6/H} \cong S_6$ this translates to saying that $1 \in \mathrm{Fix}(\theta(H))$.

(d) Deduce that $\mathrm{Aut}(S_6) \neq \mathrm{Inn}(S_6)$.

In part (b) we proves that if $\psi \in \mathrm{Inn}(S_6)$ then $\mathrm{Fix}(\psi(H)) = \emptyset$. Thus the fact that θ from part (c) doesn't satisfy that property means θ is definitely not an inner automorphism. If we can now prove that θ is in fact an automorphism despite that, then we will be done.

Fortunately, note that $\tau \in \ker(\theta)$ iff $\tau \sigma H = \sigma H$ for all $\sigma \in S_6$. This is equivalent to saying that $\tau \in \bigcap_{\sigma \in S_6} \sigma H \sigma^{-1}$. And in particular, this proves that $\ker(\theta) \subseteq H$. But now since $\ker(\theta) \lhd S_6$ and $[S_6 : \ker(\theta)] > [S_6 : H] > 2$, we know from one more application of the fact on *page 377* that $\ker(\theta) = \{\mathrm{Id}\}$. Hence θ is injective. And by pigeonhole principle this also proves that θ is surjective. ■

Set 4 Problem 6: Prove that a group G of order 36 is not simple.

Suppose for the sake of contradiction that G is simple. Then since $s_3 := |\mathrm{Syl}_3(G)|$ divides 12 and equals $1 \pmod 3$, we know s_3 must equal either 1 or 4. But we can't have that $s_3 = 1$ since that would violate the simplicity of G. Hence, we know that $s_3 = 4$.

Next, consider the action $G \curvearrowright \mathrm{Syl}_3(G)$ and let $\phi : G \to S_{\mathrm{Syl}_3(G)} \cong S_4$ be the induced homomorphism. Since that action is transitive, we know that $\ker(\phi) \neq G$. However, we also know that $\ker(\phi) \lhd G$. So by the simplicity of G we must have that $\ker(\phi) = 1$. And this implies by the first isomorphism theorem that there is some subgroup $H = \mathrm{im}(\phi)$ of S_4 with $G \cong H$. But this is a contradiction since $|G| = 36 > 24 = |S_4|$. ■

10/29/2025

Ehh the midterm went mediocrely. I guess I'm one step closer towards flunking out of school. Anyways, right now I want to go back to studying Haar measures. So I'm going to resume what I was doing on *page 353*.

If G is a group and $E, V \subseteq G$, then we can in some sense measure the size of E relative to V by asking what is the minimum cardinality of $A \subseteq G$ such that $E \subseteq \bigcup_{x \in A} xV$. Also, if E is a compact or precompact set and V is open, then we can guarentee that this minimum cardinality is finite.

Clearly, the above construction defines a translation invariant notion of size. But it's not a measure. Can we modify this approach to actually get a measure?

Firstly, we'll switch to working with functions in $C_c^+(G)$ since functions are easier to work with than measures and we can hopefully apply the Riesz representation theorem if we get a positive result when working with functions.

Suppose $f, \phi \in C_c^+(G)$. Then $U = \{x : \phi(x) > \frac{1}{2}\|\phi\|_u\}$ is an open nonempty set. Also, $\mathrm{supp}(f)$ is compact. So there are finitely many $x_1, \ldots, x_n \in G$ with $\mathrm{supp}(f) \subseteq \bigcup_{j=1}^n x_j U$. And we in turn know that:

$$f \leq \frac{2\|f\|_u}{\|\phi\|_u} \sum_{j=1}^n L_{x_j}\phi.$$

Hence, given any $f, \phi \in C_c^+(G)$ it is well-defined to set:

$$(f : \phi) := \inf\left\{\sum_{j=1}^n c_j : f \leq \sum_{j=1}^n c_j L_{x_j}\phi \text{ for some } n \in \mathbb{N} \text{ and } x_1, \ldots, x_n \in G\right\}.$$

Note that if $f \leq \sum_{j=1}^n c_j L_{x_j}\phi$ then we have by triangle inequality that $\|f\|_u \leq \|\phi\|_u \sum_{j=1}^n c_j$. Thus, we clearly have that $0 < \frac{\|f\|_u}{\|\phi\|_u} \leq (f : \phi)$.

<u>Lemma 11.5:</u> Suppose that $f, g, \phi \in C_c^+$. Then:

(a) $(f : \phi) = (L_x f : \phi)$ for any $x \in G$.

This is because $f \leq \sum_{j=1}^n c_j L_{x_j}\phi$ iff $L_x f \leq \sum_{j=1}^n c_j L_{xx_j}\phi$.

(b) $(cf : \phi) = c(f : \phi)$ for any $c > 0$.

This is because $f \leq \sum_{j=1}^{n} c_j L_{x_j}\phi$ iff $cf \leq \sum_{j=1}^{n} cc_j L_{x_j}\phi$.

(c) $(f + g, \phi) \leq (f : \phi) + (g : \phi)$.

Suppose $f \leq \sum_{j=1}^{m} c_j L_{x_j}\phi$ and $g \leq \sum_{j=m+1}^{m+n} c_j L_{x_j}\phi$. Then $f + g \leq \sum_{j=1}^{n} c_j L_{x_j}\phi$. And by minimizing $\sum_{j=1}^{m} c_j$ and $\sum_{j=m+1}^{m+n} c_j$ this claim follows.

(d) $(f : \phi) \leq (f : g)(g : \phi)$.

If $f \leq \sum_{j=1}^{n} c_j L_{x_j} g$ and $g \leq \sum_{i=1}^{n} d_i L_{y_i}\phi$, then:

$$f \leq \sum_j c_j L_{x_j}(\sum_{i=1} d_i L_{y_i}\phi) = \sum_j c_j(\sum_i d_i L_{x_j} L_{y_i}\phi) = \sum_{j,i} c_j d_i L_{x_j y_i}\phi$$

And since $\sum_{j,i} c_j d_i = (\sum_j c_j)(\sum_i d_i)$ our claim follows. ■

Now let us fix $f_0 \in C_c^+(G)$ from here on out. It doesn't matter which function f_0 specifically is. We just want to have a base-line function to compare all other functions in $C_c^+(G)$ to.

Next, if $\phi \in C_c^+(G)$ then let us define $I_\phi(f) := \frac{(f:\phi)}{(f_0:\phi)}$ for all $f \in C_c^+(G)$.

By the last lemma, we know that I_ϕ is a sublinear functional that is invariant to left translations. Also, note that:

$$I_\phi(f) = \frac{(f:\phi)}{(f_0:\phi)} \leq \frac{(f:f_0)(f_0:\phi)}{(f_0:\phi)} \leq (f : f_0) \text{ and } I_\phi(f) = \frac{(f:\phi)}{(f_0:\phi)} \geq \frac{(f:\phi)}{(f_0:f)(f:\phi)} \geq (f_0 : f)^{-1}.$$

(To put the last line more succinctly, $I_\phi(f) \in [(f_0 : f)^{-1}, (f : f_0)]$ for all $f, \phi \in C_c^+(G)$...)

Folland's next claim is that I_ϕ is "approximately" linear when $\mathrm{supp}(\phi)$ is small.

Lemma 11.7: If $f_1, f_2 \in C_c^+(G)$ and $\varepsilon > 0$ then there is a neighborhood V of $e \in G$ such that $I_\phi(f_1) + I_\phi(f_2) \leq I_\phi(f_1 + f_2) + \varepsilon$ whenever $\mathrm{supp}(\phi) \subseteq V$.

Proof:

Fix $g \in C_c^+$ with $g = 1$ on $\mathrm{supp}(f_1 + f_2)$. Then for any $\delta > 0$ define $h_\delta = f_1 + f_2 + \delta g$ and for both i define:

$$h_\delta^{(i)}(x) = \begin{cases} f_i(x)/h_\delta(x) & \text{if } x \in \mathrm{supp}(f) \\ 0 & \text{otherwise.} \end{cases}$$

Then each $h_\delta^{(i)} \in C_c^+(G)$. So by *proposition 11.2* we know there is some neighborhood V_δ of e with $|h_\delta^{(i)}i(x) - h_\delta^{(i)}(y)| < \delta$ when $y^{-1}x \in V$ and $i \in \{1, 2\}$.

Specifically, let V_δ be a nighborhood of e such that $\|R_z h_\delta^{(i)} - h_\delta^{(i)}\|_u < \delta$ for all $z \in V_\delta$. Then $|h_\delta^{(i)}(x) - h_\delta^{(i)}(y)| < \delta$ if $x = yz$ for some $z \in V_\delta$. Or in other words, if $y^{-1}x = z \in V_\delta$ then $|h_\delta^{(i)}(x) - h_\delta^{(i)}(y)| < \delta$.

If $\phi \in C_c^+(G)$ with $\mathrm{supp}(\phi) \subseteq V_\delta$ and $h_\delta \leq \sum_{j=1}^{n} c_j L_{x_j}\phi$, then $|h_\delta^{(i)}(x) - h_\delta^{(i)}(x_j)| < \delta$ whenever $x_j^{-1}x \in \mathrm{supp}(\phi)$. Hence, we can say for all $x \in G$ that:

$$f_i(x) = h_\delta(x)h_\delta^{(i)} \leq \sum_{j=1}^{n} c_j\phi(x_j^{-1}x)h_\delta^{(i)}(x) \leq \sum_{j=1}^{n} c_j\phi(x_j^{-1}x)(h_\delta^{(i)}(x_j) + \delta)$$

But now this proves that $(f_i : \phi) \leq \sum_{j=1}^n c_j(h_\delta^{(i)}(x_j) + \delta)$ for both i.

Also, since $h_\delta^{(1)} + h_\delta^{(2)} \leq 1$, we can thus conclude that:

$$(f_1 : \phi) + (f_2 : \phi) \leq \sum_{j=1}^n c_j(1 + 2\delta).$$

And by bringing $\sum_{j=1}^n c_j$ arbitrarily close to $(h : \phi)$ and dividing by $(f_0 : \phi)$, we can now say that $I_\phi(f_1) + I_\phi(f_2) \leq (1 + 2\delta)I_\phi(h_\delta) \leq (1 + 2\delta)(I_\phi(f_1 + f_2) + \delta I_\phi(g))$.

Now we want $(1 + 2\delta)(I_\phi(f_1 + f_2) + \delta I_\phi(g)) < I_\phi(f_1 + f_2) + \varepsilon$. Equivalently this means we want $2\delta I_\phi(f_1 + f_2) + \delta(1 + 2\delta)I_\phi(g) < \varepsilon$. And fortunately, this will be guarenteed if:

$$2\delta(f_1 + f_2 : f_0) + \delta(1 + 2\delta)(g : f_0) < \varepsilon$$

So, by choosing δ small enough we can guarentee that $I_\phi(f_1) + I_\phi(f_2) \leq I_\phi(f_1 + f_2)$ whenever $\mathrm{supp}(\phi) \subseteq V_\delta$. ■

Theorem 11.8: Every locally compact group G contains a left Haar measure.

Proof:

For each $f \in C_C^+(G)$ let $X_f := [(f_0 : f)^{-1}, (f : f_0)]$. Then let $X = \prod_{f \in C_c^+(G)} X_f$. By Tychonoff's theorem we know that X is a compact Hausdorff space. Also, we have that $I_\phi \in X$ for all $\phi \in C_c^+(G)$ since $I_\phi(f) \in [(f_0 : f)^{-1}, (f : f_0)]$ for all $f, \phi \in C_c^+(G)$.

Now for each neighborhood V of e let $K(V)$ be the closure in X of $\{I_\phi : \mathrm{supp}(\phi) \subseteq V\}$. Then note that $\bigcap_{j=1}^n K(V_j) \supseteq K(\bigcap_{j=1}^n V_j) \neq \emptyset$ for all finite collections $\{V_1, \ldots, V_n\}$ of neighborhoods of e. Hence since X is compact and the collection of sets $K(V)$ has the finite intersection property, we know there exists an element I in the intersection of all the $K(V)$'s.

Next since I is either an accumulation point of or inside $\{I_\phi : \mathrm{supp}(\phi) \subseteq V\}$ for all neighborhoods V of e, we know that any neighborhood of I in X must intersect $\{I_\phi : \mathrm{supp}(\phi) \subseteq V\}$ for all neighborhoods V of e in G. Consequently, for any neighborhood V of e and any $f_1, \ldots, f_n \in C_c^+(G)$ and $\varepsilon > 0$ there exists $\phi \in C_c^+(G)$ with $\mathrm{supp}(\phi) \subseteq V$ such that $|I(f_j) - I_\phi(f_j)| < \varepsilon$ for each j.

Why?

By definition of the product topology, the following is an open neighborhood of I in X:

$$U := \{I' \in X : |I(f_j) - I'(f_j)| < \varepsilon \text{ for } j = 1, \ldots, n\}$$

Then any $I_\phi \in U \cap \{I_\phi : \mathrm{supp}(\phi) \subseteq V\}$ satisfies that $|I(f_j) - I_\phi(f_j)| < \varepsilon$ for each $j \in \{1, \ldots, n\}$.

Consequently, we can now show using lemmas 11.5 and 11.7 that I is left-invariant and satisfies that $I(af + bg) = aI(f) + bI(g)$ for all $f, g \in C_c^+$ and $a, b > 0$.

To show that $I(af + bg) = aI(f) + bI(g)$, consider any $\varepsilon > 0$ and then pick a neighborhood V of e such that whenever $\mathrm{supp}(\phi) \subseteq V$ we have that:

$$I_\phi(af + bg) \leq aI_\phi(f) + bI_\phi(g) \leq I_\phi(af + bg) + \varepsilon$$

Then by our prior reasoning there exists $\phi \in C_c^+(G)$ with $\mathrm{supp}(\phi) \subseteq V$ such that $|I(f) - I_\phi(f)| < \varepsilon$, $|I(g) - I_\phi(g)| < \varepsilon$, and $|I(af + bg) - I_\phi(af + bg)| < \varepsilon$. And hence we can get that $|I(af + bg) - aI(f) - bI(g)| < 4\varepsilon$.

By taking $\varepsilon \to 0$ this then proves that $I(af + bg) = aI(f) + bI(g)$.

Similarly, to prove that I is left-invariant let $\varepsilon > 0$ and pick ϕ such that $|I(f) - I_\phi(f)| < \varepsilon$ and $|I(L_x f) - I_\phi(L_x f)| < \varepsilon$. Then $|I(f) - I(L_x f)| < 2\varepsilon$.

We can extend I to all of $C_c(G, [0, \infty))$ by defining $I(0) = 0$. This importantly preserves the linearity and left-invariance of I. Then similarly to the proof of lemma 7.15 on *pages 57-58*, by setting $I(f) = I(f^+) - I(f^-)$ where f^+ and f^- are the positive and negative parts of f we can extend I to being a positive real linear functional on $C_c(G, \mathbb{R})$. And this extension is clearly still left-invariant since $(L_x f)^+ = L_x f^+$ and $(L_x f)^- = L_x f^-$.

Finally, we extend I to being a positive left-invariant linear functional on all of $C_c(G)$ by just setting $I(f) = I(\mathrm{Re}(f)) + iI(\mathrm{Im}(f))$. Then by applying the Riesz-representation theorem plus *proposition 11.4(b)* we get a left-invariant measure on G. And I is not the zero functional on G since $I(f) \geq (f_0 : f)^{-1} > 0$ when $f \in C_c^+(G)$. ■

I will go into more depth on Haar measures later on *page 430*.

10/30/2025

Math 220a Notes:

We'll now introduce a notion of symmetric points with respect to a given circle $\Gamma \subseteq \mathbb{C}_\infty$. Firstly, if $\Gamma = \mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}$, then clearly the intuitive definition of a symmetric point of $z \in \mathbb{C}$ would be the point $z^* := \overline{z} \in \mathbb{C}$.

Proposition: If S is a Möbius transformation and $S(\mathbb{R}) = \mathbb{R}$, then $S(z^*) = (S(z))^*$ for all $z \in \mathbb{C}$ with $S(z) \neq \infty$.

Proof:

S is uniquely determined by the points $z_1, z_2, z_3 \in \mathbb{C}_\infty$ such that $S(z_1) = 1$, $S(z_2) = 0$, and $S(z_3) = \infty$. But note that since $S(\mathbb{R}) = \mathbb{R}$ and S maps circles to circles, we know that $z_1, z_2, z_3 \in \mathbb{R}_\infty$. Hence, by our construction in the existence proof at the top of *page 337*, we know that there are real $a, b, c, d \in \mathbb{R}$ such that $S(z) = \frac{az+b}{cz+d}$. And now if $z \in \mathbb{C} - \{z_3\}$, we have that:

$$S(z)^* = \overline{S(z)} = \overline{\left(\frac{az+b}{cz+d}\right)} = \frac{a\overline{z}+b}{c\overline{z}+d} = S(\overline{z}) = S(z^*). \blacksquare$$

Next we consider letting $\Gamma \subseteq \mathbb{C}_\infty$ be any circle. Then we fix $z_1, z_2, z_3 \in \Gamma$ and define the Möbius transformation $T(z) := (z, z_1, z_2, z_3)$. Given any $z \in \mathbb{C}_\infty - \{z_3\}$ we define the symmetric point of z with respect to Γ to be $z^* = T^{-1}(\overline{T(z)})$.

In other words, z^* is a symmetric point of z relative to Γ iff $(z^*, z_1, z_2, z_3) = \overline{(z, z_1, z_2, z_3)}$.

Claim: Our definition of symmetric point relative to Γ is independent of our choice of $z_1, z_2, z_3 \in \Gamma$. In other words, for any $z \in \mathbb{C}_\infty$, if $\{z_1, z_2, z_3\}$ and $\{\widetilde{z}_1, \widetilde{z}_2, \widetilde{z}_3\}$ are two triplets of distinct points in Γ with $z_3 \neq z$ and $\widetilde{z}_3 \neq z$, then after defining $T_1(z) := (z, z_1, z_2, z_3)$ and $T_2(z) := (z, \widetilde{z}_1, \widetilde{z}_2, \widetilde{z}_3)$ we have that $T_1^{-1}(\overline{T_1(z)}) = T_2^{-1}(\overline{T_2(z)})$.

Proof:

Note that $T_1 \circ T_2^{-1} =: M$ sends $\mathbb{R}$ to $\mathbb{R}$, as does $M^{-1} = T_2 \circ T_1^{-1}$. Importantly, we can thus apply our prior proposition to see that $M^{-1}(\overline{w}) = \overline{M^{-1}(w)}$ when $w \in \mathbb{C}$ and $M^{-1}(w) = T_2(T_1^{-1}(w)) \neq \infty$. In particular, if $w = M(T_2(z)) = T_1(z)$ then we have that $M^{-1}(\overline{M(T_2(z))}) = \overline{M^{-1}(M(T_2(z)))}$ if $T_1(z) \neq \infty$ and $T_2(T_1^{-1}(T_1(z))) \neq \infty$.

So for all $z \in \mathbb{C}_\infty - \{z_3, \widetilde{z_3}\}$ we have that:

$$T_1^{-1}(\overline{T_1(z)}) = T_2^{-1}(T_2(T_1^{-1}(\overline{T_1(T_2^{-1}(T_2(z)))})))$$

$$= T_2^{-1}(M^{-1}(\overline{M(T_2(z))})) = T_2^{-1}(\overline{M^{-1}(M(T_2(z)))}) = T_2^{-1}(\overline{T_2(z)}). \blacksquare$$

As a side note, for any $z \in \mathbb{C}_\infty$ we can always choose $z_3 \in \Gamma$ such that $z \neq z_3$. Hence, we can say that every point z in $\mathbb{C}_\infty$ has a well-defined symmetric point z^* relative to Γ.

Note that since z^* is the symmetric point of z relative to Γ iff $(z^*, z_1, z_2, z_3) = \overline{(z, z_1, z_2, z_3)}$, we clearly have that $(z^*)^* = z$ (relative to Γ).

<u>(Conway) Theorem III.3.19 (The Symmetry Principle):</u> If a Möbius transformation T takes a circle Γ_1 onto the circle Γ_2, then any pair of points symmetric with respect to Γ_1 are mapped by T onto a pair of points symmetric with respect to Γ_2.

Proof:

Let z_2, z_3, z_4 be distinct points in Γ_1 such that $z \neq z_4$. Then if z and z^* are symmetric with respect to Γ_1 we have that:

$$(Tz^*, Tz_2, Tz_3, Tz_4) = (z^*, z_2, z_3, z_4) = \overline{(z, z_2, z_3, z_4)} = \overline{(Tz, Tz_2, Tz_3, Tz_4)}.$$

And since Tz_2, Tz_3, Tz_4 are three distinct points of Γ_2, we have that Tz^* and Tz are symmetric with respect to Γ_2. $\blacksquare$

Can we get an explicit formula for z^*?

Firstly we need a quick observation. If $z_1, z_2, z_3, z_4 \in \mathbb{C}$ with $z_1 \neq z_4$ then:

$$\overline{(z_1, z_2, z_3, z_4)} = (\overline{z_1}, \overline{z_2}, \overline{z_3}, \overline{z_4}).$$

Why?

From the existence proof in the theorem on page 337 we know that:

$$(z_1, z_2, z_3, z_4) = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)} \text{ and } (\overline{z_1}, \overline{z_2}, \overline{z_3}, \overline{z_4}) = \frac{(\overline{z_1} - \overline{z_3})(\overline{z_2} - \overline{z_4})}{(\overline{z_1} - \overline{z_4})(\overline{z_2} - \overline{z_3})}.$$

Now let $\Gamma = \{z : |z - a| = R\}$ be a circle. Then for any three distinct points z_2, z_3, z_4 in Γ and $z \in \mathbb{C}$ with $z \neq z_4$ we have (since Mobius transforms preserve cross ratios) that:

$$(z^*, z_2, z_3, z_4) = \overline{(z, z_2, z_3, z_4)}$$

$$= \overline{(z - a, z_2 - a, z_3 - a, z_4 - a)} = \overline{\left(\tfrac{R^2}{z-a}, \tfrac{R^2}{z_2-a}, \tfrac{R^2}{z_3-a}, \tfrac{R^2}{z_4-a}\right)}$$

$$= \left(\tfrac{R^2}{\overline{z}-\overline{a}}, \tfrac{R^2}{\overline{z_2}-\overline{a}}, \tfrac{R^2}{\overline{z_3}-\overline{a}}, \tfrac{R^2}{\overline{z_4}-\overline{a}}\right)$$

$$= \left(\tfrac{R^2}{\overline{z}-\overline{a}}, \tfrac{(z_2-a)(\overline{z_2}-\overline{a})}{\overline{z_2}-\overline{a}}, \tfrac{(z_3-a)(\overline{z_3}-\overline{a})}{\overline{z_3}-\overline{a}}, \tfrac{(z_4-a)(\overline{z_4}-\overline{a})}{\overline{z_4}-\overline{a}}\right)$$

$$= \left(\tfrac{R^2}{\overline{z}-\overline{a}}, z_2 - a, z_3 - a, z_4 - a\right)$$

$$= \left(\tfrac{R^2}{\overline{z}-\overline{a}} + a, z_2, z_3, z_4\right)$$

Hence $z^* = a + \frac{R^2}{\overline{z}-\overline{a}}$ (when $\Gamma = \{z : |z-a| = R\}$). And in particular we have that $(z^* - a)(\overline{z} - \overline{a}) = R^2$.

Perhaps unsurprisingly this indicates that the symmetric point of the center of the circle is ∞.

Suppose $\Gamma = \mathbb{R}$ and let $z_1, z_2, z_3 \in \mathbb{R}$. Then put $T(z) = (z, z_1, z_2, z_3) = \frac{az+b}{cz+d}$. As mentioned before we can choose a, b, c, d to be real-valued. Thus:

$$Tz = \frac{az+b}{cz+d} = \frac{az+b}{|cz+d|^2}(c\overline{z}+d) = \frac{1}{|cz+d|^2}(ac|z|^2 + bd + bc\overline{z} + adz)$$

And specifically focusing on the imaginary component, we have that:

$$\operatorname{Im}((z, z_1, z_2, z_3)) = \frac{(ad-bc)}{|cz+d|^2}\operatorname{Im}(z).$$

This shows that $\{z : \operatorname{Im}((z, z_1, z_2, z_3)) < 0\}$ is equal to either $\{z : \operatorname{Im}(z) < 0\}$ or $\{z : \operatorname{Im}(z) > 0\}$ depending on whether $ad - bc > 0$ or $ad - bc < 0$ respectively.

Next suppose Γ is an arbitrary circle and that $z_1, z_2, z_3 \in \Gamma$. Then if S is any Möbius transformation we have that:

$$\begin{aligned}\{z : \operatorname{Im}((z, z_1, z_2, z_3)) > 0\} &= \{z : \operatorname{Im}((S(z), S(z_1), S(z_2), S(z_3))) > 0\} \\ &= S^{-1}(\{z : \operatorname{Im}((z, S(z_1), S(z_2), S(z_3))) > 0\})\end{aligned}$$

And in particular, if S maps Γ onto $\mathbb{R}_\infty$ then $\{z : \operatorname{Im}((z, z_1, z_2, z_3)) > 0\} = S^{-1}(H)$ where $H \subseteq \mathbb{C}$ is either the upper half plane or lower half plane.

For a circle, we can indicate an orientation (i.e. a direction going around the circle) of the circle by picking an ordered triple (z_1, z_2, z_3) on Γ. Intuitively, you can think of an orientation as saying that you travel in Γ in the direction going from z_1 to z_2 without passing through z_3 in the middle.

If (z_1, z_2, z_3) is an orientation of Γ then we define the right side of Γ (with respect to our orientation) to be $\{z : \operatorname{Im}((z, z_1, z_2, z_3)) > 0\}$. Similarly, we define the left side of Γ to be $\{z : \operatorname{Im}((z, z_1, z_2, z_3)) < 0\}$.

(Conway) Theorem III.3.21 (The Orientation Principle): Let Γ_1 and Γ_2 be two circles in $\mathbb{C}_\infty$ and let T be a Möbius transformation such that $T(\Gamma_1) = \Gamma_2$. Let (z_1, z_2, z_3) be an orientation for Γ_1. Then T takes the right side and the left side of Γ_1 onto the right side and left side of Γ_2 with respect to the orientation (Tz_1, Tz_2, Tz_3).

Why? Just note that $(z, z_1, z_2, z_3) = (Tz, Tz_1, Tz_2, Tz_3)$.

One more comment I'll make before moving on:

Recall that if $\Gamma \subseteq \mathbb{C}_\infty$ is a circle and $z_1, z_2, z_3 \in \Gamma$, then $f(z) := \operatorname{Im}((z, z_1, z_2, z_3))$ is equal to zero iff $z \in \Gamma$. It follows if G_+ denotes the left side of Γ relative to (z_1, z_2, z_3) and G_- denotes the right side, G_+ and G_- partition $\mathbb{C}_\infty - \Gamma$.

Also note that since f is continuous, we can show that both G_+ and G_- are clopen in $\mathbb{C}_\infty - \Gamma$. Since $\mathbb{C}_\infty - \Gamma$ is easily shown to have at most two path components, it follows that both G_+ and G_- must be the connected components of $\mathbb{C}_\infty - \Gamma$.

A consequence of this as well as the orientation principle is that if if T maps a circle $\Gamma_1 = \{z : |z-a| = R\}$ to another circle $\Gamma_2 = \{z : |z-a'| = R'\}$, then either T maps the interior Γ_1 to the whole interior of Γ_2 or T maps the interior of Γ_1 to the whole exterior of Γ_2.

Actually I have another comment to make as well.

By our construction on *page 337*, we know that $(z, z_1, z_2, z_3) = (z, z_1, z_3, z_2)^{-1}$. In particular, since $\frac{1}{w} = \frac{\overline{w}}{|w|^2}$, this means that the orientation (z_1, z_3, z_2) has the oppositive left vs right sides as does the orientation (z_1, z_2, z_3).

The next topic covered by math 220a is proving Cauchy's formula and then doing a bunch of stuff with that such as proving that holomorphic functions are analytic. Since I already took notes on a bunch of this last Spring, I'm going to intentionally skip over a lot of this.

<u>(Conway) Proposition IV.2.1:</u> Let $\varphi : [a,b] \times [c,d] \to \mathbb{C}$ be a continuous function and define $g : [c,d] \to \mathbb{C}$ by $g(t) = \int_a^b \varphi(s,t)\mathrm{d}s$. Then g is continuous. Moreover, if $\frac{\partial \varphi}{\partial t}(s,t)\mathrm{d}s$ and is continuous on $[a,b] \times [c,d]$ then g is continuously differentiable with $g'(t) = \int_a^b \frac{\partial \varphi}{\partial t}(s,t)\mathrm{d}s$.

Why? Just use the fact that φ or $\frac{\partial \varphi}{\partial t}$ is continuous on a compact domain in order to get an upper bound and then apply the theorem from math 240a (see *page 189*).

<u>(Conway) Theorem IV.2.8:</u> Let f be holomorphic on $B_R(a)$ (the Euclidean ball of radius R about a). Then $f(z) = \sum_{n=0}^\infty a_n(z-a)^n$ for $|z-a| < R$ (where $a_n = \frac{1}{n!}f^{(n)}(a)$) and this series has radius of convergence $\geq R$.

Proof:

Let $0 < r < R$ and define $\gamma_r(t) = a + re^{it}$. Then by Cauchy's formula we know for all $z \in B_r(a)$ that $f(z) = \frac{1}{2\pi i}\int_{\gamma_r} \frac{f(w)}{w-z}\mathrm{d}w$.

But now note that $\left|\frac{z-a}{w-a}\right| = \frac{|z-a|}{r} < 1$ for all $z \in B_r(a)$ and $w \in \gamma_r([0,2\pi])$. Hence:

$$\sum_{n=0}^{\infty} \frac{(z-a)^n}{(w-a)^{n+1}} = \frac{1}{w-a} \cdot \frac{1}{1-\frac{z-a}{w-a}} = \frac{1}{w-a-z+a} = \frac{1}{w-z}.$$

It follows that $f(z) = \frac{1}{2\pi i}\int_{\gamma_r} \sum_{n=0}^{\infty} \frac{f(w)}{(w-a)^{n+1}}(z-a)^n \mathrm{d}w$.

Next let $M = \max\{f(\gamma_r(t)) : t \in [0, 2\pi]\}$. Then:

$$\sum_{n=0}^{\infty}\int_0^{2\pi} \left|\frac{f(\gamma_r(t))\gamma_r'(t)(z-a)^n}{(\gamma_r(t)-a)^{n+1}}\right| \mathrm{d}t = \sum_{n=0}^{\infty}\int_0^{2\pi} \left|\frac{f(re^{it}+a)ire^{it}(z-a)^n}{(re^{it}+a-a)^{n+1}}\right| \mathrm{d}t$$

$$= \sum_{n=0}^{\infty}\int_0^{2\pi} \frac{|f(re^{it}+a)| \cdot |z-a|^n}{r^n}\mathrm{d}t \leq 2\pi M \sum_{n=0}^{\infty}\left(\frac{|z-a|}{r}\right)^n < \infty$$

Therefore, we can apply Theorem 2.25 from Folland's real analysis book (see my math 240a paper notes) to get that for all $z \in B_r(a)$:

$$f(z) = \frac{1}{2\pi i}\int_{\gamma_r}\sum_{n=0}^{\infty}\frac{f(w)}{(w-a)^{n+1}}(z-a)^n\mathrm{d}w = \frac{1}{2\pi i}\int_0^{2\pi}\sum_{n=0}^{\infty}\frac{f(\gamma_r(t))\gamma_r'(t)(z-a)^n}{(\gamma_r(t)-a)^{n+1}}\mathrm{d}t$$
$$= \sum_{n=0}^{\infty}\frac{1}{2\pi i}\int_0^{2\pi}\frac{f(\gamma_r(t))\gamma_r'(t)(z-a)^n}{(\gamma_r(t)-a)^{n+1}}\mathrm{d}t$$
$$= \sum_{n=0}^{\infty}\left(\frac{1}{2\pi i}\int_{\gamma_r}\frac{f(w)}{(w-a)^{n+1}}\mathrm{d}w\right)(z-a)^n$$

And now it follows that $\frac{1}{2\pi i}\int_{\gamma_r}\frac{f(w)}{(w-a)^{n+1}}\mathrm{d}w = \frac{1}{n!}f^{(n)}(a)$.

I'd like to pause and note that we've essentially proven the following generalization of Cauchy's formula that:

$$f^{(n)}(a) = \frac{n!}{2\pi i}\int_{\gamma_r}\frac{f(w)}{(w-a)^{n+1}}\mathrm{d}w.$$

Since $\frac{1}{n!}f^{(n)}(a)$ doesn't depend on r, we can now safely take $r \to R$ to finish proving the theorem. ■

I don't have time right now unfortunately to review what the total variation function of a curve is and write some theorems on it. Also, the professor for math 220a has said that he views it as outside the scope of the class. So, I'll cover the following theorem by Conway but weaken it to a level I can easily prove right now.

<u>(Conway) Lemma IV.2.7:</u> Let γ be a piecewise C^1 curve in $\mathbb{C}$ and suppose that F_n and F are continuous functions on the <u>trace</u> (i.e. image) of γ (which Conway denotes as $\{\gamma\}$). If $F_n \to F$ uniformly on $\{\gamma\}$ then $\int_\gamma F = \lim_{n\to\infty}\int_\gamma F_n$.

Proof:

Let a, b be the endpoints of the domain of γ and set $A = \sup\{|\gamma'(t)| : t \in [a,b]\}$. Then for any $\varepsilon > 0$ we can choose $N \in \mathbb{N}$ such that $|F(\gamma(t)) - F_n(\gamma(t))| < \varepsilon$ for all $t \in [a,b]$ and $n > N$. Hence for all $n > N$:

$$|\int_\gamma F - \int_\gamma F_n| = |\int_\gamma (F - F_n)| \leq \int_a^b |F(\gamma(t)) - F_n(\gamma(t))| \cdot |\gamma'(t)|\mathrm{d}t \leq A(b-a)\varepsilon.$$

And by taking $\varepsilon \to 0$ we are done. ■

<u>Cauchy's Estimate:</u> Suppose f is holomorphic in $B_R(a)$ and $|f| \leq M$ in $B_R(a)$. Then $|f^{(n)}(a)| \leq \frac{Mn!}{R^n}$.

I already showed for all $r < R$ that $f^{(n)}(a) = \frac{n!}{2\pi i}\int_{\gamma_r}\frac{f(w)}{(w-a)^{n+1}}\mathrm{d}w$ where $\gamma_r(t) = re^{it} + a$. Thus we have that:

$$|f^{(n)}(a)| \leq \frac{n!M}{2\pi r^{n+1}}\int_{\gamma_r}\mathrm{d}w = \frac{n!M}{r^n}$$

And the proof is completed by taking $r \to R$. ■

Unfortunately I now need to do homework. So...

Exercise IV.1.10: Define $\gamma(t) = e^{it}$ for $0 \leq t \leq 2\pi$ and find $\int_\gamma z^n \mathrm{d}z$ for every integer n.

Note that $\int_\gamma z^n \mathrm{d}z = \int_0^{2\pi} (e^{it})^n \cdot ie^{it}\mathrm{d}t = i\int_0^{2\pi} e^{i(n+1)t}\mathrm{d}t$. Thus if $n = -1$ we have that $\int_\gamma z^n \mathrm{d}z = 2\pi i$. Meanwhile, if $n \neq -1$ then:

$$\int_\gamma z^n \mathrm{d}z = \frac{i}{i(n+1)}(e^{i(n+1)(2\pi)} - e^{i(n+1)(0)}) = \frac{1}{n+1}(1-1) = 0.$$

Exercise IV.1.12: Let $I(r) = \int_\gamma \frac{e^{iz}}{z}\mathrm{d}z$ where $\gamma : [0,\pi] \to \mathbb{C}$ is defined by $\gamma(t) = re^{it}$. Then show that $\lim_{r\to\infty} I(r) = 0$.

To start off, note that $\int_\gamma \frac{e^{iz}}{z}\mathrm{d}z = \int \frac{e^{i\gamma(t)}}{re^{it}} ire^{it}\mathrm{d}t = i\int_0^\pi e^{i(re^{it})}\mathrm{d}t$. Therefore:

$$\begin{aligned}|I(r)| = |\int_0^\pi e^{i(re^{it})}\mathrm{d}t| \leq \int_0^\pi |e^{i(re^{it})}|\mathrm{d}t &= \int_0^\pi e^{\mathrm{Re}(ire^{it})}\mathrm{d}t \\ &= \int_0^\pi e^{-r\sin(t)}\mathrm{d}t = 2\int_0^{\pi/2} e^{-r\sin(t)}\mathrm{d}t.\end{aligned}$$

Then to show that $2\int_0^{\pi/2} e^{-r\sin(t)}\mathrm{d}t \to 0$ as $r \to \infty$, we first consider the inequality that $\sin(t) \geq \frac{2}{\pi}t$ when $t \in [0, \pi/2]$.

To show this inequality, let $f(t) = \sin(t) - \frac{2}{\pi}t$ and then note that $f(0) = f(\pi/2) = 0$. Since f is differentiable with $f'(t) = \cos(t) - \frac{2}{\pi}$, we thus know by the mean value theorem that there is some $t_0 \in (0, \pi/2)$ such that $f'(t_0) = 0$.

Next note that $f''(t) = -\sin(t) < 0$ for all $t \in (0, \frac{\pi}{2})$. Thus f' must be strictly decreasing on $[0, \frac{\pi}{2}]$. And in particular, we can now say that $f'(t) > 0$ on $[0, t_0]$ and $f'(t) < 0$ on $[t_0, \pi/2]$. Finally, this means that $f(t) \geq f(0) = 0$ when $t \in [0, t_0]$ and $f(t) \geq f(\pi/2) = 0$ when $t \in [t_0, \pi/2]$.

This proves that $\sin(t) - \frac{2}{\pi}t \geq 0$ when $t \in [0, \pi/2]$.

Now we can say that $2\int_0^{\pi/2} e^{-r\sin(t)}\mathrm{d}t \leq 2\int_0^{\pi/2} e^{-\frac{2r}{\pi}t}\mathrm{d}t = \frac{-\pi}{r}(e^{\frac{-2r}{\pi}\cdot\frac{\pi}{2}} - 1) = \frac{\pi}{r} - \frac{\pi}{r}e^{-r}$. And the last expression obviously goes to 0 as $r \to \infty$. Hence we've shown that $|I(r)| \to 0$ as $r \to \infty$. ■

Exercise IV.1.22: Show that if F_1 and F_2 are primitives for $f : G \to \mathbb{C}$ and G is connected then there is a constant c such that $F_1(z) = c + F_2(z)$ for each $z \in G$.

I guess to start off I'll define what a primitive is. If $f : G \to \mathbb{C}$ is continuous and F is $\mathbb{C}$-differentiable in G with $F' = f$ then F is a primitive of f.

Suppose both $F_1'(z) = f(z)$ and $F_2'(z) = f(z)$. Then $(F_1 - F_2)'(z) = 0$ everywhere on G. So by a prior theorem in the class, since G is connected we know for all $z \in G$ that $(F_1 - F_2)(z) = c$ for some constant $c \in \mathbb{C}$. In turn, $F_1(z) = c + F_2(z)$.

Math 200 homework:

Set 5 Problem 5: In a 15-puzzle, a player can rearrange the blocks labeled 1 through 15 by sliding numbers into the empty slot. Is there a way to go from the left position to the right position?

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

2	1	3	4
5	6	7	8
9	10	11	12
13	14	15	

The answer is no. And to see this, let the left image define a labeling of the possible positions of the blocks such that block i is in position i and the empty slot is in position 16. Then every possible board layout can be described by a permutation $\sigma \in S_{16}$ where $\sigma(i)$ denotes which space block i is in for $i \leq 15$ and $\sigma(16)$ denotes the position of the empty slot. Note that the left board position is associated to the identity permutation and the right board position (which I'll denote as σ') is associated with the 2-cycle $(1\ \ 2)$.

Now sliding a block into the empty slot of a board layout represented by σ is equivalent to composing a certain 2-cycle with σ. Hence, being able to go from the left position to the right position is equivalent to finding an integer $n > 0$ and a certain decomposition of transpositions $\sigma' = \tau_1, \ldots, \tau_n \mathrm{Id} = \tau_1, \ldots, \tau_n$ satisfying properties I don't really want to try and describe. Also, since $\sigma' = (1\ \ 2)$ is an odd permutation, we know that n in this hypothetical decomposition must be odd.

But now I claim that if after n moves we have that the empty slot is in the bottom right, we must have that n is even. To see this, let u, d, l, r be the number of up, down, left, and right translations of the empty slot. Also let $(0,0)$ denote the starting position of the empty slot and let (x,y) denote how many columns to the left and how may rows upwards the empty slot winds up. Then:

$$u(0,1) - d(0,1) + l(1,0) - r(1,0) = (l-r, u-d).$$

So if the empty slot ends where it started, we know that $l - r = 0$ and $u - d = 0$. Hence the total number of moves is $n = l + r + u + d = 2(l+d)$ (which is even).

This proves that no series of moves takes the left board position to the right board position.

Set 5 Problem 6: Suppose G is a finite group of order $2^k m$ where k is a positive integer and m is an odd number. Suppose G has a cyclic Sylow 2-subgroup. Then prove that G has a characteristic subgroup of order m.

We shall proceed by induction on k.

Let $\phi : G \to S_G$ be the induced homomorphism given by the action $G \curvearrowright G$ by left-translation. Since G has a cyclic Sylow 2-subgroup we know that that G has an element g of order 2^k. Then in turn, we know that $\phi(g)$ has a cycle decomposition consisting of m disjoint 2^k-cycles. Since m is odd and cycles of even length can be gotten via an odd number of transpositions, it follows that $\phi(g)$ is an odd permutation. So, letting $\varepsilon : S_G \to \{\pm 1\}$ denote the sign permutation we have that $\varepsilon \circ \phi : G \to \{\pm 1\}$ is a surjective map.

Next let $N := \ker(\varepsilon \circ \phi)$. By the first isomorphism theorem we know that

$$[G : N] = |\mathrm{im}(\varepsilon \circ \phi)| = 2.$$

Also suppose $\theta : G \to G$ is any automorphism. Then $o(\theta^{-1}(g)) = o(g)$ and so the cycle type of $\phi(\theta^{-1}(g))$ is the same as $\phi(g)$ (namely $o(g)$ repeated $|G|/o(g)$ many times). In turn, we know that $\varepsilon(\phi(\theta^{-1}(g))) = \varepsilon(\phi(g))$. So $\theta^{-1}(g) \in N$ iff $g \in N$. This proves that $g \in \theta(N)$ iff $g \in N$. So N is a characteristic subgroup of G with order $2^{k-1}m$.

If $k = 1$, this proves our base case. Meanwhile, if $k > 1$, then note that N also has a cyclic Sylow 2-subgroup.

Let P be a cyclic Sylow 2-subgroup of G with generator a. Then for any Sylow 2-subgroup Q of N we know that there exists $x \in G$ with $Q \subseteq xPx^{-1}$. In particular, this means that Q is a subgroup of the cyclic group $\langle xax^{-1} \rangle$. It follows that Q is also a cyclic group since all subgroups of cyclic groups are cyclic.

Thus by induction we can conclude that N has a characteristic subgroup H of order m. And all we need to do now is show that H is also a characteristic subgroup of G. Fortunately, note that if $\theta \in \mathrm{Aut}(G)$ then we know that $\theta|_N \in \mathrm{Aut}(N)$ since $\theta(N) = N$. So, $\theta(H) = \theta|_N(H) = H$ for all $\theta \in \mathrm{Aut}(G)$.

Thanks Hugo (the TA) for the corrections.

Set 5 Problem 2: Recall *problem 5 on homework 4*. We're now going to prove that $\mathrm{Inn}(S_n) = \mathrm{Aut}(S_n)$ when $n \geq 7$.

(a) Suppose ϕ is an automorphism of S_n which sends transpositions to transpositions. In other words, $\phi((a\ \ b))$ is a 2-cycle for every $1 \leq a < b \leq n$. Then prove that ϕ is an inner automorphism. (For this part it is enough to assume $n \geq 5$).

Consider the complete graph K_n on $\{1, \ldots, n\}$. Then there is an obvious bijective correspondance between the edges of this graph and the 2-cycles in S_n. Also, since ϕ is an automorphism sending 2-cycles to 2-cycles in S_n, we can thus define a bijective map $\widetilde{\phi}$ on the edges of K_n sending (i, j) to the edge associated with $\phi((i\ \ j))$.

Notice that for any two distinct transpositions τ_1 and τ_2 we have that $\tau_1\tau_2 = \tau_2\tau_1$ if and only if both transpositions have disjoint supports. But also note since ϕ is a group automorphism that $\phi(\tau_2)\phi(\tau_1) = \phi(\tau_1)\phi(\tau_2)$ iff $\tau_2\tau_1 = \tau_1\tau_2$. Since $\phi(\tau)$ is a 2-cycle for all 2-cycle $\tau \in S_n$, we can thus conclude that $\phi(\tau_1)$ and $\phi(\tau_2)$ have disjoint supports iff τ_1 and τ_2 have disjoint supports. Consequently, we know that two edges (a, b) and (c, d) in K_n are adjacent iff $\widetilde{\phi}((a, b))$ and $\widetilde{\phi}((c, d))$ are adjacent.

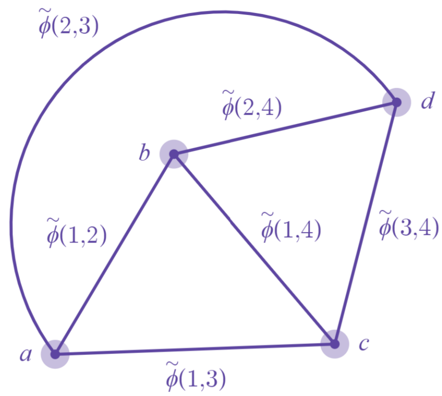
Now we show that $\widetilde{\phi}$ induces a graph automorphism of K_n.

To start off, suppose i_1, i_2, i_3 are distinct elements in $\{2, \ldots, n\}$. Then since $\widetilde{\phi}((1, i_1))$ and $\widetilde{\phi}((1, i_2))$ are adjacent to each other, we know there are distinct elements $a, b, c \in \{1, \ldots, n\}$ with $\widetilde{\phi}((1, i_1)) = (a, b)$ and $\widetilde{\phi}((1, i_2)) = (a, c)$. But now note that since $\widetilde{\phi}((1, i_3))$ is distinct from but adjacent to both (a, b) and (a, c), we must either have that $\widetilde{\phi}((1, i_3)) = (b, c)$ or we must have there exists $d \notin \{1, a, b, c\}$ such that $\widetilde{\phi}((1, k)) = (a, d)$. If $n \geq 5$ though then the former can't happen.

Why?

Let i_4 be distinct from $1, i_1, i_2, i_3$. Then if the former is true we must have that $\widetilde{\phi}(1, i_4)$ is adjacent to (a, b), (a, c), and (b, c). But that's impossible.

As a side note since I already wasted an hour and a half proving this before realizing it wasn't necessary, if $n = 4$ then this configuration of $\widetilde{\phi}$ would preserve adjacent edges. Although perhaps the other properties from ϕ being a group automorphism would guarentee this configuration never happens?



It now follows that there is some element $a \in \{1, \ldots, n\}$ such that a is a vertex of $\widetilde{\phi}((1, i))$ for all $i \neq 1$. Hence, we now define $\sigma(1) := a$ and set $\sigma(i)$ to whatever makes it so that $\widetilde{\phi}((1, i)) = (\sigma(1), \sigma(i))$. Clearly $\sigma \in S_n$.

Our next claim is that $\widetilde{\phi}((i, j)) = (\sigma(i), \sigma(j))$ for all distinct $i, j \in \{1, \ldots, n\}$. This is obvious from the definition of σ if either i or j is 1. Otherwise, $\widetilde{\phi}((i, j))$ is adjacent to both $\widetilde{\phi}((1, i)) = (\sigma(1), \sigma(i))$ and $\widetilde{\phi}((1, j)) = (\sigma(1), \sigma(j))$. And since $\widetilde{\phi}$ is a bijection of the edges of K_n and $\widetilde{\phi}((1, \sigma^{-1}(k))) = (\sigma(1), k)$ for any $k \in \{1, \ldots, n\}$, we know that $\widetilde{\phi}((i, j))$ can't equal $(\sigma(1), k)$ for any k. Hence, we must have that $\widetilde{\phi}((i, j)) = (\sigma(i), \sigma(j))$ as claimed.

With that, we finally have shown the existence of $\sigma \in S_n$ such that:

$$\phi((i\ \ j)) = (\sigma(i)\ \ \sigma(j)) = \sigma(i\ \ j)\sigma^{-1} \text{ for all 2-cycles } (i\ \ j) \in S_n.$$

Since the 2-cycles generate all of S_n, we can conclude that $\phi(\tau) = \sigma\tau\sigma^{-1}$ for all $\tau \in S_n$.

In fact if G is a group, $S \subseteq G$ is symmetric with $\langle S \rangle = G$, and $\theta_1, \theta_2 \in \mathrm{Aut}(G)$ satisfy that $\theta_1(x) = \theta_2(x)$ for all $x \in S$, then $\theta_1 = \theta_2$.

This is because for any $g \in G$, we may write $g = x_1 x_2 \cdots x_k$. And now:

$$\theta_1(g) = \theta_1(x_1) \cdots \theta_1(x_k) = \theta_2(x_1) \cdots \theta_2(x_k) = \theta_2(g).$$

(b) Suppose $\phi \in \mathrm{Aut}(S_n)$. Prove for all $\sigma_1, \sigma_2 \in S_n$ that $\phi(\sigma_1)$ and $\phi(\sigma_2)$ are conjugate if and only if σ_1 and σ_2 are conjugate. (This is true for all automorphisms of any arbitrary group.)

If there exists $\tau \in S_n$ such that $\tau\sigma_1\tau^{-1} = \sigma_2$, then $\phi(\tau)\phi(\sigma_1)(\phi(\tau))^{-1} = \phi(\sigma_2)$. Similarly, if there exists $\tau \in S_n$ such that $\tau\phi(\sigma_1)\tau^{-1} = \phi(\sigma_2)$ then do the prior argument using ϕ^{-1}.

(c) Let T_k be the set of permutations with cycle type $(\underbrace{2 \geq \cdots \geq 2}_{k} \geq \underbrace{1 \geq \cdots \geq 1}_{n-2k})$.

For instance, T_1 is the set of 2-cycles. Prove that when $k \leq n/2$ then:

$$|T_k| = \tfrac{n(n-1)\cdots(n-2k+1)}{k!2^k} \geq \tfrac{n(n-1)}{2} \cdot \tfrac{(2k-2)!}{k!2^{k-1}}$$

To start off, there are $\binom{2k}{2,\ldots,2} = \frac{(2k)!}{2^k}$ many ways of partitioning a set with $2k$ elements into k labeled subsets with 2 elements. Meanwhile, any permutation in S_{2k} with cycle type $(2 \geq \cdots \geq 2)$ can be considered a partition of $\{1, \ldots, 2k\}$ into k *unlabeled* subsets with 2 elements. Hence we get that there are $\frac{(2k)!}{k!2^k}$ many permutations in S_{2k} with cycle type $(2 \geq \cdots \geq 2)$.

Since there are $\binom{n}{2k}$ many ways to choose which elements we don't want fixed, it follows that $|T_k| = \binom{n}{2k} \cdot \frac{(2k)!}{k!2^k} = \frac{n(n-1)\cdots(n-2k+1)}{k!2^k} = \frac{n(n-1)}{2} \cdot \frac{(n-2)\cdots(n-2k+1)}{k!2^{k-1}}$.

Finally, we finish by showing that $(n-2)\cdots(n-2k+1) \geq (2k-2)!$ when $k \leq n/2$.

Note that if $k = 1$ then both sides of the inequality will be 1 (I realize that is not the clearest from how I wrote this $(n-2)\cdots(n-2k+1)$). Meanwhile, if $k > 1$ then our proposed inequality is equivalent to saying that:

$$\prod_{j=2}^{2k-1}(n-j) = \prod_{i=n-2k+1}^{n-2} i = \prod_{i=1}^{2k-2}(n-2k+i) \geq \prod_{j=1}^{2k-2} j.$$

Clearly this is true when $k < n/2$.

(d) Prove that for every $\phi \in \mathrm{Aut}(S_n)$ there exists an integer k such that $\phi(T_1) = T_k$.

By part (b) we know that $\phi(T_1)$ must equal a set of permutations all with the same cycle type. Also note that since ϕ is an automorphism, $o(\phi(\sigma)) = o(\sigma)$ for all $\sigma \in S_n$. Therefore, if $\sigma \in T_1$ and $\phi(\sigma)$ has order type $(k_1 \geq \cdots \geq k_\ell)$ then:

$$\mathrm{lcm}(k_1, \ldots, k_\ell) = 2.$$

But, this is only possible if $\phi(\sigma) \in T_k$ for some k. Hence, we've proven that $\phi(T_1) \subseteq T_k$ for some integer k. To show the other inclusion, apply part (b) again to see that $\phi^{-1}(\tau)$ must have the same order type for all $\tau \in T_k$. In turn $\phi^{-1}(T_k) \subseteq T_1$ and this proves that $T_k = \phi(T_1)$.

(e) Prove that when $n \geq 7$, for every $\phi \in \mathrm{Aut}(S_n)$ we have that $\phi(T_1) = T_1$. Therefore by part (a), we have that $\mathrm{Aut}(S_n) = \mathrm{Inn}(S_n)$.

Note that $|T_1| = \frac{n(n-1)}{2}$. Meanwhile, we claim that $\frac{(2k-2)!}{k!2^{k-1}} > 1$ whenever $k \geq 3$.

After all, it is easy to check that $\frac{(2k-2)!}{k!2^{k-1}} = 1$ when $k = 1$ or 3. Meanwhile, $\frac{(2j-2)(2j-3)}{2j} > \frac{(2j-2)(2j-3)}{(2j-3)} = 2j - 2 > 1$ for all $j \geq 2$. Hence, we can easily show by induction that $\frac{(2k-2)!}{k!2^{k-1}} > 1$ when $k > 3$.

Hence, by part (c) we know that $|T_k| > |T_1|$ for all $k > 3$. And so, all we need to show is that $|T_1| \neq |T_2|$ and $|T_1| \neq |T_3|$.

Fortunately, note that:

- $|T_2| = \frac{n(n-1)(n-2)(n-3)}{2!2^2} = \frac{n(n-1)}{2} \cdot \frac{(n-2)(n-3)}{4}$,
- $|T_3| = \frac{n(n-1)(n-2)(n-3)(n-4)(n-5)}{3!\cdot 2^3} = \frac{n(n-1)}{2} \cdot \frac{(n-2)(n-3)(n-4)(n-5)}{24}$

Thus we will know $|T_1| \neq |T_2|$ and $|T_1| \neq |T_3|$ if n is large enough that $f(n) \coloneqq (n-2)(n-3) > 4$ and $g(n) \coloneqq (n-2)(n-3)(n-4)(n-5) > 24$. Since these polynomials of n are strictly increasing once $n > 5$, and $f(7) = 20 > 4$ and $g(7) = 120 > 24$, we've thus shown that $|T_1| \neq |T_k|$ for any $k > 1$ when $n \geq 7$.

Since $|\phi(T_1)| = |T_1|$, it follows by part (d) that $\phi(T_1) = T_1$. ■

As a side note, if $n = 5$ then we don't need to worry about T_3 since $3 > 5/2$. And since $(5-2)(5-3) = 6 > 4$, we know that $|T_2| \neq |T_1|$ when $n = 5$. Hence, we can similarly conclude using part (a) that $\mathrm{Inn}(S_5) = \mathrm{Aut}(S_5)$.

Set 5 Problem 3: For every group G, the group of outer automorphisms is $\mathrm{Out}(G) = \frac{\mathrm{Aut}(G)}{\mathrm{Inn}(G)}$.

Note that for any $g, h \in G$ and $\theta \in \mathrm{Aut}(G)$, we have that:

$$\theta(h\theta^{-1}(g)h^{-1}) = \theta(h)g(\theta(h))^{-1}.$$

This shows show that $\mathrm{Inn}(G) \lhd \mathrm{Aut}(G)$. Hence, our definition of $\mathrm{Out}(G)$ makes sense.

Let $\mathrm{Cl}(G)$ be the set of conjugacy classes of G.

(a) Prove that $(\theta\mathrm{Inn}(G)) \cdot [a] \coloneqq [\theta(a)]$ is a well-defined action of $\mathrm{Out}(G)$ on $\mathrm{Cl}(G)$ where $[g]$ is the conjugacy class of $g \in G$.

To start off, we need to show this is well-defined. Luckily, we already know from part (b) of the last problem that if $[a] = [b]$ then we have that $[\theta(a)] = [\theta(b)]$ for any $\theta \in \mathrm{Aut}(G)$. So now we just need to show that if $\theta_1\mathrm{Inn}(G) = \theta_2\mathrm{Inn}(G)$ then $[\theta_1(a)] = [\theta_2(a)]$. Note that because $\theta_1 \circ \theta_2^{-1} \in \mathrm{Inn}(G)$ and $\theta_1 = (\theta_1 \circ \theta_2^{-1}) \circ \theta_2$, there must exist $h \in G$ with $\theta_1(g) = h\theta_2(g)h^{-1}$ for all $g \in G$. In particular, this means that $\theta_1(a) = h\theta_2(a)h^{-1}$. So $\theta_1(a) \in [\theta_2(a)]$.

Next we show that $(\theta\mathrm{Inn}(G)) \cdot [a]$ is actually a group action.

- If $[a] = [b]$ then $[hah^{-1}] = [hbh^{-1}]$ for all $h \in G$. Hence $\mathrm{Inn}(G) \cdot [a] = [a]$ for all $a \in G$.
- $$\begin{aligned}(\theta_1\mathrm{Inn}(G)) \cdot ((\theta_2\mathrm{Inn}(G)) \cdot [a]) &= (\theta_1\mathrm{Inn}(G)) \cdot [\theta_2(a)] \\ &= [(\theta_1 \circ \theta_2)(a)] = ((\theta_1 \circ \theta_2)\mathrm{Inn}(G)) \cdot [a].\end{aligned}$$

(b) Argue why $f : \mathrm{Cl}(G) \to \mathbb{Z} \times \mathbb{Z}$ given by $f([g]) \coloneqq (o(g), |[g]|)$ is fixed along an $\mathrm{Out}(G)$-orbit.

Suppose $(\theta\mathrm{Inn}(G)) \cdot [a] = [b]$. Then we know that $[\theta(a)] = [b]$. Hence, $o(a) = o(\theta(a)) = o(b)$ since the order of elements of G is conserved by automorphisms of G (and in particular by conjugation).

Next note that by the orbit-stabilizer theorem (applied to the action $G \curvearrowright G$ by conjugation), we have that $|[g]| = [G : C_G(g)]$ for all $g \in G$. So if we can show that $|C_G(a)| = |C_G(b)|$ then we will be done. Fortunately, note that if $\phi \in \mathrm{Aut}(G)$ then $\phi(C_G(a)) = C_G(\phi(a))$. Hence if $b = h\theta(a)h^{-1}$ then:

$$|C_g(b)| = |C_g(h\theta(a)h^{-1})| = |h\theta(C_g(a))h^{-1}| = |C_g(a)|.$$

(c) Prove that $\mathrm{Aut}(S_n) = \mathrm{Inn}(S_n)$ if $n \neq 6$.

By the way I'm now doing this after the due date...

You can see the side note I left before the start of problem 3 for why we are already done when $n \geq 5$. As for showing the cases for when $n = 1, 2, 3, 4$, the professor's provided hint is just to modify the argument of part (a) of problem 2. So, I guess I'll start by showing that all automorphisms of S_n take transpositions to other transpositions when $n < 5$.

Let $T_k \subseteq S_n$ be as in problem 2 for each integer k. You could use part (b) of this problem or just reuse the arguments from problem 2 to see that if $\phi \in \mathrm{Aut}(S_n)$ then $\phi(T_1) = T_k$ for some integer k. But now note that:

$|T_1| = \frac{n(n-1)}{2}$, $|T_2| = \frac{n(n-1)(n-2)(n-3)}{8}$, and $|T_3| = \frac{n(n-1)(n-2)(n-3)(n-4)(n-5)}{48}$.

So we can manually calculate out the following table:

	$n=1$	$n=2$	$n=3$	$n=4$	$n-5$	$n=6$
$\lvert T_1\rvert$	0	1	3	6	10	15
$\lvert T_2\rvert$	0	0	0	3	15	45
$\lvert T_3\rvert$	0	0	0	0	0	15

From this table hopefully it is clear why we wouldn't expect every automorphism of S_6 to take transpositions to transpositions. After all, there isn't a combinatorial reason why an automorphism of S_6 wouldn't map T_1 to T_3 instead. That said, we know by this table plus the facts that $|\phi(T_1)| = |T_1|$ and $\phi(T_1) = T_k$ for some k that $\phi(T_1)$ must equal T_1 when $n \leq 5$. In particular, this means that when $n \leq 4$ then all automorphisms map transpositions to transpositions.

Next, we return to our proof of part (a) in problem 2. The only part where we relied on n being larger than 5 was when we were showing that there exists $a \in \{1, \ldots, n\}$ such that $\widetilde{\phi}((1,i))$ includes a for any $i \neq 1$. Fortunately when $n = 1$ or $n = 2$ then this is trivially true. Also, when $n = 3$ then there are only two 2-cycles containing 1. And as their images under $\widetilde{\phi}$ are adjacent, we once again can find our desired a in $\{1, 2, 3\}$.

Finally, we get to the case that $n = 4$. If a, b, c, d are the four elements of $\{1, 2, 3, 4\}$ and we have that $\widetilde{\phi}((1,2)) = (a, b)$ and $\widetilde{\phi}((1,3)) = (a, c)$, then recall that if $\widetilde{\phi}((1,4)) \neq (a, d)$ then we must have that $\widetilde{\phi}((1,4)) = (b, c)$.

But that would imply that:

$$\phi((1\ \ 4\ \ 3\ \ 2)) = \phi((1\ \ 2)(1\ \ 3)(1\ \ 4)) = (a\ \ b)(a\ \ c)(b\ \ c) = (a\ \ c) = \phi((1\ \ 3)).$$

Since ϕ is an injective function, this is a contradiction. Hence, we conclude that even when $n = 4$ we must still have that $\widetilde{\phi}((1,4)) = (a, d)$ and not (b, c).

(d) Prove that $\operatorname{Aut}(S_n) \cong S_n$ when n is not 2 or 6.

By applying problem 2 from the second problem set (see *pages 284-285*) to S_n where $n \neq 6$, we thus know that:

$$S_n/Z(S_n) \cong \operatorname{Inn}(S_n) = \operatorname{Aut}(S_n).$$

But now note that if $n \neq 2$ then $Z(S_n)$ is trivial. To see this, first note there is nothing to prove when $n = 1$ and hence we can focus on when $n \geq 3$.

We claim for any permutation $\sigma \neq \mathrm{Id}$ in S_n that there exists a permutation $\tau \in S_n$ such that $\tau\sigma \neq \sigma\tau$. After all, let i, j be distinct elements of $\{1, \ldots, n\}$ such that $\sigma(j) = i$. Then let k be a third element distinct from i and j and define $\tau = (j \;\; k)$. Now $\tau \circ \sigma(j) = \tau(i) = i$ but $\sigma \circ \tau(j) = \sigma(k) \neq \sigma(j) = i$.

This shows that $\tau\sigma \neq \sigma\tau$. And importantly, this means that:

$$\sigma \notin C_{S_n}(\tau) \subseteq Z(S_n).$$

So when $n \neq 2$ and $n \neq 6$, we have that:

$$S_n \cong S_n/Z(S_n) \cong \operatorname{Inn}(S_n) = \operatorname{Aut}(S_n). \blacksquare$$

11/3/2025

More math 200a notes:

A normal series of G is a list of subgroups such that $G = G_0 \triangleright G_1 \triangleright G_2 \triangleright \cdots \triangleright G_m$. We call the normal series a composition series if $G_m = \{1\}$ and G_i/G_{i+1} is simple and nontrivial for $0 \leq i < m$. Also, given a composition series $G = G_0 \triangleright \cdots \triangleright G_m = \{1\}$ we say that each G_i/G_{i+1} is a composition factor of G.

Jordan-Hölder Theorem: Suppose G is a finite group.

(a) G has a composition series.

Proof:
Let $G_0 \triangleright \cdots \triangleright G_m = \{1\}$ be a longest possible normal series of G with $G_i \supsetneqq G_{i+1}$ for each i. Note that we know there is a maximum length for any such normal series of G since we clearly can't have a normal series of length greater than $|G|$.

We claim this is a composition series. After all, if not there exists i such that G_i/G_{i+1} is not simple. But now there exists a normal subgroup $\{1\} \subsetneqq \overline{N} \subsetneqq G_i/G_{i+1}$. And by the correspondance theorem this means that there exists $N \triangleleft G_i$ with $G_{i+1} \subsetneqq N \subsetneqq G_i$. But now we have a contradiction since $G_0 \triangleright \cdots \triangleright G_i \triangleright N \triangleright G_{i+1} \triangleright \cdots G_m$ is a longer normal series.

(b) If $G_0 \triangleright \cdots \triangleright G_m = \{1\}$ and $H_0 \triangleright \cdots \triangleright H_\ell = \{1\}$ are both composition series of G then $\ell = m$ and there exists $\sigma \in S_n$ such that $H_{i-1}/H_i \cong G_{\sigma(i)-1}/G_{\sigma(i)}$.

Proof:
We'll say $(G_0, \ldots, G_m) \sim (H_0, \ldots, H_\ell)$ if $m = \ell$ and $G_i/G_{i+1} \cong H_i/H_{i+1}$ after some reordering of the H_i. We shall proceed by induction on $|G|$ to show that if $G_0 \triangleright \cdots \triangleright G_m = \{1\}$ and $H_0 \triangleright \cdots \triangleright H_\ell = \{1\}$ are composition series of G then $(G_0, \ldots, G_m) \sim (H_0, \ldots, H_\ell)$.

(For our base case, note that if G is simple then the only composition series of G is $G \triangleright \{1\}$. In particular, this means that when $|G| \leq 3$ our theorem is trivially true.)

Now for our first induction case, suppose $G_1 = H_1$. Then by our induction hypothesis we know that $(G_1, G_2, \ldots, G_m) \sim (H_1, H_2, \ldots, H_m)$. And by concatenating on $G = G_0 = H_0$ we are done.

Next suppose $G_1 \neq H_1$. Then $G_1 \subsetneq G_1H_1$. But also $G_1 \triangleleft G_1H_1 \triangleleft G$ as both $G_1 \triangleleft G$ and $H_1 \triangleleft G$.

After all, if $x \in G$ then $xghx^{-1} = xgx^{-1}xhx^{-1} \in G_1H_1$ for any $g \in G_1$ and $h \in H_1$ since both $G_1 \triangleleft G$ and $H_1 \triangleleft G$.

Now $\{1\} \neq G_1H_1/G_1 \triangleleft G/G_1$ by the correspondence theorem. And since G/G_1 is simple this implies that $G_1H_1 = G$. Hence, by the second isomorphism theorem we have that $G_0/G_1 = G_1H_1/G_1 \cong H_1/G_1 \cap H_1$. By analogous reasoning we can also show that $H_0/H_1 = G_1H_1/H_1 \cong G_1/G_1 \cap H_1$.

Now suppose $G_1 \cap H_1 = K_1 \triangleright K_1 \triangleright \cdots \triangleright K_s = \{1\}$ is a composition series of $G_1 \cap H_1$. Then $G_1 \triangleright K_0 \triangleright \cdots K_s$ and $H_1 \triangleright K_0 \triangleright \cdots \triangleright K_s$ are composition series of G_1 and H_1. So by induction, we have that:

$(G_1, G_2, \ldots, G_m) \sim (G_1, K_0, \ldots, K_s)$ and $(H_1, H_2, \ldots, H_m) \sim (H_1, K_0, \ldots, K_s)$.

Finally, note that $(G_0, G_1, K_0, \ldots, K_s) \sim (H_0, H_1, K_0, \ldots, K_s)$. This is because $G_0/G_1 \cong H_1/K_0$ and $H_0/H_1 \cong G_1/K_0$. And since we can see by concatenating on $G = G_0 = H_0$ that:

$(G_0, G_1, G_2, \ldots, G_m) \sim (G_0, G_1, K_0, \ldots, K_s)$ and
$(H_0, H_1, H_2, \ldots, H_m) \sim (H_0, H_1, K_0, \ldots, K_s)$,

it's now clear that $(G_0, \ldots, G_m) \sim (H_0, \ldots, H_\ell)$. ■

I'm now going to show that if $n \geq 5$ and $N \triangleleft S_n$ and $1 \subsetneq N \subsetneq S_n$ then $N = A_n$.

Suppose $N \triangleleft S_n$ and $N \neq A_n$. Then $N \cap A_n \triangleleft A_n$. And since A_n is simple we know that either $N \cap A_n = \{\mathrm{Id}\}$ or $N \cap A_n = A_n$. Suppose the latter is true. Then we know that $A_n \subsetneq N < S_n$. But now $[S_n : N] < [S_n : A_n] = 2$, this must mean that $S_n = N$.

Meanwhile, suppose $N \cap A_n = \{\mathrm{Id}\}$. Then if $N \neq \{\mathrm{Id}\}$ we must have that $N - \{\mathrm{Id}\}$ consists entirely of odd permutations. And since the composition of two odd permutations is even, this also means that for any $\sigma, \tau \in N - \{\mathrm{Id}\}$ we must have that $\sigma\tau = \mathrm{Id}$.

But now note that because N is closed under conjugation, we have for any $\sigma \in N - \{\mathrm{Id}\}$ that every permutation with the same cycle type as σ is in N. Also, every $\sigma \in N - \{\mathrm{Id}\}$ must have order 2 and thus it's cycle decomposition must consist of only 2-cycles. This now let's us split into the following two cases for when $N \neq \{\mathrm{Id}\}$.

Suppose the transposition $(a\ \ b) \in N$. Then let c be distinct from a and b and note $(b\ \ c) \in N$. But now we have a contradiction since $(a\ \ b)(b\ \ c) = (a\ \ b\ \ c) \in N$ and $(a\ \ b\ \ c) \neq \mathrm{Id}$. Meanwhile suppose $\sigma = (a\ \ b)(c\ \ d)(e\ \ f)\cdots \in N$. Then $\tau = (a\ \ c)(b\ \ d)(e\ \ f)\cdots \in N$ as well. And now $\tau\sigma(a) = d \neq a$. This contradicts that $\tau\sigma = \mathrm{Id}$. ■

Here is an example of a decomposition series. Suppose $n \in \mathbb{Z}$ and $n = p_1 p_2 \cdots p_m$ where all the p_i are prime. Then:

$$\mathbb{Z}/n\mathbb{Z} \rhd p_1\mathbb{Z}/n\mathbb{Z} \rhd p_1p_2\mathbb{Z}/n\mathbb{Z} \rhd \cdots \rhd \frac{p_1\cdots p_m\mathbb{Z}}{n\mathbb{Z}}$$

I already mentioned in a note at the end problem 2 on the second problem set (see *pages 285-286*) what a commutator and derived subgroup is. Slightly more generally, given any $H_1, H_2 < G$ we define $[H_1, H_2] := \langle [h_1, h_2] : h_1 \in H_1 \text{ and } h_2 \in H_2 \rangle$.

Given any group G, the derived series of G is given by $G^{(0)} := G$ and $G^{(i+1)} := [G^{(i)}, G^{(i)}]$. Also $G^{(1)}$ is sometimes given the special name of the commutator subgroup.

Lemma: If $N \lhd G$ then G/N is abelian if and only if $[G, G] < N$.

This is because $ghN = hgN$ if and only if $g^{-1}h^{-1}gh = [g^{-1}, h^{-1}] \in N$. So if G/N is abelian then all commutators of G must be in N.

As a side note, this also says that $G/[G, G]$ is the largest abelian quotient of G. In particular, if G/N is also abelian then there is a surjective group homomorphism $G/[G, G] \twoheadrightarrow G/N$ given by $x[G, G] \mapsto xN$.

We call $G^{\mathrm{ab}} := G/[G : G]$ the abelianization of G.

Lemma: If H_1, H_2 are characteristic subgroups of G then so is $[H_1, H_2]$.

Proof:

For all $\phi \in \mathrm{Aut}(G)$ we have that $\phi(H_1) = H_1$ and $\phi(H_2) = H_2$. Therefore:

$$\phi([h_1, h_2]) = \phi(h_1)\phi(h_2)\phi(h_1)^{-1}\phi(h_2)^{-1} = [\phi(h_1), \phi(h_2)] \in [H_1, H_2].$$

Using the above fact for both ϕ and ϕ^{-1} we can thus show that:

$$\phi(\{[h_1, h_2] : h_1 \in H_1 \text{ and } h_2 \in H_2\}) = \{[h_1, h_2] : h_1 \in H_1 \text{ and } h_2 \in H_2\}$$

And now we are done by noting that:

$$\phi(\langle [h_1, h_2] : h_1 \in H_1 \text{ and } h_2 \in H_2 \rangle) = \langle \phi(\{[h_1, h_2] : h_1 \in H_1 \text{ and } h_2 \in H_2\}) \rangle. \blacksquare$$

As a side note, by identical reasoning we could have shown that if $H_1, H_2 \lhd G$ then $[H_1, H_2] \lhd G$.

Corollary: The ith. derived subgroup $G^{(i)}$ is a characteristic subgroup of G.

Lemma: If $H < G$ then $H^{(i)} < G^{(i)}$ (where $G^{(i)}$ denotes the ith. derived subgroup).

Proof:

Suppose $N_1, N_2 < G$ with $N_1 < N_2$. Then every element generating $[N_1, N_1]$ is also in the set generating $[N_2, N_2]$. So $[N_1, N_1] < [N_2, N_2]$. Using this fact, we can inductively show that $H^{(i)} < G^{(i)}$ for all i. ■

A group G is called solvable if $G^{(n)} = \{1\}$ for some integer n.

Proposition: G is solvable iff there exists a normal series $G = G_0 \rhd G_1 \rhd \cdots \rhd G_n = \{1\}$ such that G_i/G_{i+1} is abelian for all i.

$(\Longrightarrow)$

If $G^{(n)} = \{1\}$ for some n then $G = G^{(0)} \rhd G^{(1)} \rhd \cdots \rhd G^{(n)} = \{1\}$ is a normal series. Also $G^{(i)}/G^{(i+1)}$ is abelian for all i as $[G^{(i)}, G^{(i)}] = G^{(i+1)}$.

$(\Longleftarrow)$

Suppose $G = G_0 \rhd G_1 \rhd \cdots \rhd G_n = \{1\}$ is a normal series and G_i/G_{i+1} is abelian for each i. By induction on i we can see that $G^{(i)} \subseteq G_i$ for each i. After all $G_0 = G^{(0)}$. Meanwhile, if $G^{(i)} < G_i$ then we know that $G^{(i+1)} < [G_i, G_i]$. But since G_i/G_{i+1} is abelian we must have that $[G_i, G_i] < G_{i+1}$. Hence $G^{(i+1)} < G_{i+1}$.

In particular this shows that $G^{(n)} < G_n = \{1\}$. Hence G is solvable. ■

Proposition: If G is solvable, $N \lhd G$, and $H < G$, then G/N and H are solvable.

Proof:

Clearly H is solvable since if $G^{(n)} = \{1\}$ then $H^{(n)} < G^{(n)} = \{1\}$.

Meanwhile, note that $(G/N)^{(i)} = (G^{(i)}N)/N$.

Why?

We claim that if $\varphi : G \to G'$ is a surjective group homomorphism then $\varphi(G^{(i)}) = (G')^{(i)}$. To prove this, first note that $\varphi(G^{(0)}) = (G')^{(0)}$ because φ is surjective. Next, suppose we have already shown that $\varphi(G^{(i)}) = (G')^{(i)}$. Then because $\varphi([g_1, g_2]) = [\varphi(g_1), \varphi(g_2)]$ for all $g_1, g_2 \in G^{(i)}$, we have that:

$$\varphi(G^{(i+1)}) = \varphi([G^{(i)}, G^{(i)}]) \subseteq [(G')^{(i)}, (G')^{(i)}] = (G')^{(i+1)}.$$

Meanwhile, the fact that $(G')^{(i+1)} \subseteq \varphi(G^{(i+1)})$ is because for every $h_1, h_2 \in (G')^{(i)}$ there exists $g_1, g_2 \in G^{(i)}$ with $\varphi([g_1, g_2]) = [h_1, h_2]$.

Applying this fact to the natural surjective homomorphism, $\pi : G \to G/N$, we have that $(G/N)^{(i)} = \pi(G^{(i)}) = G^{(i)}N/N$.

Since G is solvable, we know that eventually $G^{(i)} \subseteq N$. And so $(G/N)^{(i)}$ is eventually trivial. ■

Proposition: If G is nontrivial, solvable, and simple, then G is a cyclic group of prime order.

Proof:

Since G is solvable, $G^{(1)} \neq G$. And as G is simple, we know that $G^{(1)} = \{1\}$. This proves that G is abelian. Hence, all of its subgroups are normal and we once again have by the simplicity of G that $\langle x \rangle = G$ for all $x \neq 1$.

If $o(x) = \infty$ or $p \mid o(x)$ with $p \neq o(x)$ and p prime, then $G = \langle x^p \rangle \neq \langle x \rangle = G$. This is a contradiction. So, we conclude G is a cyclic group of prime order. ■

Proposition: Suppose G is a finite group. Then G is solvable if and only if all its composition factors are cyclic groups of prime order.

$(\Longrightarrow)$
Suppose $G = G_0 \rhd \cdots \rhd G_c = \{1\}$ is a composition series of G. If G is solvable, then we know by a proposition on the prior page that G_i is solvable for all i. Then by that same proposition we know that G_i/G_{i+1} is solvable for all i. And since G_i/G_{i+1} is also simple and nontrivial, we have that G_i/G_{i+1} is a cyclic group of prime order for all i.

$(\Longleftarrow)$
Suppose $G = G_0 \rhd G_1 \rhd \cdots \rhd G_c = \{1\}$ is a composition series such that G_i/G_{i+1} is a cyclic group of prime order for all i. Then G_i/G_{i+1} is abelian for all i. Hence, G is solvable. ■

Math 220a Notes:

To start, off here is an exercise that I partially turned in for math 220a (although I later completed the exercise after it was due).

Exercise IV.2.2: Prove the following analogue of Leibniz's rule:
Let $G \subseteq \mathbb{C}$ be an open set and let γ be a piecewise C^1 curve in G. Suppose that $\varphi : \{\gamma\} \times G \to \mathbb{C}$ is a continuous function and define $g : G \to \mathbb{C}$ by $g(z) = \int_\gamma \varphi(w,z)\mathrm{d}w$. Then g is continuous. And if $\frac{\partial \varphi}{\partial z}$ exists for each (w,z) in $\{\gamma\} \times G$ and is continuous then g is holomorphic with $g'(z) = \int_\gamma \frac{\partial \varphi}{\partial z}(w,z)\mathrm{d}w$.

Say that γ is defined on the interval $[a,b]$. Then:
$$g(z) = \int_\gamma \varphi(w,z)\mathrm{d}w = \int_a^b \varphi(\gamma(t),z)\gamma'(t)\mathrm{d}t.$$

Now because γ is piecewise C^1, we know that there are $a = t_0 < t_1 < \ldots < t_n = b$ such that γ restricted to $[t_{i-1}, t_i]$ has a continuously defined derivative. In turn, by using the extreme value theorem on each of those intervals and then taking a max, we can find $M > 0$ such that $|\gamma'(t)| \leq M$ for all $t \in [a,b]$ where γ' is defined. It follows that:
$$\begin{aligned}|g(z_1) - g(z_2)| &= |\int_a^b (\varphi(\gamma(t),z_1) - \varphi(\gamma(t),z_2))\gamma'(t)\mathrm{d}t| \\ &\leq M \int_a^b |\varphi(\gamma(t),z_1) - \varphi(\gamma(t),z_2)|\mathrm{d}t\end{aligned}$$

Now fix $z_2 \in G$. If $r > 0$ is small enough so that $\overline{B_r(z_2)} \subseteq G$, then we have that $[a,b] \times \overline{B_r(z_2)}$ is compact and $f(t,z) = |\varphi(\gamma(t),z)|$ is continuous on $[a,b] \times \overline{B_r(z_2)}$. It follows that f is uniformly continuous on $[a,b] \times \overline{B_r(z_2)}$. Hence, for any $\varepsilon > 0$ we can find $0 < \delta \leq r$ such that if $|t' - t| < \delta$ and $|z' - z| < \delta$ then $|f(t', z') - f(t, z)| < \varepsilon$. In particular, this means that if $|z_1 - z_2| < \delta$ then:

$$|f(t, z_1) - f(t, z_2)| = |\varphi(\gamma(t), z_1) - \varphi(\gamma(t), z_2)| < \varepsilon \text{ for all } t \in [a, b].$$

Then for all $z_1 \in B_\delta(z_2)$ we have that $|g(z_1) - g(z_2)| \leq M(b-a)\varepsilon$. And by taking $\varepsilon \to 0$ we've proven that $\lim_{z_1 \to z_2} g(z_1) = g(z_2)$. Also since z_2 was arbitrary, we know that g is continuous.

Now if we can show that $g'(z) = \int_\gamma \frac{\partial \varphi}{\partial z}(w, z)\mathrm{d}w$ then we will be able to conclude by our prior reasoning that $g'(z)$ is continuous on G.

So to start off, note that our desired identity holds if and only if:

$$\lim_{z_1 \to z_2} \int_a^b \left(\frac{\varphi(\gamma(t), z_1) - \varphi(\gamma(t), z_2)}{z_1 - z_2} - \frac{\partial \varphi}{\partial z}(\gamma(t), z_2) \right) \gamma'(t)\mathrm{d}t = 0$$

By noting that $|\gamma'(t)| \leq M$ (from the last section) for all $t \in [a, b]$, we can in turn say our theorem is proved if for any $z_2 \in G$:

$$\lim_{z_1 \to z_2} \int_a^b \left(\frac{\varphi(\gamma(t), z_1) - \varphi(\gamma(t), z_2)}{z_1 - z_2} - \frac{\partial \varphi}{\partial z}(\gamma(t), z_2) \right) \mathrm{d}t = 0$$

For ease of nation I will now denote $f(t, z) := \varphi(\gamma(t), z)$. Note that f is continuous on $[a, b] \times G$. Also importantly, $\frac{\partial f}{\partial z}(t, z) = \frac{\partial \varphi}{\partial z}(\gamma(t), z)$ is continuous on $[a, b] \times G$. So now our problem is about showing that for all $z_2 \in G$:

$$\lim_{z_1 \to z_2} \int_a^b \left(\frac{f(t, z_1) - f(t, z_2)}{z_1 - z_2} - \frac{\partial f}{\partial z}(t, z_2) \right) \mathrm{d}t = 0.$$

Fix $z_2 \in G$ and pick $r > 0$ such that $\overline{B_r(z_2)} \subseteq G$. Then for any $\varepsilon > 0$, because $[a, b] \times \overline{B_r(z_2)}$ is compact, and $\frac{\partial f}{\partial z}$ is continuous on that compact set, we can find $\delta > 0$ such that $|\frac{\partial f}{\partial z}(t', z') - \frac{\partial f}{\partial z}(t, z)| < \varepsilon$ whenever $|t' - t| < \delta$ and $|z' - z| < \delta$. In particular, this lets us conclude for any fixed t and $z_1 \in B_\delta(z_2)$ that:

$$\begin{aligned} \left| \int_{[z_2, z_1]} \frac{\partial f}{\partial z}(t, w) - \frac{\partial f}{\partial z}(t, z_2)\mathrm{d}w \right| &\leq \int_0^1 |\tfrac{\partial f}{\partial z}(t, sz_1 + (1-s)z_2) - \tfrac{\partial f}{\partial z}(t, z_2)| \cdot |z_1 - z_2|\mathrm{d}s \\ &< \varepsilon |z_1 - z_2| \end{aligned}$$

But also note that $F(z) = f(t, z) - z\frac{\partial f}{\partial z}(t, z_2)$ is a primitive for $\frac{\partial f}{\partial z}(t, z) - \frac{\partial f}{\partial z}(t, z_2)$. Therefore, we have that:

$$\begin{aligned} \left| \int_{[z_2, z_1]} \frac{\partial f}{\partial z}(t, w) - \frac{\partial f}{\partial z}(t, z_2)\mathrm{d}w \right| &= |F(z_1) - F(z_2)| \\ &= |f(t, z_1) - f(t, z_2) - (z_1 - z_2)\tfrac{\partial f}{\partial z}(t, z_2)|. \end{aligned}$$

And now we know for all $z_1 \in B_\delta(z_2)$ and $t \in [a, b]$ that: $\frac{|f(t, z_1) - f(t, z_2) - (z_1 - z_2)\frac{\partial f}{\partial z}(t, z_2)|}{|z_1 - z_2|} < \varepsilon$.

Or in other words, when $|z_1 - z_2| < \delta$ then:

$$\left| \int_a^b \left(\frac{f(t, z_1) - f(t, z_2)}{z_1 - z_2} - \frac{\partial f}{\partial z}(t, z_2) \right) \mathrm{d}t \right| \leq \int_a^b \left| \frac{f(t, z_1) - f(t, z_2) - (z_1 - z_2)\frac{\partial f}{\partial z}(t, z_2)}{z_1 - z_2} \right| \mathrm{d}t < \varepsilon(b-a).$$

And by taking $\varepsilon \to 0$ we are done. ■

I'd just like to add that this probably could have been done quicker using the dominated convergence theorem. However, I didn't think the professor would like me using it.

If $f : G \to \mathbb{C}$ is analytic and $a \in G$ satisfies that $f(a) = 0$ then a is a zero of f of multiplicity $m > 0$ if there is an analytic function $g : G \to \mathbb{C}$ such that $f(z) = (z-a)^m g(z)$ where $g(a) \neq 0$.

To see that the multiplicity of a zero is uniquely defined, suppose $f(z) = (z-a)^{m_1}g(z)$ and $f(z) = (z-a)^{m_2}h(z)$ where $m_1 < m_2$ and g and h are analytic functions that are nonzero at a. Then $g(z) = (z-a)^{m_2 - m_1}h(z)$ whenever $z \neq a$. And by taking the limit as $z \to a$ we get a contradiction as $0 \neq g(a) = 0 \cdot h(a) = 0$. So, we conclude that it is impossible for a zero to have two different multiplicities.

Theorem: Let $G \subseteq \mathbb{C}$ be a region and let $f : G \to \mathbb{C}$ be an analytic function. Then either f is zero everywhere on G or the zero set of f (denoted Z_f) has no limit points in G and each $a \in Z_f$ is a zero of some multiplicity.

For a proof of this see my notes from last Spring.

Corollary: If f and g are analytic functions on a region G that equal each other on some set with a limit point, then $f = g$ everywhere on G.

Proof:

$f - g$ is analytic on G and has a non discrete zero set. Hence $f - g$ must equal zero everywhere on G. ■

Observe that if f is a polynomial of degree $m > 0$ (so that $f(z) = a_0 + a_1 z + \ldots + a_m z^m$ where $a_m \neq 0$), then in the absolute worst case:

$$|f(z)| \geq |a_m| \cdot |z|^m (1 - \left|\frac{a_{m-1}}{a_m}\right| \frac{1}{|z|} - \cdots - \left|\frac{a_0}{a_m}\right| \frac{1}{|z|^m})$$

And clearly if we make $|z|$ large enough then the ugly expression in the parentheses goes to 1. Hence there exists $r > 0$ such that when $|z| > r$ then $|f(z)| \geq \frac{|a_m|}{2}|z|^m$. And this proves that $|f(z)| \to \infty$ as $|z| \to \infty$ unless f is a constant polynomial.

We can prove a similar statement for analytic functions.

Liouville's Theorem: If f is a bounded entire (meaning analytic on all of $\mathbb{C}$) function, then f is a constant.

Proof:

Let $M < \infty$ be such that $|f(z)| \leq M$ for all $z \in \mathbb{C}$. By Cauchy's estimate, we know that $|f'(z)| \leq \frac{M}{r^n}$ for all $z \in \mathbb{C}$ and $r > 0$. In particular, by taking $r \to \infty$ we get that $f'(z) = 0$ for all $z \in \mathbb{C}$. Hence, f is constant. ■

Fundamental Theorem of Algebra: If f is a polynomial then either f is constant or f has at least one zero.

Proof:
Assume for the sake of contradiction that f is non-constant but has no zeros. Then $g = 1/f$ is an entire function. Also, since $|f(z)| \to \infty$ as $|z| \to \infty$, we know that g vanishes at infinity. But this implies g is bounded. So by Liouville's theorem, g must be a constant function. Yet that then implies f is constant (which is a contradiction). ■

I'll note that I think this proof is nicer then the one in my notes from last spring.

Before continuing on, here is a homework exercise that I didn't do two weeks ago but that will be used in the the proof of the next theorem.

Exercise III.3.17: Let G be a region and suppose $f : G \to \mathbb{C}$ is analytic such that $f(G)$ is a subset of a circle (not one going through ∞). Show that f is constant.

Suppose there exists $a \in \mathbb{C}$ and $R > 0$ such that $|f(z) - a|^2 = R^2$ for all z. Then after identifying $\mathbb{R}^2$ with $\mathbb{C}$ and setting $f(x+iy) - a := u(x,y) + iv(x,y)$, we have that $(u(x,y))^2 + (v(x,y))^2 = R^2$. Differentiating with respect to x and y then shows that:

$$2u\frac{\partial u}{\partial x} + 2v\frac{\partial v}{\partial x} = 0 = 2u\frac{\partial u}{\partial y} + 2v\frac{\partial v}{\partial y}$$

And by applying the Cauchy-Riemann equations and dividing by 2 we arrive at the following system of equations:

$$\begin{bmatrix} v(x,y) & -u(x,y) \\ u(x,y) & v(x,y) \end{bmatrix} \begin{bmatrix} \frac{\partial u}{\partial x}(x,y) \\ \frac{\partial v}{\partial x}(x,y) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

But now because $u^2 + v^2 = R^2 > 0$, we know the matrix $\begin{bmatrix} v(x,y) & -u(x,y) \\ u(x,y) & v(x,y) \end{bmatrix}$ always has full rank.

Hence, the only way our above system of equations can hold is if $\frac{\partial u}{\partial x}(x,y) = \frac{\partial v}{\partial x}(x,y) = 0$ for all $(x,y) \in \mathbb{R}^2$. And by the Cauchy-Riemann equations this proves that $(f-a)'(z) = 0$ for all $z \in G$. ■

Actually while I'm at it here is another miscellaneous result that may be useful to have in the back of my head just in case the professor asks me to prove the circle version of Cauchy's integral formula tomorrow.

If $|z| < 1$ then $\int_0^{2\pi} \frac{e^{it}}{e^{it} - z} \mathrm{d}t = 2\pi$.

Proof:
Let $F(t,s) := \frac{e^{it}}{e^{it} - sz}$ and $g(s) := \int_0^{2\pi} F(t,s)\mathrm{d}t$. Then if we can prove that $g(1) = 2\pi$.

Fortunately, note that $g(0) = 2\pi$. Also, by Leibniz's rule we have that:

$$g'(s) = \int_0^{2\pi} \frac{\partial F}{\partial s}(t,s)\mathrm{d}t = \int_0^{2\pi} \frac{ze^{it}}{(e^{it} - sz)^2}\mathrm{d}t.$$

But now note that $\frac{\partial}{\partial t}[iz(e^{it} - sz)^{-1}] = \frac{ze^{it}}{(e^{it} - sz)^2}$. Hence:

$$g'(s) = iz(e^{i2\pi} - sz)^{-1} - iz(e^{i0} - sz)^{-1} = 0.$$

This proves g is constant. So $g(1) = 2\pi$. ■

The Maximum Modulus Theorem: If G is a region (i.e. an open and connected set) and $f : G \to \mathbb{C}$ is an analytic function such that there is a point $a \in G$ with $|f(a)| \geq |f(z)|$ for all z in a neighborhood $U \subseteq G$ of a, then f is constant.

Proof:

Pick $r > 0$ such that $\overline{B_r(a)} \subseteq U$ and then define $\gamma(t) = a + re^{it}$. By Cauchy's integral formula, we have that:

$$f(a) = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{w-a}\mathrm{d}w = \frac{1}{2\pi i}\int_0^{2\pi} \frac{f(a+re^{it})\cdot ire^{it}}{a+re^{it}-a}\mathrm{d}t = \frac{1}{2\pi}\int_0^{2\pi} f(a+re^{it})\mathrm{d}t$$

In turn, because $|f(a)| \geq |f(a+re^{it})|$ for all t we get that:

$$|f(a)| \leq \frac{1}{2\pi}\int_0^{2\pi} |f(a+re^{it})|\mathrm{d}t \leq |f(a)|$$

This proves that $\int_0^{2\pi} |f(a)| - |f(a+re^{it})|\mathrm{d}t = 0$. And since $|f(a)| - |f(a+re^{it})| \geq 0$ for all t, this implies that $|f(a)| = |f(a+re^{it})|$ for all t.

But now note that r was arbitrary. So, we've actually proven that f maps some open ball $B_R(a) \subseteq U \subseteq G$ to the circle of radius $|f(a)|$ about the origin. And by exercise III.3.17 that implies that f is constant on $B_R(a)$. Finally, since $f(z) = f(a)$ for all z on a set with a limit point in G, we know that $f(z) = f(a)$ everywhere on G. ■

As a side note, this proves that if f is analytic on a bounded region G and continuous on $\overline{G}$, then f must attain it's maximum on ∂G.

Here is one other quick theorem to have in my back pocket I guess. (I have a stronger theorem written in my notes from last Spring but the proof of this statement is quicker).

(Conway) Proposition IV.2.15: Let f be analytic on $B_r(a)$ and suppose that γ is a closed piecewise C^1 curve in $B_r(a)$. Then $\int_\gamma f = 0$.

Proof:

Write $f(z) := \sum_{n=0}^{\infty} c_n(z-a)^n$. Then f has a radius of convergence of at least r. Also, suppose $F(z) := \sum_{n=0}^{\infty} \frac{c_n}{n+1}(z-a)^{n+1}$. Then since $\sqrt[n]{n+1} \to 1$ as $n \to \infty$, we have that F also has a radius of convergence of at least r. And as F is a primitive of f on $B_r(a)$, we are done. ■

Math 200a notes:

If G is a group, then $N_1 \supseteq N_2 \supseteq \cdots \supseteq N_k$ is called a central series if $N_i \lhd G$ for all i and $N_{i-1}/N_i \subseteq Z(G/N_i)$ for all $i \geq 2$.

Lemma: If $L, N \lhd G$ and $N \subseteq L$, then $L/N \subseteq Z(G/N)$ iff $[L, G] \subseteq N$.

($\Longrightarrow$)
As $[L, G] = \langle [y, x] : y \in L, x \in G \rangle$, to show the claim it suffices to argue why $[y, x] \in N$ for all $y \in L$ and $x \in G$. Fortunately, since $yN \in Z(G/N)$ for all $y \in L$, we have that $yxN = xyN$ for all $x \in G$ and $y \in L$. Or in other words, $[y, x]N = yxy^{-1}x^{-1}N = N$ for all $y \in L$ and $x \in G$.

($\Longleftarrow$)
Note that $yxN = xyN$ for all $y \in L$ and $x \in G$ if and only if $y^{-1}x^{-1}yx = [y^{-1}, x^{-1}] \in N$ for all $y \in L$ and $x \in G$. The latter will hold though since $[L, G] \subseteq N$.

Corollary: $N_1 \supseteq \cdots \supseteq N_c$ is a central series if and only if $N_i \lhd G$ and $[N_i, G] \subseteq N_{i+1}$ for all i.

Let G be a group.

- $\gamma_1(G) := G$ and $\gamma_{i+1}(G) := [\gamma_i(G), G]$ is called the lower central series.
- $Z_0(G) = \{1\}$ and $Z_{i+1}(G)/Z_i(G) := Z(G/Z_i(G))$ is called the upper central series.

  Side note: By correspondance theorem there is a unique normal subgroup $N \lhd G$ containing $Z_i(G)$ such that $N/Z_i(G) = Z(G/Z_i(G))$. So, this is an acceptable definition of $Z_{i+1}(G)$.

Note that by the lemma right above plus a lemma at the bottom of *page 378* $\{\gamma_i(G)\}$, we can easily see that $\{\gamma_i(g)\}$ is a central series (as i increases) and $\gamma_i(G)$ is a characteristic subgroup for all i.

Meanwhile, it's clear from the definition of a central series that $\{Z_i(G)\}$ (as i decreases) is a central series. And, I also claim that $Z_i(G)$ is a characteristic subgroup for all i.

$Z_0(G)$ is trivially a characteristic subgroup. So now we proceed by induction on i. To start off, given any $\theta \in \mathrm{Aut}(G)$, since $Z_i(G)$ is a characteristic subgroup we know that θ uniquely defines an automorphism $\overline{\theta}$ on $G/Z_i(G)$ by $\overline{\theta}(gZ_i(G)) := \theta(g)Z_i(G)$. This is well-defined because:

$$\begin{aligned} gZ_i(G) = hZ_i(G) \Longrightarrow gh^{-1} \in Z_i(G) &\Longrightarrow \theta(gh^{-1})Z_i(G) = Z_i(G) \\ &\Longrightarrow gZ_i(G) = hZ_i(G) \end{aligned}$$

In problem 4 of the sixth problem set I go into more detail on showing that $\overline{\theta}$ is an automorphism.

Next, note that if G' is any group then $Z(G')$ is a characteristic subgroup of G'.

This is because if $\phi \in \mathrm{Aut}(G')$, then $xy = yx$ for all $x \in G'$ implies that $\phi(x)\phi(y) = \phi(y)\phi(x)$ for all $x \in G'$. Hence $\phi(Z(G')) \subseteq Z(G')$. And by repeating the same reasoning with ϕ^{-1} we get that $\phi(Z(G')) = Z(G')$.

In particular, this means that $\overline{\theta}(Z(G/Z_i(G))) = Z(G/Z_i(G))$ for all $\theta \in \mathrm{Aut}(G)$. So if $Z_{i+1}(G)/Z_i(G) = Z(G/Z_i(G))$ then $\theta(x)Z_i(G) \in Z_{i+1}(G)/Z_i(G)$ for all $x \in Z_{i+1}(G)$ and $\theta \in \mathrm{Aut}(G)$. And in particular this means that $\theta(Z_{i+1}(G)) \subseteq Z_{i+1}(G)$ for all $\theta \in \mathrm{Aut}(G)$.

Since this reasoning holds for both θ and θ^{-1}, this proves $\theta(Z_{i+1}(G)) = Z_{i+1}(G)$ for all $\theta \in \mathrm{Aut}(G)$. Hence $Z_{i+1}(G)$ is a characteristic subgroup.

As a side note, by the same reasoning we can show that if N is a characteristic subgroup of G and L/N is a characteristic subgroup of G/N, then L is a characteristic subgroup of G.

Conversely, we can show that if N, L are characteristic subgroups of G and $N \subseteq L$, then L/N is a characteristic subgroup of G/N.

Theorem: Suppose G has a central series $G = N_1 \supseteq \cdots \supseteq N_{c+1} = \{1\}$. Then for all i we have that $\gamma_i(G) \subseteq N_i \subseteq Z_{c+1-i}(G)$.

Proof:

First, we proceed by induction on i to show that $\gamma_i(G) \subseteq N$. For our base case, note that $\gamma_1(G) = G = N_1$. Meanwhile, note that:

$$\gamma_i(G) \subseteq N_i \Longrightarrow \gamma_{i+1}(G) = [\gamma_i(G), G] \subseteq [N_i, G] \subseteq N_{i+1}.$$

Next we proceed by induction on j to show that $N_{c+1-j} \subseteq Z_j(G)$. For our base case, note that $Z_0(G) = \{1\} = N_{c+1-0}$. Meanwhile, suppose $N_{c+1-j} \subseteq Z_j(G)$ and consider the group homomorphism $\pi : G/N_{c+1-j} \to G/Z_j(G)$ by $xN_{c+1-j} \mapsto xZ_j(G)$. It is easy to check π is well-defined.

Now $\pi(Z(G/N_{c+1-j})) \subseteq Z(G/Z_j(G))$.

In fact, given any surjective group homomorphism $\phi : G \to G'$ we have that $\phi(Z(G)) \subseteq Z(G')$.

And since $Z_{j+1}(G)/Z_j(G) = Z(G/Z_j(G))$ and $N_{c-j}/N_{c-j+1} \subseteq Z(G/N_{c+1-j})$, this means that $\pi(N_{c-j}/N_{c+1-j}) \subseteq Z_{j+1}(G)/Z_j(G)$. So $N_{c-j} \subseteq Z_{j+1}(G)$. ■

Corollary: $Z_c(G) = G$ iff $\gamma_{1+c}(G) = \{1\}$.

$(\Longrightarrow)$

$\{1\} = Z_0(G) \subseteq \cdots \subseteq Z_c(G) = G$ being a central series implies that $\gamma_i(G) \subseteq Z_{c+1-i}(G)$ for all i. In particular, $\gamma_{c+1}(G) \subseteq Z_0(G) = \{1\}$.

$(\Longleftarrow)$

$G = \gamma_1(G) \supseteq \cdots \supseteq \gamma_{c+1}(G) = 1$ being a central series implies that $\gamma_i(G) \subseteq Z_{c+1-i}(G)$ for all i. In particular, $G = \gamma_1(G) \subseteq Z_c(G)$. ■

G is called nilpotent if there is a nonnegative integer c such that $\gamma_{1+c}(G) = \{1\}$. In that case we say the nilpotency class of G is c.

Corollary: G is nilpotent iff there exists a central series:

$$G = N_1 \supseteq N_2 \supseteq \cdots \supseteq N_{c+1} = \{1\}.$$

Lemma: Let G be a nilpotent group. Then $H \subsetneqq G \Longrightarrow H \subsetneqq N_G(H)$.

Proof:

Suppose $G = G_1 \supseteq G_2 \supseteq \cdots \supseteq G_{c+1} = \{1\}$ is a central series. Then since $H \neq G$, we know there exists i such that $G_{i+1} \subseteq H \subsetneqq G_i$. And, we claim that $G_i \subseteq N_G(H)$.

To see this, note for all $x \in G_i$ and $h \in H$ that $xhx^{-1}h^{-1} = [x,h] \in G_{i+1}$ as $[G, G_i] \subseteq G_{i+1}$. In particular, this means $[x,h] \in H$. So $[x,h]h = xhx^{-1} \in H$. And this proves that $xHx^{-1} \subseteq H$ for all $x \in G_i$.

Applying this same result to x^{-1}, we get that $x^{-1}Hx \subseteq H$. Hence $H \subseteq xHx^{-1}$ and this finishes the proof that $xHx^{-1} = H$ for all $x \in G_i$. So, $G_i \subseteq N_G(H)$. ■

Corollary:

- If G is a finite nilpotent group then for all $p \mid G$ we have that $|\mathrm{Syl}_p(G)| = 1$.
- If G is a finite nilpotent group then $G \cong \prod\limits_{\substack{p \mid |G| \\ P \in \mathrm{Syl}_p(G)}} P$.

For a proof of this, see *problems 2 and 3 on the fourth homework set*.

Proposition: Every finite p-group is nilpotent.

Proof:

We proceed by strong induction on $|G|$, noting that if $|G| = 1$ then the theorem is obvious.

Suppose $G \neq \{1\}$ is a finite p-group. Then we know that $Z(G) \neq \{1\}$. Hence $|G/Z(G)| < |G|$ and $G/Z(G)$ is a p-group. By induction we know that $G/Z(G)$ is nilpotent. So, there exists a central series $G/Z(G) = \overline{G_1} \supseteq \cdots \supseteq \overline{G_c} = \{Z(G)\}$.

By the correspondance theorem, for each i there exists a unique $G_i < G$ such that $\overline{G_i} = G_i/Z(G)$. And so, we can consider the series:

$$G = G_1 \supseteq G_2 \cdots \supseteq G_c = Z(G) \supseteq G_{c+1} = \{1\}.$$

Because $\{\overline{G_i}\}$ is a central series, we know $[\overline{G}, \overline{G_i}] \subseteq \overline{G_{i+1}}$ for all $i < c$. Also, by a slightly generalized argument to that of the second to last proposition on *page 379*, we can say that $[G, G_i]/Z(G) = [G/Z(G), G_i/Z(G)] = [\overline{G_1}, \overline{G_i}]$ for each $i < c$. Thus $[G, G_i]/Z(G) \subseteq G_{i+1}/Z(G)$ for all $i < c$. And this implies $[G, G_i] \subseteq G_{i+1}$ for all $i < c$.

Since we also know that $[G, Z(G)] = \{1\}$, we know that $[G, G_c] \subseteq G_{c+1}$. Therefore, $G = G_1 \supseteq \cdots \supseteq G_{c+1} = \{1\}$ is a central series of G. And that finishes the proof that G is nilpotent. ■

Lemma: If $G_1, \ldots, G_k$ is nilpotent then so is $\prod_{i=1}^k G_i$.

Proof:

$\gamma_j(\prod_{i=1}^k G_i) = \prod_{i=1}^k \gamma_j(G_i)$ for all j. To see why, note if $h = (h_1, \ldots, h_k) \in \gamma_j(\prod_{i=1}^k G_i)$ and $g = (g_1, \ldots, g_k) \in \prod_{i=1}^k$ then $[h, g] = ([h_1, g_1], \ldots, [h_k, g_k])$. Therefore, we can show that $\gamma_j(\prod_{i=1}^k G_i) = \prod_{i=1}^k \gamma_j(G_i) \Longrightarrow \gamma_{j+1}(\prod_{i=1}^k G_i) = \prod_{i=1}^k \gamma_{j+1}(G_i)$

But now we can just pick j large enough that $\gamma_j(G_i) = \{1\}$ for all i. And then we are done. ■

Theorem: Let G be a finite group. Then the following are equivalent.

1. G is nilpotent.
2. For all prime p dividing $|G|$, $|\mathrm{Syl}_p(G)| = 1$.
3. $G \cong \prod_{i=1}^{m} P_i$ where P_i is a finite p_i-group and each p_i is a prime.

One quick note is that if $G^{(i)}$ is the ith. derived subgroup of G, then $G^{(i)} < \gamma_{i+1}(G)$. To see why, note that $G^{(0)} = G = \gamma_1(G)$. Then if $G^{(i)} < \gamma_{i+1}(G)$ we have that:

$$G^{(i+1)} = [G^{(i)}, G^{(i)}] < [G^{(i)}, G] < [\gamma_i(G), G] = \gamma_{i+1}(G).$$

This shows that all nilpotent groups are solvable. The converse is not true though.

Claim: Suppose G is a group with $Z(G) = \{1\}$. Then G is not nilpotent.

Proof:
If G is nilpotent, then $\{Z_i(G)\}$ must be a strictly increasing central series such that $\{1\} = Z_0(G) \subseteq Z_1(G) \subseteq \cdots \subseteq Z_c(G) = G$. But $Z_1(G) = Z(G)$. So we know that $Z(G) \neq \{1\}$. ■

Consequently, we have that S_3 is not nilpotent. Yet by problem 2(c) on the sixth problem set, we do know that S_3 is solvable.

Let $\mathrm{Max}(G) := \{M : M \lneq G \text{ and } \nexists H \lneq G \text{ with } M < H\}$. The subgroups in $\mathrm{Max}(G)$ are called maximal.

Note that $\theta(M) \in \mathrm{Max}(G)$ if $M \in \mathrm{Max}(G)$ and $\theta \in \mathrm{Aut}(G)$ (this is just an application of the correspondance theorem). In particular, this measn $\bigcap_{M \in \mathrm{Max}(G)} M$ is a characteristic subgroup.

We call $\Phi(G) := \bigcap_{M \in \mathrm{Max}(G)} M$ the Frattini subgroup.

Theorem: Suppose G is a finite group. Then G is nilpotent if and only if all maximal subgroups are normal.

$(\Longrightarrow)$
If $M \in \mathrm{Max}(G)$ then we know that $M \lneq G$. And since G is nilpotent we know that $M \lneq N_G(M)$. But now because M is maximal, we know that $N_G(M) = G$. So $M \lhd G$.

$(\Longleftarrow)$
Let $P \in \mathrm{Syl}_p(G)$ and suppose $N_G(P) \neq G$. Then there would exist $M \in \mathrm{Max}(G)$ such that $N_G(P) \subseteq M$. And since $M \lhd G$ and $P \in \mathrm{Syl}_p(M)$, we'd know by Frattini's trick (see *problem 2 on the third problem set*) that $MN_G(P) = G$. However, as $N_G(P) \subseteq M$ and $M \neq G$, this is a contradiction.

So, we conclude all Sylow p-subgroups of G are normal for any prime p dividing $|G|$. This proves that G is nilpotent. ■

$(M, \cdot)$ is called a monoid if $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in M$ and there exists $e \in M$ (called the netural (or identity) element) such that $e \cdot a = a = a \cdot e$ for all a.

Given a set $X \neq \emptyset$ (our alphabet), we define the language in our alphabet X to be the collection $\mathcal{L}(X)$ of formal strings of elements of X. Concatenation of strings turns $\mathcal{L}(X)$ into a monoid.

We claim $\mathcal{L}(X)$ is a free monoid generated by X.

What does it mean to say $C(X)$ is a free $*$ generated by X?

Let $*$ be a placeholder for the name of some algebraic structure (like a ring, group, monoid, etc). Then saying $C(X)$ is a free $*$ generated by X is equivalent to saying that:

1. there is an embedding $i : X \hookrightarrow C(X)$;
2. if M is a $*$ and $f : X \to M$ is any function, then there is a unique $*$-homomorphism $\hat{f} : C(X) \to M$ such that the below diagram commutes:

$$\begin{tikzcd} & C(X) \arrow[d, "\hat{f}"] \\ X \arrow[ur, hook, "i"] \arrow[r, "f"'] & M \end{tikzcd}$$

Equivalently, if $\mathcal{F}$ denotes the forgetful functor (i.e. if M has some algebraic structure then $\mathcal{F}(M)$ is just the set M without it's structure), then we are saying that $\mathrm{Hom}(X, \mathcal{F}(M)) \cong \mathrm{Hom}(C(X), M)$.

To see why $\mathcal{L}(X)$ is a free monoid generated by X:

Let $i : X \hookrightarrow \mathcal{L}(X)$ be given by $i(x) \mapsto x$. Then given any $f : X \to M$ define $\widehat{f}(\emptyset) = e_M$ (where $\emptyset$ denotes the string of length zero and e_M is the neutral element of M) and $\widehat{f}(x_1 \cdots x_n) = f(x_1) \cdots f(x_n)$. Note $\widehat{f}$ is forced to be that since $\widehat{f}$ is a monoid homomorphism with $\widehat{f}(x) = f(x)$ for all $x \in X$.

Let $\{G_i\}_{i \in I}$ be a family of groups. By modifying each G_i by some tagging process (i.e. replacing G_i with $\{i\} \times G_i$) we can assume each G_i is disjoint. We say a group P is a free product of $\{G_i\}_{i \in I}$ if:

1. for each $i \in I$ there is an injective group homomorphism $j_i : G_i \hookrightarrow P$;
2. For every group G and family of homomorphisms $\{f_i : G_i \to G\}_{i \in I}$, there is a unique homomorphism $\widehat{f} : P \to G$ such that the below diagram commutes for each $i \in I$:

$$\begin{tikzcd} & P \arrow[d, "\widehat{f}"] \\ X \arrow[ur, hook, "j_i"] \arrow[r, "f_i"'] & G \end{tikzcd}$$

Note that if $*$ is a placeholder for the name of any other algebraic structure, then we can similarly define what it means to be a free product for that structure. That said, for now I'll avoid going into too much detail since I've heard math 200b will focus a lot more on actual category theory.

One observation I'd like to make is that if P and Q are both free products of $\{G_i\}_{i\in I}$ then $P \cong Q$. Hence, it makes sense to talk about *the* free product of a collection of groups.

To see this, let us denote each of the embeddings $G_i \hookrightarrow P$ and $G_i \hookrightarrow Q$ which define our free products as $j_i^{(p)}$ and $j_i^{(q)}$ respectively. Then importantly, there are unique group homomorphisms $\widehat{f} : P \to Q$ and $\widehat{g} : Q \to P$ such that $\widehat{f} \circ j_i^{(p)} = j_i^{(q)}$ and $\widehat{g} \circ j_i^{(q)} = j_i^{(p)}$ for all i.

$$\begin{tikzcd} & P \arrow[d, "\widehat{f}"] \\ G_i \arrow[ur, "j_i^{(p)}"] \arrow[r, hook, "j_i^{(q)}"'] & Q \end{tikzcd} \qquad \begin{tikzcd} & Q \arrow[d, "\widehat{g}"] \\ G_i \arrow[ur, "j_i^{(q)}"] \arrow[r, hook, "j_i^{(p)}"'] & P \end{tikzcd}$$

But now in paricular we have that $\widehat{f} \circ (\widehat{g} \circ j_i^{(q)}) = j_i^{(q)}$ and $\widehat{g} \circ (\widehat{f} \circ j_i^{(p)}) = j_i^{(p)}$ for all i. So, $\widehat{f} \circ \widehat{g}$ and $\widehat{g} \circ \widehat{f}$ are the unique group homomorphisms such that the below diagrams commute for all i:

$$\begin{tikzcd} & P \arrow[d, "\widehat{g}\circ\widehat{f}"] \\ G_i \arrow[ur, "j_i^{(p)}"] \arrow[r, hook, "j_i^{(p)}"'] & P \end{tikzcd} \qquad \begin{tikzcd} & Q \arrow[d, "\widehat{f}\circ\widehat{g}"] \\ G_i \arrow[ur, "j_i^{(q)}"] \arrow[r, hook, "j_i^{(q)}"'] & Q \end{tikzcd}$$

And since Id satisfies the same property, we must have by uniqueness that $\widehat{f} \circ \widehat{g} = \mathrm{Id}$ and $\widehat{g} \circ \widehat{f} = \mathrm{Id}$. Hence $\widehat{f}$ and $\widehat{g}$ are isomorphisms with $\widehat{f}^{-1} = \widehat{g}$.

Note that similar reasoning shows that a free $*$ generated by X is unique up to $*$-isomorphism.

Our goal now is to give a somewhat easy to work with construction of the free product of $\{G_i\}_{i\in I}$. To start off, we'll take $X = \bigcup_{i\in I} G_i$ and let $\mathcal{L}(X)$ be the language in alphabet X.

Note that if $a \in G_i$, then $a^2 \in \mathcal{L}(X)$ is a word with 1 letter whereas $aa \in \mathcal{L}(X)$ is a word with 2 letters. It would be nice if we can identify those two words with each other. So, we define an equivalence relation $\sim$ such that:

- if $a, b \in G_i$ with $ab = c$ (in G_i) and also ω_1, ω_2 are words in $\mathcal{L}(X)$, then $\omega_1 ab\omega_2 \sim \omega_1 c\omega_2$ in $\mathcal{L}(X)$;
- if e_i denotes the neutral element of G_i, then we want $\omega_1 e_i \omega_2 \sim \omega_1\omega_2$;

- $\omega \sim \omega'$ if and only if you can apply the prior two properties finitely many times to go from ω to ω'. (You can easily show that this is a well-defined equivalence relation.)

Note that it is pretty easy to see that if $\omega \sim \omega'$ then $g\omega \sim g\omega'$ and $\omega g \sim \omega' g$ for all $g \in X$. Hence, we get the following result:

Lemma: Suppose $\omega_1 \sim \omega_1'$ and $\omega_2 \sim \omega_2'$. Then $\omega_1\omega_2 \sim \omega_1'\omega_2'$.

Why? $\omega_1 \sim \omega_1' \implies \omega_1\omega_2 \sim \omega_1'\omega_2$ and $\omega_2 \sim \omega_2' \implies \omega_1'\omega_2 \sim \omega_1'\omega_2'$. Thus $\omega_1\omega_2 \sim \omega_1'\omega_2'$ by transitivity.

Let $\mathcal{F}(X) := \mathcal{L}(X)/_\sim$. Then our prior lemma tells us that $[\omega]\cdot[\omega'] := [\omega\omega']$ is a well-defined operation.

Claim: $(\mathcal{F}(X), \cdot)$ is a group.

Proof:

Associativity of $\cdot$ is because concatenation is associative. Also $[\emptyset]$ is the identity. And finally, if $\omega = x_1 \cdots x_n$ then $[\omega]^{-1} = [x_n^{-1} \cdots x_1^{-1}]$ is the inverse of $[\omega]$.

And finally, we claim that $\mathcal{F}(X)$ is a free product of $\{G_i\}_{i\in I}$.

To start off, note that if $\{f_i : G_i \to G\}_{i\in I}$ is a family of group homomorphisms, then we can define $h : X \to G$ to be the function with $g \mapsto f_i(g)$ for all $i \in I$ and $g \in G_i$. Then in turn, there is a unique monoid homomorphism $h' : \mathcal{L}(X) \to G$ with $h'(g) = h(g)$ for all $g \in X$.

Importantly h' is constant over equivalence classes of $\sim$.

To prove this it suffices to argue that $h'(\omega_1 ab\omega_2) = h'(\omega_1 c\omega_2)$ if $ab = c$ in G_k for some $k \in I$, and similarly that $h'(\omega_1 e\omega_2) = h'(\omega_1\omega_2)$ if e is the neutral element in G_k for some $k \in I$. Luckily:

- $$\begin{aligned} h'(\omega_1 ab\omega_2) = h'(\omega_1)h'(a)h'(b)h'(\omega_2) &= h'(\omega_1)f_k(a)f_k(b)h'(\omega_2) \\ &= h'(\omega_1)f_k(ab)h'(\omega_2) \\ &= h'(\omega_1)h'(c)h'(\omega_2) = h'(\omega_1 c\omega_2) \end{aligned}$$

- $$\begin{aligned} h'(\omega_1 e\omega_2) = h'(\omega_1)h'(e)h'(\omega_2) &= h'(\omega_1)f_k'(e)h'(\omega_2) \\ &= h'(\omega_1)h'(\omega_2) = h'(\omega_1\omega_2). \end{aligned}$$

Thus, h' induces a well-defined map $\mathcal{F}(X) \to G$ such that $[g] \mapsto f_i(g)$ for all $i \in I$ and $g \in G_i$. And clearly no other map can have this property.

Now if we define $j_k : G_k \to \mathcal{F}(X)$ by $j_k(g) = [g]$ for all $k \in I$, then all that's left to do in order to prove $\mathcal{F}(X)$ is a free product of $\{G_i\}_{i\in I}$ is show that each j_k is in fact injective. Fortunately, the prior reasoning let's us prove that $\mathcal{F}(X)$ has enough distinct equivalence classes so that $[g] \neq [g']$ when $g \neq g'$ in G_i.

Specifically, given any fixed $k \in I$ let $f_i : G_i \to G_k$ be the trivial homomorphism for all $i \neq k$ and let $f_k : G_k \to G_k$ be the identity. Then there is a homomorphism $h' : \mathcal{F}(X) \to G_k$ such that $h'([g]) = g$ for all $g \in G_k$. And in particular since $h'([g]) \neq h'([g'])$ when $g \neq g'$ in G_k, we are done. ■

We typically denote $\mathcal{F}(X)$ as $*_{i\in I} G_i$. Furthermore, $\mathcal{F}(X)$ is also called the amalgam product of the G_i. And in category theory contexts it is called the coproduct of the G_i (denoted $\coprod_{i\in I} G_i$).

By the way I'm incredibly grateful to Hagan for giving me their lecture notes for 11/7 lecture since I was too sick to attend.

Math 200a Homework

Set 6 Problem 1: Let G be a group.

(a) Prove that $[G,G]$ is a characteristic subgroup.

For all $\phi \in \mathrm{Aut}(G)$ we have that $\phi([g_1,g_2]) = [\phi(g_1),\phi(g_2)] \in [G,G]$. Therefore, we can easily see $\phi([G,G]) \subseteq [G,G]$. Also, by applying that reasoning to ϕ^{-1} we get that $\phi^{-1}([G,G]) \subseteq [G,G]$. Thus $\phi(\phi^{-1}([G,G])) = [G,G] \subseteq \phi([G,G])$.

(b) Prove that for a normal subgroup N of G, G/N is abelian precisely when $[G,G] \subseteq N$.

We have that $ghN = hgN$ iff $g^{-1}h^{-1}gh = [g^{-1},h^{-1}] \in N$. Thus, G/N is abelian precisely when $[g^{-1},h^{-1}] \in N$ for all $g,h \in G$. But that happens if and only if $[G,G] \subseteq N$.

(c) Prove that $[S_n,S_n] = A_n$ for every integer $n \geq 3$.

To start off, we know that $[\sigma,\tau] \in A_n$ for all $\sigma,\tau \in S_n$. This is because:

$$\mathrm{sgn}([\sigma,\tau]) \equiv 2\mathrm{sgn}(\sigma) + 2\mathrm{sgn}(\tau) \equiv 0 \pmod 2.$$

As a result of that plus part (a), we know that $[S_n,S_n] \lhd A_n$. At the same time, note that $[(1\ \ 2),(2\ \ 3)] = (1\ \ 2)(2\ \ 3)(1\ \ 2)(2\ \ 3) = (1\ \ 3\ \ 2)$. Therefore, since $[S_n,S_n] \lhd A_n$ and $[S_n,S_n]$ has a 3-cycle, we have that $[S_n,S_n] = A_n$.

Set 6 Problem 2: Suppose $n \geq 5$ and $m \geq 2$ are integers.

(a) Find the composition factors of S_n.

We know that $S_n \rhd A_n \rhd \{\mathrm{Id}\}$ is a composition series for S_n. After all, $S_n/A_n \cong C_2$ (the cyclic group of order 2) since C_2 is the only group of order 2. Meanwhile $A_n/\{\mathrm{Id}\} \cong A_n$. And both C_2 and A_n are simple.

It follows that C_2 and A_n are the composition factors of S_n.

(b) Prove that if N is a non-trivial proper normal subgroup of S_n then $N = A_n$.

Suppose $N \lhd S_n$ and $N \neq A_n$. Then $N \cap A_n \lhd A_n$. And since A_n is simple we know that either $N \cap A_n = \{\mathrm{Id}\}$ or $N \cap A_n = A_n$. Suppose the latter is true. Then we know that $A_n \subsetneq N < S_n$. But now $[S_n : N] < [S_n : A_n] = 2$, this must mean that $S_n = N$.

Meanwhile, suppose $N \cap A_n = \{\mathrm{Id}\}$. Then for the sake of contradiction suppose $N \neq \{\mathrm{Id}\}$. We must have that $N - \{\mathrm{Id}\}$ consists entirely of odd permutations. And since the composition of two odd permutations is even, this also means that for any $\sigma, \tau \in N - \{\mathrm{Id}\}$ we must have that $\sigma\tau = \mathrm{Id}$. But also note that because N is closed under conjugation, we have for any $\sigma \in N - \{\mathrm{Id}\}$ that every permutation with the same cycle type as σ is in N. Also, every $\sigma \in N - \{\mathrm{Id}\}$ must have order 2 and thus it's cycle decomposition must consist of only 2-cycles. This now let's us split into the following two cases for when $N \neq \{\mathrm{Id}\}$.

Suppose the transposition $(a\ \ b) \in N$. Then let c be distinct from a and b and note $(b\ \ c) \in N$. But now we have a contradiction since $(a\ \ b)(b\ \ c) = (a\ \ b\ \ c) \in N$ and $(a\ \ b\ \ c) \neq \mathrm{Id}$. Meanwhile suppose $\sigma = (a\ \ b)(c\ \ d)(e\ \ f)\cdots \in N$. Then $\tau = (a\ \ c)(b\ \ d)(e\ \ f)\cdots \in N$ as well. And now $\tau\sigma(a) = d \neq a$. This contradicts that $\tau\sigma = \mathrm{Id}$.

(c) Find out for what values of m is S_m solvable.

If $m \geq 5$ then we know that S_m isn't solvable. This is because S_m is solvable iff all it's composition factors are cyclic groups of prime order. That said, A_m is a composition factor of S_m (when $m \geq 5$) that isn't a group with prime order (not to mention that it's not cyclic either).

Meanwhile if $m = 4$, then S_4 is solvable. To see this, we first find a normal subgroup of A_4.

Consider $P \in \mathrm{Syl}_2(A_4)$. Then the order of every element of P must be a power of 2. Yet at the same time, if $o(\sigma) = 4$ then we know that $\sigma \in S_4$ is a 4-cycle and thus odd. Hence, $P - \{\mathrm{Id}\}$ contains only even permutations of order 2. Since we also know that P has 4 elements, this actually guarentees that $P = \{\mathrm{Id}, (1\ \ 2)(3\ \ 4), (1\ \ 3)(2\ \ 4), (1\ \ 4)(2\ \ 3)\}$. And since we've shown P is uniquely determined, we know that $P \lhd A_4$.

Going a step further, we can easily verify that if $Q = \{\mathrm{Id}, (1\ \ 2)(3\ \ 4)\}$ then $Q \lhd P$. So, at last we have the normal series $S_4 \rhd A_4 \rhd P \rhd Q \rhd \{\mathrm{Id}\}$. And since $[S_4 : A_4] = 2$, $[A_4 : P] = 3$, $[P : Q] = 2$, and $[Q : \{\mathrm{Id}\}] = 2$, we know that every composition factor of S_4 is a cyclic group of prime order.

Similarly, if $m = 3$ then we have that S_3 is solvable. This is because $|A_3| = 6/2 = 3$. Hence, A_3 is a simple group since it is cyclic group with prime order 3. And in turn, S_3 has the composition factors $S_3/A_3 \cong C_2$ and $A_3/\{\mathrm{Id}\} \cong C_3$.

Finally, if $m = 2$ then we have that S_2 is cyclic with order 2. Hence, S_2 is abelian and we trivially have that S_2 is solvable.

Set 6 Problem 3: Suppose the following is an S.E.S.: $\{1\} \to G_1 \to G_2 \to G_3 \to \{1\}$. Prove that G_2 is solvable iff G_1 and G_3 are solvable.

Recall on *page 379* that we showed that if $\varphi : G \to G'$ is a surjective group homomorphism then $\varphi(G^{(i)}) = (G')^{(i)}$.

For the grader: Note that $\varphi(G^{(0)}) = (G')^{(0)}$ because φ is surjective. Next, suppose we have already shown that $\varphi(G^{(i)}) = (G')^{(i)}$. Then because $\varphi([g_1, g_2]) = [\varphi(g_1), \varphi(g_2)]$ for all $g_1, g_2 \in G^{(i)}$, we have that:

$$\varphi(G^{(i+1)}) = \varphi([G^{(i)}, G^{(i)}]) \subseteq [(G')^{(i)}, (G')^{(i)}] = (G')^{(i+1)}.$$

Meanwhile, the fact that $(G')^{(i+1)} \subseteq \varphi(G^{(i+1)})$ is because for every $h_1, h_2 \in (G')^{(i)}$ there exists $g_1, g_2 \in G^{(i)}$ with $\varphi([g_1, g_2]) = [h_1, h_2]$.

It follows that if G and G' are isomorphic then G is solvable iff G' is solvable. And in particular, this means that if $\{1\} \to H_1 \to H_2 \to H_3 \to \{1\}$ is an S.E.S. that is isomorphic to $\{1\} \to G_1 \to G_2 \to G_3 \to \{1\}$, then G_i is solvable iff H_i is solvable for all i. For this reason, we can without loss of generality now assume we are working with a standard S.E.S. $\{1\} \to N \to G \to G/N \to \{1\}$ where $N \lhd G$.

$(\Longrightarrow)$
We already showed this direction on *page 379*.

$(\Longleftarrow)$
Suppose N and G/N are both solvable and then let $N = K_0 \rhd K_1 \rhd \cdots \rhd K_\ell = \{\mathrm{Id}\}$ and $G/N = \overline{H}_0 \rhd \overline{H}_1 \rhd \cdots \rhd \overline{H}_m = \{1\}$ be composition series for N and G/N. By the correspondance theorem, we can find $G = H_0 > H_1 > \cdots > H_m = N$ such that $H_i/N = \overline{H}_i$ for all i. Furthermore, we then know also by the correspondence theorem that $H_i \lhd H_{i-1}$ for all i since $H_i/N = \overline{H}_i \lhd \overline{H}_{i-1} = H_{i-1}/N$ for all i. And finally, note by the third isormophism theorem that $H_{i-1}/H_i \cong (H_{i-1}/N)/(H_i/N) = \overline{H}_{i-1}/\overline{H}_i$.

Hence, $G = H_0 \rhd H_1 \rhd \cdots \rhd H_{m-1} \rhd H_m = K_0 \rhd K_1 \rhd \cdots \rhd K_\ell = \{1\}$ is a normal series of G such that every factor H_{i-1}/H_i and K_{i-1}/K_i is a cyclic group of prime order. It follows that G is solvable.

Set 6 Problem 4: Prove there is no finite group G such that $[G, G] \cong S_4$.
Suppose to the contrary that there exists a finite group G with $[G, G] \cong S_4$. Then recall from my work on problem 2 of this problem set that:
$P := \{\mathrm{Id}, (1\ \ 2)(3\ \ 4), (1\ \ 3)(2\ \ 4), (1\ \ 4)(2\ \ 3)\}$ is the unique Sylow 2-subgroup of A_4.

It follows that P is a characteristic subgroup of A_4. Also note that since all automorphisms of S_4 take transpositions to transpositions, we have that A_4 is a characteristic subgroup of S_4. And in turn, we can conclude that P is a characteristic subgroup of S_4.

Now consider the action $G \curvearrowright [G, G]$ by conjugation. This gives a group homomorphism $\phi : G \to \mathrm{Aut}([G, G])$. And if $\eta : [G, G] \to S_4$ is an isomorphism, we know there exists a group homomorphism $\phi' : G \to \mathrm{Aut}(S_4)$ given by $\phi'(g) = \eta \circ (\phi(g)) \circ \eta^{-1}$.

But note that if $\theta \in \mathrm{Aut}(S_4)$ then there is a well-defined automorphism $\overline{\theta}$ on A_4/P given by $\overline{\theta}(\sigma P) \mapsto \theta(P)$.

To see that $\overline{\theta}$ is well defined, suppose $\sigma_1 P = \sigma_2 P$. Then $\sigma_1 \sigma 2^{-1} \in P$, and because P is a characteristic subgroup we have that $\theta(\sigma_1)\theta(\sigma_2)^{-1} = \theta(\sigma_1\sigma_2^{-1}) \in P$ as well. Hence, we can conclude that $\theta(\sigma_1)P = \theta(\sigma_2)P$. And since A_4 is a characteristic subgroup of S_4, we know that $\theta(\sigma_i) \in A_4$ for each i. This finishes the proof that $\overline{\theta}$ is a well-defined function from A_4/P to A_4/P

To see that it's injective, suppose $\overline{\theta}(\sigma P) = P$. Then $\theta(\sigma)P = P$ (or equivalently $\theta(\sigma) \in P$.) And because P is a characteristic subgroup of S_4, we in turn know that $\sigma \in P$. So our proposed automorphism $\overline{\theta}$ has a trivial kernel.

Finally, by a simple cardinality argument we know $\overline{\theta}$ is surjective.

Furthermore $\overline{\theta_1 \circ \theta_2}(\sigma P) = (\theta_1 \circ \theta_2(\sigma))P = \overline{\theta_1}(\theta_2(\sigma)P) = \overline{\theta_1}(\overline{\theta_2}(\sigma P))$. Thus, the map $\theta \mapsto \overline{\theta}$ is a group homomorphism. And we've in turn proven that there exists a group homomorphism $\varphi : G \to \mathrm{Aut}(A_4/P)$.

However, A_4/P is a cyclic group of order 3 and in turn $\mathrm{Aut}(A_4/P)$ is a cyclic group of order 2. This shows that $\mathrm{Aut}(A_4/P)$ is abelian. And since φ is a group homomorphism from G to an abelian group, we must have that $[G, G] \subseteq \ker(\varphi)$.

Why?

Suppose $\varphi(g_1)\varphi(g_2) = \varphi(g_2)\varphi(g_1)$ for all $g_1, g_2 \in G$. Then:

$$\varphi([g_1, g_2]) = \varphi(g_1 g_2 g_1^{-1} g_2^{-1}) = 1 \text{ for all } g_1, g_2 \in G.$$

This suggests the action of $[G, G] \curvearrowright A_4/P$ given $g \cdot (\sigma P) = \eta(g)\sigma(\eta(g))^{-1}$ is trivial. And since $\eta([G, G]) = S_4$, this is the same as saying the action $S_4 \curvearrowright A_4/P$ by conjugation is trivial.

This is a contradiction though as $(1\ \ 2)(1\ \ 2\ \ 3)(1\ \ 2)P = (1\ \ 3\ \ 2)P \neq (1\ \ 2\ \ 3)P$. To see this, just note that $(1\ \ 3\ \ 2)(1\ \ 2\ \ 3)^{-1} = (1\ \ 3\ \ 2)(1\ \ 3\ \ 2) = (1\ \ 2\ \ 3) \notin P$.

Set 6 Problem 5: Prove that the collection of functions $D_\infty \coloneqq \{ax + b : a \in \{\pm 1\}, b \in \mathbb{Z}\}$ under composition is an infinite solvable group which is generated by two elements of order 2. Find the center $Z(D_\infty)$ of D_∞.

First we show that D_∞ is a group.

To start off, if $a_1 x + b_1$ and $a_2 x + b_2$ are functions in D_∞, then $a_1 a_2 \in \{\pm 1\}$ and $a_1 b_2 + b_1 \in \mathbb{Z}$ hence $a_1(a_2 x + b_2) + b_1 \in D_\infty$. So the composition operator is well-defined.

Associativity holds because function composition is associative. Also, $x \in D_\infty$ is the identity element. And if $ax + b \in D_\infty$ then $ax - ab \in D_\infty$ is its inverse. To see this, note that $a^2 = 1$. So:

$$a(ax - ab) + b = x - b + b = x \text{ and } a(ax + b) - ab = x + ab - ab = x$$

Next we show that D_∞ is solvable. Let $N \coloneqq \{x + b : b \in \mathbb{Z}\} \subseteq D_\infty$. It's easy to see that N is a subgroup of D_∞ and that $N \cong (\mathbb{Z}, +)$ (and is abelian). I also claim that N is normal.

To see why, note that:

$$(ax + b) \circ (x + c) \circ (ax + b)^{-1} = a((ax - ab) + c) + b = x - b + ac + b = x + ac.$$

Furthermore, it is easy to see that the coset $(-x) \circ N$ contains all of $D_\infty - N$. So, $D_\infty \rhd N \rhd \{\mathrm{Id}\}$ is a normal series such that D_∞/N has order 2 (and is thus cyclic and abelian) and $N/\{\mathrm{Id}\} \cong N \cong (\mathbb{Z}, +)$ is abelian. Hence, D_∞ is solvable.

Thirdly, we show D_∞ is generated by two elements of order 2.

Let $r(x) = -x$ and $s(x) = -x + 1$. Then $r(r(x)) = -(-x) = x$ and $s(s(x)) = -(-x + 1) + 1 = x - 1 + 1 = x$. Hence, both r and s have order 2.

That said, note that $(r \circ s)(ax + b) = -(-(ax + b) + 1) = ax + (b - 1)$. And similarly $(r \circ s)^{-1}(ax + b) = (s \circ r)(ax + b) = -(-(ax + b)) + 1 = ax + (b + 1)$. Hence, we can show by induction that for all $ax + b \in D_\infty$ we have that:

$$(r \circ s)^b \circ (ax + b) = ax$$

If $a = 1$ this implies that $(r \circ s)^{-b} = ax + b$. Meanwhile, if $a = -1$ then this implies that $r \circ (r \circ s)^b \circ (ax + b) = x$. So $(r \circ s)^{-b} \circ r = (ax + b)$. Either way, this shows that for any $ax + b \in D_\infty$ there is some way of composing r and s together to get that element. So $\langle r, s \rangle = D_\infty$.

Finally, we show that D_∞ has a trivial center. Note that $(cx + d) \in Z(D_\infty)$ if and only if $cx + (b + ad - bc) = (ax + b) \circ (cx + d) \circ (ax - ab) = cx + d$ for all $ax + b \in D_\infty$. Or in other words $cx + d \in Z(D_\infty)$ iff $b + ad - bc = d$ for all $a \in \{\pm 1\}$ and $b \in \mathbb{Z}$.

But now if $c = -1$, we must have that $2b + d = d$ for all $b \in \mathbb{Z}$. This is clearly impossible. Meanwhile if $c = 1$ then $b + ad - bc = ad$ for all b. And now $ad = d$ for all $a \in \{\pm 1\}$ iff $d = 0$. Hence, $cx + d \in Z(D_\infty)$ iff $c = 1$ and $d = 0$.

Math 220a lecture notes:

For a closed piecewise C^1 path γ, the index / winding number of γ with respect to $a \in \mathbb{C} - \{\gamma\}$ is:

$$n(\gamma; a) = \frac{1}{2\pi i} \int_\gamma \frac{\mathrm{d}w}{w - a}.$$

I think Rudin does a cleaner job here of proving that $n(\gamma; a)$ is an integer then Conway does. So if you want to see a proof of that fact look at my notes from last spring.

Why do we call it the winding number?

Suppose $\gamma : [0, 1] \to \mathbb{C}$ is a closed curve and that $0 \in \mathbb{C} - \{\gamma\}$. (Note that after reparametrizing and maybe translating γ we can always guarentee this). Then we can find a partition $a = t_0 < t_1 < \cdots < t_n = b$ such that $z = \gamma(t)$ lies in a sector with $\arg(z) \in [\theta_{j-1}, \theta_j]$ for all $t \in [t_{j-1}, j]$ (where $|\theta_j - \theta_{j-1}| < 2\pi$). And having done that, we define a branch of the logarithm for each sector by $\log_j(z) = \log|z| + i\arg(z)$ (where $\arg(z) \in [\theta_{j-1}, \theta_h]$).

Now:

$$\begin{aligned}\int_0^1 \frac{\gamma'(s)}{\gamma(s)-0}\mathrm{d}s = \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \frac{\gamma'(s)}{\gamma(s)}\mathrm{d}s &= \sum_{j=1}^n (\log_j(\gamma(t_j)) - \log_j(\gamma(t_{j-1})))\\ &= \sum_{j=1}^n (\log|\gamma(t_j)| + i\arg(\gamma(t_j)) - \log|\gamma(t_{j-1})| - i\arg(\gamma(t_{j-1})))\\ &= i\sum_{j=1}^n (\arg(\gamma(t_j)) - \arg(\gamma(t_{j-1})))\end{aligned}$$

Therefore the winding number $n(\gamma; w)$ calculates the number of times the curve γ loops around w (keeping track of which direction).

A hopefully clear corollary of this is the following:

(Conway) Proposition 4.3:

(a) If γ is a piecewise C^1 closed curve and $-\gamma$ is a reparametrization of γ traveling in the reverse direction, then $n(\gamma;a) = -n(-\gamma;a)$ for all $a \notin \{\gamma\}$.

(b) Suppose γ and ω are piecewise C^1 closed curves rooted at the same point and let $\gamma + \omega$ denote the concatenation of the two curves. Then for any $a \notin \{\gamma\} \cup \{\omega\}$ we have that $n(\gamma+\omega;a) = n(\gamma;a) + n(\omega;a)$.

Math 220a Homework:

Exercise IV.3.1: Let f be an entire function (i.e. analytic on all $\mathbb{C}$) and suppose there are constants $M, R > 0$ and an integer $n \geq 1$ with $|f(z)| \leq M|z|^n$ for all $|z| > R$. Then f is a polynomial of degree $\leq n$.

Recall that for any $a \in \mathbb{C}$, if $\gamma_r(t) := a + re^{it}$ then $f^{(n+1)}(a) = \frac{(n+1)!}{2\pi i}\int_{\gamma_r} \frac{f(w)}{(w-a)^{n+2}}\mathrm{d}w$.

In particular, if r is large enough so that $|\gamma_r(t)| > R$ for all t, then:

$$\begin{aligned}|f^{(n+1)}(a)| = \tfrac{(n+1)!}{2\pi}\left|\int_{\gamma_r} \tfrac{f(w)}{(w-a)^{n+2}}\mathrm{d}w\right| &\leq \tfrac{(n+1)!}{2\pi}\int_0^{2\pi}\left|\tfrac{|f(a+re^{it})|}{(re^{it})^{n+2}} \cdot ire^{it}\right|\mathrm{d}t \\ &\leq \tfrac{(n+1)!}{2\pi}\int_0^{2\pi}\tfrac{M|a+re^{it}|^n}{r^{n+2}}r\mathrm{d}t \\ &\leq \tfrac{(n+1)!}{2\pi}\cdot\tfrac{M(|a|+r)^n}{r^{n+1}}\int_0^{2\pi}\mathrm{d}t = M(n+1)!\cdot\tfrac{(|a|+r)^n}{r^{n+1}}\end{aligned}$$

And by taking $r \to \infty$, we thus have that $f^{(n+1)}(a) = 0$. And as a was arbitrary, we have that $f^{(n+1)} = 0$ everywhere. It then also follows that $f^{(k)} = 0$ for all $k > n$.

And now any power series given for f will be a polynomial.

Exercise IV.3.6: Let G be a region and suppose that $f : G \to \mathbb{C}$ is analytic and $a \in G$ satisfies that $|f(a)| \leq |f(z)|$ for all $z \in G$. Show that either $f(a) = 0$ or f is constant.

Suppose $|f(a)| > 0$. Then $f(z) \neq 0$ for any $z \in G$ and so $\frac{1}{f}$ is analytic on G. Furthermore, we know $|\frac{1}{f(z)}| \leq |\frac{1}{f(a)}|$ for all $z \in G$. So by the maximal modulus theorem we know that $\frac{1}{f}$ is contant on G. This is equivalent to saying that f is constant on G.

Exercise IV.3.8: Let G be a region and let f and g be analytic functions on G such that $f(z)g(z) = 0$ for all $a \in G$. Show that either $f = 0$ or $g = 0$.

Suppose f is not the zero function and let U be a precompact open set whose closure is contained in G. Since $\overline{U} \subseteq G$ is compact and the zero set of f has no limit points in G, there must be finitely many points $w_1, \ldots, w_n$ in $\overline{U}$ where $f = 0$.

But now for any $a \in U - \{w_1, \ldots, w_n\}$ we know there is some $r > 0$ such that the open ball $B_r(a)$ is a subset of $U - \{w_1, \ldots, w_n\}$ (since the latter set is open). Since $f(z) \neq 0$ for all $z \in B_r(a)$, we must have that $g(z) = 0$ for all $z \in B_r(a)$. And as the set $B_r(a)$ has a limit point in G, we know $g = 0$ everywhere on G.

Exercise IV.3.10: Show that if f and g are analytic functions on a region G such that $\overline{f}g$ is analytic then either f is constant or $g = 0$.

I'd like to start by noting that Conway has a typo where he wrote $\overline{fg}$ instead of $\overline{f}g$. The way that you can spot that that is a typo is that $\overline{fg}$ is analytic iff fg is constant. After all, consider the proposition below:

Proposition: Suppose $G \subseteq \mathbb{C}$ is a region and $h : G \to \mathbb{C}$ is analytic. Then $\overline{h}$ analytic iff h is a constant function.

Proof:

If h is constant then so is $\overline{h}$ and hence $\overline{h}$ is analytic.

Conversely, write $h(x+iy) = u(x,y) + iv(x,y)$. Then by applying the Cauchy-Riemann equations to both h and $\overline{h}$ we have that $u_x = v_y$, $u_y = -v_x$, $u_x = -v_y$, and $u_y = v_x$. This forces u_x, u_y, v_x, v_y to all be zero. Hence h is constant on G.

But clearly it should be possible for fg to be constant while f itself is not constant and g is not identically zero. Hence, we know there is a typo in how Conway originally wrote the problem.

Now let's actually get to the actual exercise I copied down. Write:

$f(x+iy) = r(x,y) + is(x,y)$ and $g(x+iy) = u(x,y) + iv(x,y)$.

Then $\overline{f}g = (ru + sv) + i(rv - su)$. And since $\overline{f}g$ is analytic we can apply the the Cauchy-Riemann equations to get that:

- $r_x u + r u_x + s_x v + s v_x = r_y v + r v_y - s_y u - s u_y$
- $r_y u + r u_y + s_y v + s v_y = s_x u + s u_x - r_x v - r v_x$

But note that since we can also apply the Cauchy-Riemann equations to f and g individually, we know that:

- $r_y v + r v_y - s_y u - s u_y = -s_x v + r u_x - r_x u + s v_x$
- $s_x u + s u_x - r_x v - r v_x = -r_y u + s v_y - s_y v + r u_y$

Hence, we can conclude that $2r_x u + 2s_x v = 0$ and $2r_y u + 2s_y v = 0$. Or to put in other words, we know the following system of equations must be satisfied:

$$\begin{bmatrix} r_x & s_x \\ r_y & s_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Now suppose $g \neq 0$ everywhere on G and pick $a \in G$ with $g(a) \neq 0$. By the continuity of g we can find some $r > 0$ such that $B_r(a) \subseteq G$ and $g(z) \neq 0$ for any $z \in B_r(a)$. The significance of this is that there is some subset of G with limit points in G such that $(u,v) \neq 0$ on that subset.

But now by the Cauchy-Riemann equations applied to f, we know that:

$\det\left(\begin{bmatrix} r_x & s_x \\ r_y & s_y \end{bmatrix}\right) = (r_x)^2 + (s_x)^2 = (s_y)^2 + (r_y)^2 \geq 0$ with equality iff $f' = 0$.

Since $(u,v) \neq 0$ on $B_r(a)$, we must have that the matrix $\left[\begin{smallmatrix} r_x & s_x \\ r_y & s_y \end{smallmatrix}\right]$ is not full rank on $B_r(a)$. Hence, $f' = 0$ on $B_r(a)$. And from there we know f is constant on $B_r(a)$ since $B_r(a)$ is a region. And as f is constant on a set with a limit point in G, we know that f is constant everywhere. ■

Exercise IV.4.3: Let $p(z)$ be a polynomial of degree n and let $R > 0$ be sufficiently large so that $p(z) \neq 0$ for all z with $|z| > R$. If $\gamma(t) = Re^{it}$ then show:

$\int_\gamma \frac{p'(z)}{p(z)} \mathrm{d}z = 2\pi i n$ for some $n \in \mathbb{Z}$.

Write $p(z) = c\prod_{i=1}^{n}(z - a_i)$ where $c \neq 0$ and $a_0, \ldots, a_n$ are the (perhaps nondistinct) zeros of p. Then by the product rule we have that:

$$p'(z) = c\sum_{i=1}^{n}(z - a_i)'(\prod_{\substack{j\in\{1,\ldots,n\}\\ j\neq i}}(z - a_j)) = c\sum_{i=1}^{n}(\prod_{\substack{j\in\{1,\ldots,n\}\\ j\neq i}}(z - a_j))$$

Hence:

$$\int_\gamma \frac{p'(z)}{p(z)}\mathrm{d}z = \int_\gamma (c\sum_{i=1}^{n}(\prod_{\substack{j\in\{1,\ldots,n\}\\ j\neq i}}(z - a_j))) \cdot (c\prod_{i=1}^{n}(z - a_i))^{-1}\,\mathrm{d}z = \int_\gamma \sum_{i=1}^{n}\frac{1}{z - a_i}\mathrm{d}z$$

And since all the a_i are in the inside component of $\mathbb{C} - \{\gamma\}$, we know that:

$$\int_\gamma \sum_{i=1}^{n}\frac{1}{z-a_i}\mathrm{d}z = \sum_{i=1}^{n}\int_\gamma \frac{1}{z-a_i}\mathrm{d}z = \sum_{i=1}^{n} n(\gamma; a_i)2\pi i = \sum_{i=1}^{n} 2\pi i = 2\pi i n.$$

Misc Problem From the Exam: Suppose $u : \mathbb{R}^2 \to \mathbb{R}$ is harmonic with $u \geq 0$. Then u is constant.

Proof:

Let $v : \mathbb{R}^2 \to \mathbb{R}$ be the harmonic conjugate of u. (We know v exists by *(Conway) Theorem 2.30*). Then $f(x + iy) = u(x, y) + iv(x, y)$ is holomorphic on all of $\mathbb{C}$. And if we set $h(z) = e^{-f(z)}$, then we know h is also holomorphic on all of $\mathbb{C}$ with:

$|h(z)| = e^{-u(z)} \leq e^0 = 1$ for all $z \in \mathbb{C}$.

By Liouville's theorem we thus know that $h = e^{-f(z)}$ is constant. In turn, $|h| = e^{-u(z)}$ is constant. And finally, we have that $-\log|h| = u(x, y)$. So u is constant.

Math 200a Notes: (Again thanks to Hagan for their 11/7 lecture notes...)

Proposition: Let X be a nonempty set. Then there exists a free group $\mathcal{F}(X)$ satisfying the following universal property.

Set Grp

$$\begin{tikzcd} X \arrow[r, hook, "i"] \arrow[dr, "\forall f"'] & \mathcal{F}(X) \arrow[d, dashed, "\exists!\widehat{f}"] \\ & G \end{tikzcd}$$

(The meaning of this diagram is that for all functions from X to G, there exists a unique group homomorphism $\widehat{f}$ such that the below diagram commutes.

Set and Grp are the names of the categories of sets and groups respectively.)

Proof:
For all $x \in X$ let $C_x := \{x^n : n \in \mathbb{Z}\} \cong \mathbb{Z}$ (where x^n is a formal power and $x^{n_1} \cdot x^{n_2} := x^{n_1+n_2}$). Note that $C_x \cap C_{x'} = \emptyset$ if $x \neq x'$. So define $\mathcal{F}(X) = *_{x\in X} C_x$. Then define $j : X \hookrightarrow \mathcal{F}(X)$ by $j(x) = j_x(x^1)$ where $j_{x'} : C_{x'} \hookrightarrow *_{x\in X} C_x$ is the natural embedding.

Suppose $f : X \to G$ is any function. Then define $f_x : C_x \to G$ by $f_x(x^n) = (f(x))^n$ for all $n \in \mathbb{Z}$ and $x \in X$. By the universal property of the free product, we know there is a unique group homomorphism $\widehat{f} : *_{x\in X} C_x \to G$ such that the below diagram commutes for all $x' \in X$:

$$\begin{tikzcd} C_{x'} \arrow[r, hook, "j_{x'}"] \arrow[dr, "f_{x'}"'] & *_{x\in X} C_x \arrow[d, "\widehat{f}"] \\ & G \end{tikzcd}$$

In particular, the diagram below commutes since $\widehat{f}(j(x)) = \widehat{f}(j_x(x^1)) = f_x(x^1) = f(x)$:

$$\begin{tikzcd} X \arrow[r, hook, "j"] \arrow[dr, "f"'] & \mathcal{F}(X) \arrow[d, "\widehat{f}"] \\ & G \end{tikzcd}$$

And the uniqueness of $\widehat{f}$ is because each $f_x : C_x \to G$ is uniquely determined by the value it takes at x^1. ■

Ping-Pong Lemma: Suppose $G \curvearrowright X$ is a group action such that G_1, G_2 are subgroups of G with $|G_1| \geq 2$ and $|G_2| \geq 3$. If there exists $X_1, X_2 \subseteq X$ with $X_1 \not\subseteq X_2$ and $X_2 \not\subseteq X_1$ (their intersection is allowed to be nonempty) such that $(G_1 - \{1\}) \cdot X_2 \subseteq X_1$ and $(G_2 - \{1\}) \cdot X_1 \subseteq X_2$, then $\langle G_1, G_2 \rangle \cong G_1 * G_2$.

Note, $\langle G_1, G_2 \rangle$ is just the group generated by the subset $G_1 \cup G_2$. Meanwhile, $G_1 * G_2$ is the free product of G_1 and G_2.

Proof:
Let $f_i : G_i \hookrightarrow \langle G_1 \cup G_2 \rangle$ be the natural embedding. Then by the universal property of free products, there exists a unique group homomorphism $\widehat{f} : G_1 * G_2 \to \langle G_1 \cup G_2 \rangle$ such that the below diagrams commute:

$$\begin{tikzcd} G_1 \arrow[r, hook, "j_1"] \arrow[dr, "f_1"'] & G_1 * G_2 \arrow[d, "\widehat{f}"] \\ & \langle G_1 \cup G_2 \rangle \end{tikzcd} \qquad \begin{tikzcd} G_2 \arrow[r, hook, "j_2"] \arrow[dr, "f_2"'] & G_1 * G_2 \arrow[d, "\widehat{f}"] \\ & \langle G_1 \cup G_2 \rangle \end{tikzcd}$$

Note that as $G_1 \cup G_2$ is in the image of $\widehat{f}$, we must have that $\widehat{f}$ is surjective. Hence, all we need to show is that $\widehat{f}$ is injective.

Suppose $\omega \in \ker(\widehat{f})$. If $\omega \neq \emptyset$ then we can guarentee that after reducing until we can't anymore, we'll have that $\omega = x_1 \cdots x_m$ where $m \geq 1$, no x_k is 1_{G_1} or 1_{G_2}, and $x_k \in G_i \Longrightarrow x_{k+1} \notin G_i$ for all k. In particular, we have four cases:

Case 1:
Let $\omega = a_1 b_1 a_2 b_2 \cdots a_n b_n a_{n+1}$ where each $a_k \in G_1 - \{1\}$ and each $b_k \in G_2 - \{1\}$. Then since $\widehat{f}(\omega) = 1$ we know $a_1 b_1 a_2 b_2 \cdots a_n b_n a_{n+1} \cdot x = x$ for all $x \in X$. But if $x \in X_2 - X_1$ then:

$$a_{n+1} \cdot x \in X_1 \Longrightarrow b_n a_{n+1} \cdot x \in X_2 \Longrightarrow \cdots \Longrightarrow a_1 b_1 a_2 b_2 \cdots a_n b_n a_{n+1} \cdot x \in X_1$$

Note, this is the ping-ponging in the name of the lemma...

This is a contradiction.

Case 2:
Let $\omega = b_1 a_1 b_2 a_2 \cdots b_n a_n b_{n+1}$ where each $a_k \in G_1 - \{1\}$ and each $b_k \in G_2 - \{1\}$. Then we get a contradiction analogously to in case 1.

Case 3:
Let $\omega = a_1 b_1 a_2 b_2 \cdots a_n b_n$ where each $a_k \in G_1 - \{1\}$ and each $b_k \in G_2 - \{1\}$. Since $|G_2| \geq 3$, we know there exists $b \notin \{1, b_n^{-1}\}$. Then:

$$\omega \in \ker(\widehat{f}) \Longrightarrow b^{-1} \omega b \in \ker(\widehat{f}).$$

But $b^{-1} \omega b = b^{-1} a_1 b_1 a_2 \cdots b_{n-1} a_n (b_n b)$ where $b_n b \neq 1$. So, we can now get a contradiction using the reasoning of case 2.

Case 4:
Let $\omega = b_1 a_1 b_2 a_2 \cdots b_n a_n$ where each $a_k \in G_1 - \{1\}$ and each $b_k \in G_2 - \{1\}$. Then by analogous reasoning to case 3 using some $b \notin \{1, b_1^{-1}\}$ we get another contradiction.

All of this proves that $\ker(\widehat{f}) = \{\emptyset\}$. So, $\widehat{f}$ is injective. ■

11/13/2025

Some Math 220a Notes:

(Conway) Lemma IV.5.1: Let γ be a piecewise C^1 curve and suppose φ is a function defined and continuous on $\{\gamma\}$. For each $m \geq 1$ let $F_m(z) = \int_\gamma \varphi(w)(w-z)^{-m} \mathrm{d}w$ for $z \notin \{\gamma\}$. Then each F_m is analytic on $\mathbb{C} - \{\gamma\}$ and $F_m'(z) = m F_{m+1}(z)$.

Firstly here is a quick proof:

By *Exercise IV.2.2* we know each F_m is continuously differentiable with:
$F_m'(z) = \int_\gamma \frac{\partial}{\partial z}[\varphi(w)(w-z)^{-m}]\mathrm{d}w = m \int_\gamma \varphi(w)(w-z)^{-(m+1)}\mathrm{d}w = m F_{m+1}(z)$.

Here is a longer proof:

Recall that:

$$x^m - y^m = (x-y)\sum_{k=1}^m x^{m-k}y^{k-1}.$$

In particular, this means that:

$$\frac{1}{(w-z)^m} - \frac{1}{(w-a)^m} = \left(\frac{1}{w-z} - \frac{1}{w-a}\right)\sum_{k=1}^m \frac{1}{(w-z)^{m-k}(w-a)^{k-1}}$$

$$= \left(\frac{(w-a)-(w-z)}{(w-z)(w-a)}\right)\sum_{k=1}^m \frac{1}{(w-z)^{m-k}(w-a)^{k-1}} = (z-a)\sum_{k=1}^m \frac{1}{(w-z)^{m-k+1}(w-a)^k}$$

So for any fixed $a \in \mathbb{C} - \{\gamma\}$ we have that $\frac{F_m(z)-F_m(a)}{z-a} = \sum\limits_{k=1}^m \int_\gamma \frac{\varphi(w)(w-a)^{-k}}{(w-z)^{m-k+1}}\mathrm{d}w$.

In turn, $F_m'(a) = \sum\limits_{k=1}^m \lim\limits_{z\to a}\int_\gamma \frac{\varphi(w)(w-a)^{-k}}{(w-z)^{m-k+1}}\mathrm{d}w$.

And upon showing that $f_k(z) := \int_\gamma \frac{\varphi(w)(w-a)^{-k}}{(w-z)^{m-k+1}}\mathrm{d}w$ is continuous at a for each k, we get that:

$$F_m'(a) = \sum_{k=1}^m \lim_{z\to a}\int_\gamma \frac{\varphi(w)(w-a)^{-k}}{(w-z)^{m-k+1}}\mathrm{d}w = \sum_{k=1}^m \int_\gamma \frac{\varphi(w)(w-a)^{-k}}{(w-a)^{m-k+1}}\mathrm{d}w$$

$$= \sum_{k=1}^m F_{m+1}(z) = mF_{m+1}(z).\ \blacksquare$$

<u>(Conway) Cauchy's Integral Formula (First Version):</u> Let G be an open subset of the plane and $f : G \to \mathbb{C}$ an analytic function. If γ is a closed piecewise C^1 curve in G such that $n(\gamma; w) = 0$ for all $z \in \mathbb{C} - G$, then for $z \in G - \{\gamma\}$ we have that:

$$n(\gamma; z)f(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{w-z}\mathrm{d}w.$$

Proof:

We want to show that $0 = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{w-z}\mathrm{d}w - f(z)\frac{1}{2\pi i}\int_\gamma \frac{1}{w-z}\mathrm{d}w = \frac{1}{2\pi i}\int_\gamma \frac{f(w)-f(z)}{w-z}\mathrm{d}w$ for all $z \in G - \{\gamma\}$. Therefore consider defining $\varphi : G \times G \to \mathbb{C}$ by:

$\varphi(z,w) = \frac{f(w)-f(z)}{w-z}$ when $z \neq w$ and $\varphi(z,w) = f'(z)$ when $z = w$.

Exercise IV.5.1: φ is continuous and for each fixed w the map $z \mapsto \varphi(z,w)$ is analytic.

First we show φ is continuous. Suppose $z_0 \neq w_0$. Then we know there exists $r > 0$ such that $B_r(z_0) \cap B_r(w_0) = \emptyset$. And hence on the neighborhood $B_r(z_0) \times B_r(w_0)$ of (z_0, w_0) we have that $\varphi(z,w) = \frac{f(w)-f(z)}{w-z}$ is a quotient of two continuous functions with the denominator never equaling zero. So φ is continuous at (z_0, w_0).

Meanwhile, suppose $z_0 = w_0$. Since f' is continuous, we know that if $H = \{(z,z) : z \in G\}$ then:

$$\lim_{(z,w)\to(z_0,w_0)} \varphi|_H(z,w) = \lim_{z\to z_0} f'(z) = f'(z_0) = \varphi(z_0,w_0).$$

Hence, it now suffices to show that $\lim_{(z,w)\to(z_0,z_0)} \frac{f(w)-f(z)}{w-z} = f'(z_0)$ (where we restrict ourselves to picking $(z,w) \notin H$). Luckily, note that:

$$\left|\frac{f(w)-f(z)}{w-z} - f'(z_0)\right| \leq \left|\frac{f(w)-f(z)}{w-z} - f'(z)\right| + |f'(z) - f'(z_0)|.$$

Since f' is continuous we know that $|f'(z) - f'(z_0)| \to 0$ as $(z, w) \to (z_0, w_0)$. At the same time, note that when $|w - z| = r$ is small enough so that $B_r(z) \subseteq G$ then:

$$f(w) = f(z) + f'(z)(w - z) + (w - z)^2 \sum_{n=0}^{\infty} \tfrac{f^{(n+2)}(z)}{(n+2)!}(w - z)^n$$

In particular:

$$\begin{aligned} \left|\tfrac{f(w)-f(z)}{w-z} - f'(z)\right| &= |w - z| \cdot \left|\sum_{n=0}^{\infty} \tfrac{f^{(n+2)}(z)}{(n+2)!}(w - z)^n\right| \\ &\leq |w - z| \cdot \sum_{n=0}^{\infty} \tfrac{|f^{(n+2)}(z)|}{(n+2)!}|w - z|^n \end{aligned}$$

So suppose $R > 0$ is small enough so that $\overline{B_{2R}(z_0)} \subseteq G$ and then use extreme value theorem to find $M > 0$ such that $\max(|f(z)|) \leq M$ for all $z \in \overline{B_{2R}(z_0)}$. When $w, z \in B_{R/2}(z_0)$ we can use Cauchy's estimate to conclude that:

$$\tfrac{|f^{(n+2)}(z)|}{(n+2)!} \leq \tfrac{\frac{M(n+2)!}{R^{n+2}}}{(n+2)!} = \tfrac{M}{R^{n+2}} \text{ for all } n.$$

And then since $|w - z| < R$ we have that:

$$\left|\tfrac{f(w)-f(z)}{w-z} - f'(z)\right| \leq \tfrac{M}{R^2}|w - z| \cdot \sum_{n=0}^{\infty} \left(\tfrac{|w-z|}{R}\right)^n = \tfrac{M|w-z|}{R^2(1-\frac{|w-z|}{R})} = \tfrac{M|w-z|}{R(R-|w-z|)}.$$

Clearly the latter expression goes to zero as $(z, w) \to (z_0, w_0)$. Hence, we've finished proving that $\left|\frac{f(w)-f(z)}{w-z} - f'(z)\right| + |f'(z) - f'(z_0)| \to 0 + 0$. And from there it is clear that φ is continuous at (z_0, w_0).

Next we show that the map $z \mapsto \varphi(z, w)$ is analytic for each fixed w. To start off, when $z \neq w$ have that $\varphi(\cdot, w)$ is a quotient of two analytic functions and is thus analytic with:

$$\tfrac{\partial\varphi}{\partial z}(z, w) = \tfrac{-f'(z)(w-z)+(f(w)-f(z))z}{(w-z)^2}$$

Meanwhile, to show that $\varphi(\cdot, w)$ is analytic at $z = w$ note that there exists $r > 0$ such that:

$$f(z) = \sum_{n=0}^{\infty} \tfrac{f^{(n)}}{n!}(z - w)^n \text{ when } |z - w| < r.$$

In turn, when $z \neq w$ we have that:

$$\varphi(z, w) = \tfrac{f(w)-f(z)}{w-z} = \tfrac{1}{w-z}\left(f(w) - \sum_{n=0}^{\infty} \tfrac{f^{(n)}(w)}{n!}(z - w)^n\right) = \sum_{n=1}^{\infty} \tfrac{f^{(n)}(w)}{n!}(z - w)^{n-1}.$$

Also we can easily check that:

$$\varphi(z, w) = f'(w) = \sum_{n=1}^{\infty} \tfrac{f^{(n)}(w)}{n!}(z - w)^{n-1} \text{ when } z = w.$$

Hence $\varphi(\cdot, w)$ can be represented by a power series at $z = w$, proving that $\varphi(\cdot, w)$ is analytic at $z = w$.

Returning back to our original goal, let $\Omega = \{w \in \mathbb{C} : n(\gamma; w) = 0\}$. Note that since $G^{\mathsf{C}} \subseteq \Omega$ by assumption, we have that $\Omega \cup G = \mathbb{C}$. Also, since $n(\gamma; w)$ is a continuous *integer-valued* function of w and $\mathbb{C} - \{\gamma\}$ is open, we know that Ω is open in $\mathbb{C}$.

Define $g : \mathbb{C} \to \mathbb{C}$ by $g(z) = \int_\gamma \varphi(z, w)\mathrm{d}w$ if $z \in G$ and $g(z) = \int_\gamma (w - z)^{-1} f(w)\mathrm{d}w$ if $z \in \Omega$. Note that if $z \in G \cap \Omega$ then:

$$\int_\gamma \varphi(z, w)\mathrm{d}w = \int_\gamma \frac{f(w) - f(z)}{w - z}\mathrm{d}w = \int_\gamma \frac{f(w)}{w - z}\mathrm{d}w - f(z)n(\gamma; z)2\pi i = \int_\gamma \frac{f(w)}{w - z}\mathrm{d}w.$$

Thus, g is well-defined. Also, by *exercise IV.2.2* we have that g is analytic on all of $\mathbb{C}$. But note that $K := \mathbb{C} - \Omega$ is closed and bounded, and hence compact. Also $K \subseteq G$ and contains $\{\gamma\}$. So, we can find a precompact open set $V \subseteq \mathbb{C}$ such that $K \subseteq V \subseteq \overline{V} \subseteq \mathbb{C}$. Then there exists $M_1 \geq 0$ such that $|\varphi(z)| \leq M_1$ for all $z \in \overline{V}$. Also, there exists $M_2 \geq 0$ such that $|f(w)| \leq M_2$ for all $w \in \{\gamma\}$. And thirdly, since $\{\gamma\}$ is compact and V^{C} is a closed set disjoint from $\{\gamma\}$, we know that:

$$d := \inf\{|w - z| : w \in \{\gamma\}, z \in V^{\mathsf{C}}\} > 0.$$

Hence, we can conclude that $|g(z)| \leq \max(M_1, \frac{M_2}{d}) \cdot \int_a^b |\gamma'(t)|\mathrm{d}t < \infty$ for all $z \in \mathbb{C}$.

This proves that g is a bounded entire function. So, we know that g is constant. And since clearly $\int_\gamma \frac{f(w)}{w - z}\mathrm{d}w \to 0$ as $|z| \to \infty$, we know $g = 0$ everywhere. This proves that $\int_\gamma \frac{f(w) - f(z)}{w - z}\mathrm{d}w = 0$ for all $z \in G - \{\gamma\}$. ■

<u>(Conway) Cauchy's Integral Formula (Second Version):</u> Let G be an open subset of the plane and $f : G \to \mathbb{C}$ an analytic function. If $\gamma_1, \ldots, \gamma_m$ are closed piecewise C^1 curves in G such that $n(\gamma_1; a) + \ldots + n(\gamma_m; a) = 0$ for all $a \in \mathbb{C} - G$, then for any $z \in G - \bigcup_{k=1}^m \{\gamma_k\}$ we have that:

$$f(z) \sum_{k=1}^m n(\gamma_k; z) = \sum_{k=1}^m \frac{1}{2\pi i} \int_{\gamma_k} \frac{f(w)}{w - z}\mathrm{d}w$$

Proof:
Similarly to the proof of our prior theorem, note that our proposed formula holds iff $0 = \frac{1}{2\pi i} \sum_{k=1}^m \int_{\gamma_k} \frac{f(w) - f(z)}{w - z}\mathrm{d}w$ for all $z \in G - \bigcup_{k=1}^m \{\gamma_k\}$. So define $\varphi(z, w)$ like before. Then upon letting $\Omega = \{z \in \mathbb{C} : \sum_{k=1}^m n(\gamma_k; z) = 0\}$, define

$$g(z) := \sum_{k=1}^m \int_{\gamma_k} \varphi(z, w)\mathrm{d}w \text{ when } z \in G \text{ and } g(z) := \sum_{k=1}^m \int_{\gamma_k} \frac{f(w)}{w - z}\mathrm{d}w \text{ when } z \in \Omega$$

By nearly identical reasoning to before, we can show that g is well-defined, entire, and equal to zero everywhere. ■

I'd like to add that the previous two results are way stronger than what I proved in my notes last Spring. After all, Conway has made no assumptions about G being convex. That said, if G is convex then we'd automatically have that $n(\gamma; z) = 0$ for all $z \notin G$ since G^{C} would be a subset of the unbounded component of $\{\gamma\}^{\mathsf{C}}$.

<u>(Conway) Cauchy's Theorem (First Version):</u> Let $G \subseteq \mathbb{C}$ be an open set and $f : G \to \mathbb{C}$ be analytic. If $\gamma_1, \ldots, \gamma_m$ are closed piecewise C^1 curves in G such that $n(\gamma_1; a) + \ldots + n(\gamma_m; a) = 0$ for all $a \in \mathbb{C} - G$ then:

$$\sum_{k=1}^{m} \int_{\gamma_k} f(w)\mathrm{d}w = 0$$

Proof:
Pick $a \in G - \bigcup_{k=1}^{m}\{\gamma_k\}$. Note that such a point exists because $\bigcup_{k=1}^{m}\{\gamma_k\}$ is compact and a subset of G. Thus if $G = \mathbb{C}$ we know $\bigcup_{k=1}^{m}\{\gamma_k\}$ is bounded and not all of $\mathbb{C}$. Meanwhile, if $G \subsetneq \mathbb{C}$ then we know G isn't clopen. So we can't have that $G = \bigcup_{k=1}^{m}\{\gamma_k\}$.

Next, apply Cauchy's Integral Formula (second version) to $f(z)(z-a)$. This shows for all $z \in G - \{\gamma\}$ that:

$$f(z)(z-a)\sum_{k=1}^{m} n(\gamma_k; z) = \sum_{k=1}^{m} \tfrac{1}{2\pi i}\int_{\gamma_k} \tfrac{f(w)(w-a)}{w-z}\mathrm{d}w$$

And in particular, by plugging in $z = a$ we have that:

$$0 = \sum_{k=1}^{m} \tfrac{1}{2\pi i}\int_{\gamma_k} \tfrac{f(w)(w-a)}{w-a}\mathrm{d}w = \tfrac{1}{2\pi i}\sum_{k=1}^{m}\int_{\gamma_k} f(w)\mathrm{d}w. \blacksquare$$

Exercise IV.5.4: We can conversely say that Cauchy's integral formula follows from Cauchy's theorem.

Suppose $G \subseteq \mathbb{C}$ is open and $\gamma_1, \ldots, \gamma_m$ are closed piecewise C^1 curves in G such that $n(\gamma_1; a) + \ldots + n(\gamma_m; a) = 0$ for all $a \in \mathbb{C} - G$. Then given any fixed $z \in G$ define $\varphi(w) := \frac{f(w)-f(z)}{w-z}$ when $w \neq z$ and $\varphi(z) = f'(z)$.

φ is trivially analytic when $w \neq z$. Meanwhile, note that when w is sufficiently close but not equal to z we have that:

$$\varphi(w) = \tfrac{f(w)-f(z)}{w-z} = \tfrac{1}{w-z}\left(\left(\textstyle\sum_{n=0}^{\infty}\tfrac{f^{(n)}(z)}{n!}(w-z)^n\right) - f(z)\right) = \textstyle\sum_{n=1}^{\infty}\tfrac{f^{(n)}(z)}{n!}(w-z)^{n-1}$$

Yet the latter power series also equals $f'(z) = \varphi(w)$ when $w = z$. So, we've shown that φ is analytic at $w = z$ as well.

By Cauchy's theorem we know that $\sum_{k=1}^{m}\int_{\gamma_k}\varphi(w)\mathrm{d}w = 0$. But in particular if $z \notin \bigcup_{k=1}^{m}\{\gamma_k\}$ then we have that:

$$0 = \sum_{k=1}^{m}\int_{\gamma_k}\tfrac{f(w)-f(z)}{w-z}\mathrm{d}w \Longrightarrow \sum_{k=1}^{m}\int_{\gamma_k}\tfrac{f(w)}{w-z}\mathrm{d}w = f(z)\sum_{k=1}^{m}\int_{\gamma_k}\tfrac{1}{w-z}\mathrm{d}w = f(z)2\pi i\sum_{k=1}^{m} n(\gamma; z).$$

Divide both sides by $2\pi i$ and we are done. $\blacksquare$

Exercise IV.5.5: Let γ be a closed piecewise C^1 curve in a region G and suppose $a \notin \{\gamma\}$. Then if $n \geq 2$ we have that $\int_\gamma (w-a)^{-n}\mathrm{d}w = 0$.

To start off, note that if $\gamma_r(t) = a + re^{ikt}$ for $t \in [0, 2\pi]$ where $k \in \mathbb{Z}$, $r > 0$, and $a \in \mathbb{C}$; then $n(\gamma; a) = k$. So fix $k = -n(\gamma; a)$ and let V be the connected component of $\mathbb{C} - \{\gamma\}$ containing a. Since $V \subseteq \mathbb{C}$ is open, we know there is a compact set F such that $a \in F \subseteq V$. And in turn, we can let $G = F^{\mathsf{C}}$. That way, $(z-a)^{-n}$ is analytic on G and when r is sufficiently large we have that $\{\gamma\}$ and $\{\gamma_r\}$ are contained in G with $n(\gamma; z) + n(\gamma_r; z) = 0$ for all $z \in V \supseteq \mathbb{C} - G$.

Thus, by Cauchy's integral formula we know that:

$$0 = \int_\gamma (w-a)^{-n}\mathrm{d}w + \int_{\gamma_r}(w-a)^{-n}\mathrm{d}w$$
$$= \int_\gamma (w-a)^{-n}\mathrm{d}w - \int_0^{2\pi}(re^{-in(\gamma;a)t})^{-n} rin(\gamma;a)e^{-in(\gamma;a)t}\mathrm{d}t$$

But note that $\left|\int_0^{2\pi}(re^{-in(\gamma;a)t})^{-n} rin(\gamma;a)e^{-in(\gamma;a)t}\mathrm{d}t\right| \leq \int_0^{2\pi} r^{-n+1}n(\gamma;a)\mathrm{d}t = 2\pi r^{-n+1}n(\gamma;a)$.

Thus by taking $r \to \infty$ we can make $\int_0^{2\pi}(re^{-in(\gamma;a)t})^{-n} rin(\gamma;a)e^{-in(\gamma;a)t}\mathrm{d}t$ arbitrarily small. And this forces $\int_\gamma (w-a)^{-n}\mathrm{d}w = 0$. ■

Exercise IV.5.8: Let G be a region and suppose $f_n : G \to \mathbb{C}$ is analytic for each $n \geq 1$. Suppose that $\{f_n\}_{n\in\mathbb{N}}$ converges uniformly to $f : G \to \mathbb{C}$. Then show that f is analytic.

Because each f_n is continuous and $f_n \to f$ uniformly, we know f is continuous. Also, if Δ is any triangular path in G (satisfying that the bounded component of $\mathbb{C} - \{\Delta\}$ is a subset of G), then we know by Cauchy's theorem plus the fact that $f_n \to f$ uniformly that:

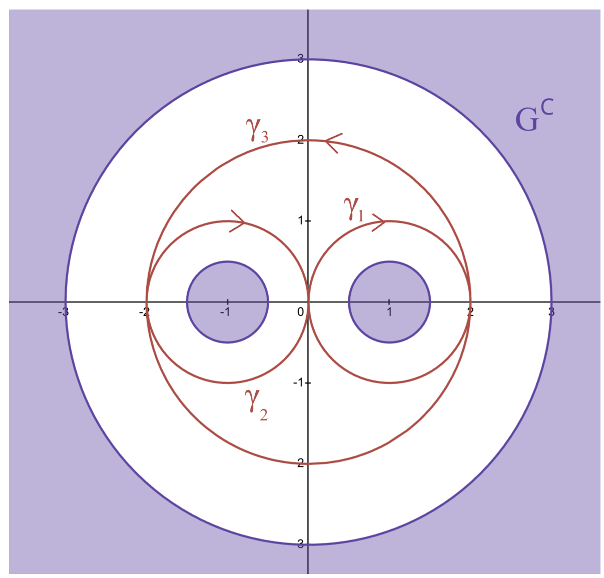
$$\int_\Delta f(w)\mathrm{d}w = \lim_{n\to\infty}\int_\Delta f_n(w)\mathrm{d}w = \lim_{n\to\infty} 0 = 0$$

Now the fact that f is analytic follows from Morera's theorem ([note to me and not the grader] see my notes from last spring).

Exercise IV.5.3: Let $B_\pm := B_{1/2}(\pm 1)$ and $G := B_3(0) - (B_+ \cup B_-)$. Let $\gamma_1, \gamma_2, \gamma_3$ be curves whose traces are $\{|z-1| = 1\}$, $\{|z+1| = 1\}$, and $\{|z| = 2\}$ respectively. Then give $\gamma_1, \gamma_2, \gamma_3$ orientations such that $n(\gamma_1; w) + n(\gamma_2; w) + n(\gamma_3; w) = 0$ for all $w \in \mathbb{C} - G$.

Parametrize γ_1 by $1 + e^{-it}$ where $t \in [0, 2\pi]$. Similarly parametrize γ_2 by $-1 + e^{-it}$ where $t \in [0, 2\pi]$. And finally, parametrize γ_3 by $2e^{+it}$ for $t \in [0, 2\pi]$.

Hopefully it is now obvious enough that $n(\gamma_1; w) + n(\gamma_2; w) + n(\gamma_3; w) = 0$ for all $w \in \mathbb{C} - G$.



Exercise IV.5.9: Show that if $f : \mathbb{C} \to \mathbb{C}$ is a continuous function such that f is analytic on $\mathbb{C} - [-1, +1]$ then f is an entire function.

To start off, here is some notation. Let $\Delta(a, b, c)$ be the closure of the triangular region with vertices a, b, c. Then let $\partial\Delta(a, b, c)$ denote the polygonal path $[a, b, c, a]$.

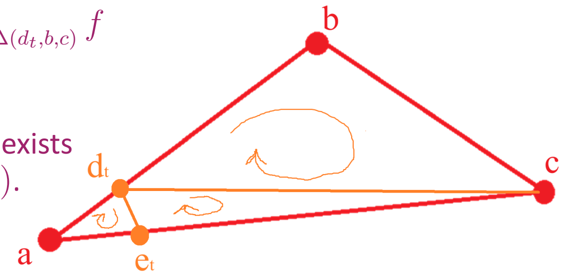
If $\Delta(a, b, c) \cap [-1, +1] = \emptyset$, then $\int_{\partial\Delta(a,b,c)} f(w)\mathrm{d}w = 0$ by Cauchy's theorem. So, our goal now is to show that we still have that $\int_{\partial\Delta(a,b,c)} f(w)\mathrm{d}w = 0$ even when $\Delta(a, b, c) \cap [-1, +1] \neq \emptyset$. That way, we can apply Morera's theorem and be done.

Lemma 1: If $\Delta(a,b,c) \cap [-1,1] = \{a\}$ then $\int_{\partial\Delta(a,b,c)} f(w)\mathrm{d}w = 0$.

Proof:

Let $d_t = ta + (1-t)b$ and $e_t = ta + (1-t)c$. Then:

$$\int_{\partial\Delta(a,b,c)} f = \int_{\partial\Delta(a,d_t,e_t)} f + \int_{\partial\Delta(e_t,d_t,c)} f + \int_{\partial\Delta(d_t,b,c)} f$$
$$= \int_{\partial\Delta(a,d_t,e_t)} f + 0 + 0.$$



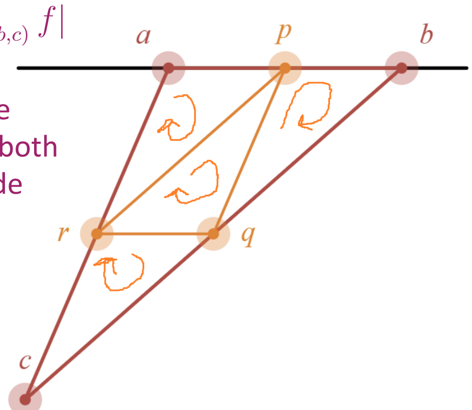
But note by extreme value theorem that there exists $M \geq 0$ with $|f(z)| \leq M$ for all $z \in \Delta(a,b,c)$. Also, by taking $t \to 1$ we can minimize the length of $\partial\Delta(a,d_t,e_t)$. So, we can make $\int_{\partial\Delta(a,d_t,e_t)} f$ arbitrarily small. And this proves the claim.

Lemma 2: If $\Delta(a,b,c) \cap [-1,1] = [a,b]$, then $\int_{\partial\Delta(a,b,c)} f(w)\mathrm{d}w = 0$.

Proof:

To start off, we introduce the following procedure. Let p, q, r be the midpoints of $[a,b]$, $[b,c]$, and $[a,c]$ respectively. Then note that:

$$\int_{\partial\Delta(a,b,c)f} = \int_{\partial\Delta(a,p,r)} f + \int_{\partial\Delta(p,b,q)} f + \int_{\partial\Delta(p,q,r)} f + \int_{\partial\Delta(r,q,c)} f$$
$$= \int_{\partial\Delta(a,p,r)} f + \int_{\partial\Delta(p,b,q)} f + 0 + 0$$



It follows that either $|\int_{\partial\Delta(a,p,r)} f| \geq \frac{1}{2}|\int_{\partial\Delta(a,b,c)} f|$ or $|\int_{\partial\Delta(p,b,q)} f| \geq \frac{1}{2}|\int_{\partial\Delta(a,b,c)} f|$. Also note that both $\Delta(a,p,r)$ and $\Delta(p,b,q)$ have half the perimeter of $\Delta(a,b,c)$. And thirdly, both of these smaller triangles have exactly one side intercepting $[-1,+1]$.

Hence, after letting $I := \int_{\partial\Delta(a,b,c)} f$ and $L :=$ the perimeter of $\Delta(a,b,c)$, we can inductively find a sequence of triangular paths $\{\partial\Delta_n\}_{n\in\mathbb{N}}$ which trace the boundary of the compact triangular sets $\{\Delta_n\}_{n\in\mathbb{N}}$ such that:

- The length of $\partial\Delta_n$ is $2^{-n}L$ for all $n \in \mathbb{N}$
- $|I| \leq 2^n|\int_{\partial\Delta_n} f|$ for all $n \in \mathbb{N}$
- $\Delta(a,b,c) \supseteq \Delta_1 \supseteq \Delta_2 \supseteq \cdots$

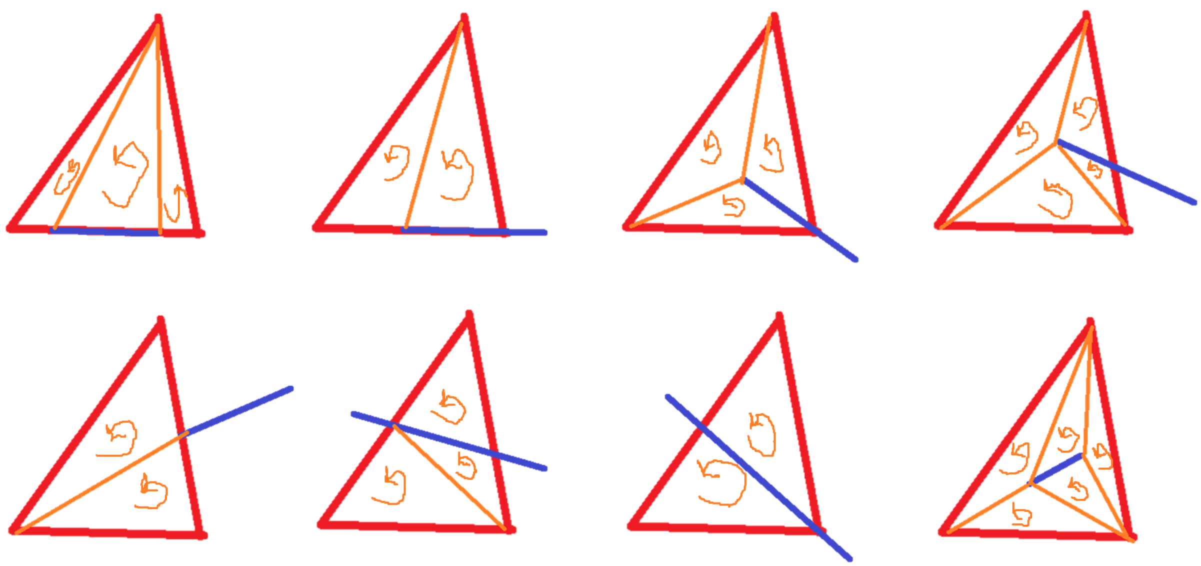
But note by the finite-intersection property that there must exist $z_0 \in \bigcap_{n\in\mathbb{N}} \Delta_n$. Then, because f is continuous at z_0, we know for any $\varepsilon > 0$ that there exists $\delta > 0$ such that $|f(w) - f(z_0)| < \varepsilon$ when $|w - z_0| < \delta$. Also, because $z \mapsto f(z_0)$ is trivially holomorphic everywhere, we know that $\int_{\Delta_n} f(z_0)\mathrm{d}w = 0$ for all n. Hence, we can say that:

$$\int_{\Delta_n} f(w)\mathrm{d}w = \int_{\Delta_n} f(w)\mathrm{d}w - \int_{\Delta_n} f(z_0)\mathrm{d}w = \int_{\Delta_n} f(w) - f(z_0)\mathrm{d}w$$

But now as $n \to \infty$ we clearly have that $\mathrm{diam}(\Delta_n) \leq \mathrm{perimeter}(\Delta_n) \to 0$. Therefore, if n is sufficiently big we can guarentee that $|w - z| < \delta$ for all $w \in \partial\Delta_n$. And then $|I| \leq 2^n|\int_{\partial\Delta_n} f(w)\mathrm{d}w| = 2^n|\int_{\partial\Delta_n} f(w) - f(z_0)\mathrm{d}w| \leq 2^n \cdot \varepsilon 2^{-n} = \varepsilon$.

Taking $\varepsilon \to 0$ then finishes the proof of this claim.

Now there are eight other cases. But luckily all of them can be addressed by applying the previous two lemmas. In the below pictures I'll use red lines to denote the triangle $\Delta(a, b, c)$, orange lines to indicate added subtriangles (as well as which way I'm integrating around them), and a blue line will denote the forbidden set $[-1, +1]$ I'm trying to work around.



Going from left to right and then top to bottom, the eight cases above are:

- $[-1, 1]$ intersects $\partial\Delta(a, b, c)$ at infinitely many points.
  (1) $[-1, 1]$ doesn't pass through a vertex of the triangle but lies on one of the line segments.
  (2) $[-1, 1]$ does pass through a vertex of the triangle and lies along a line segment.
  Note that lemma 2 deals with the case that $[-1, 1]$ passes through two vertices.

- $[-1, 1]$ intersects $\partial\Delta(a, b, c)$ at one point.
  (3) $[-1, 1]$ intersects a vertex of the triangle and then goes into the interior.
  (4) $[-1, 1]$ intersects a non-vertex point and then goes into the interior.
  (5) $[-1, 1]$ intersects a non-vertex point but doesn't extend into the interior.
  Note that lemma 1 deals with the case that $[-1, 1]$ intersects a vertex but doesn't extend into the interior.

- $[-1, 1]$ intersects $\partial\Delta(a, b, c)$ at two points.
  (6) $[-1, 1]$ intersects a vertex and a non-vertex point of $\partial\Delta(a, b, c)$.
  (7) $[-1, 1]$ intersects two non-vertex points of $\Delta(a, b, c)$.

- $[-1, 1]$ intersects $\partial\Delta(a, b, c)$ at three points.
  (8) $[-1, 1]$ is contained within the interior of $\Delta(a, b, c)$.
  Note that the case that $[-1, 1]$ is contained within the complement of $\Delta(a, b, c)$ is already dealt with by Cauchy's theorem.

As a side note, we could have replaced $[-1, 1]$ with any arbitrary line segment (or even any full line). Also, we could have replace $\mathbb{C}$ with any region containing the line segment.

Math 200a Homework:

Set 7 Problem 1: Suppose G is a finite group and let $\Phi(G) := \bigcap_{M \in \mathrm{Max}(G)} M$ denote the Frattini subgroup of G. Recall from *page 388* that $\Phi(G)$ is a characteristic subgroup.

(a) Suppose H is a subgroup of G and $H\Phi(G) = G$. Then prove that $H = G$.

Suppose $H \lneqq G$. Then we can pick $M \in \mathrm{Max}(G)$ such that $H < M$. And importantly since $H < M$ and $\Phi(G) < M$, we know that $H\Phi(G) \subseteq M$. But M is a proper subgroup of G. So $H\Phi(G) \neq G$.

(b) Suppose $S \subseteq G$. Then prove that $\langle S \rangle = G$ if and only if $\langle \pi(S) \rangle = G/\Phi(G)$ where $\pi : G \to G/\Phi(G)$ is the natural quotient map.

$(\Longrightarrow)$
We can prove a more general statement. If $\langle S \rangle = G$ and N is any normal subgroup of G, then $\langle \pi(S) \rangle = G/N$ where $\pi : G \to G/N$ is the natural quotient map. To see why, note that if $g = x_1 \cdots x_k$ where $x_1, \ldots, x_k \in S$, then $gN = (x_1N) \cdots (x_kN) \in \langle \pi(S) \rangle$.

$(\Longleftarrow)$
Note that if $f : G \to G'$ is a group homomorphism, then $\langle f(S) \rangle = f(\langle S \rangle)$. This is because if $g = y_1 \cdots y_k$ where $y_i = f(x_i)$ then $g = f(x_1 \cdots x_k) = f(\langle S \rangle)$. Hence $\langle f(S) \rangle \subseteq f(\langle S \rangle)$. Conversely, $\langle f(S) \rangle \supseteq f(\langle S \rangle)$ because if $h = x_1 \cdots x_k$ where $x_1, \ldots, x_k \in S$ then $f(h) = f(x_1) \cdots f(x_k) \in \langle f(S) \rangle$.

But now $G/\Phi(G) = \langle \pi(S) \rangle = \pi(\langle S \rangle) = (\langle S \rangle \Phi(G))/\Phi(G)$. Therefore, $G = \langle S \rangle \Phi(G)$. And by part (a) this implies that $\langle S \rangle = G$.

(c) Prove that $\Phi(G)$ is nilpotent.

Suppose P is a Sylow p-subgroup of $\Phi(G)$. Then as $\Phi(G) \lhd G$, we know by Frattini's trick that $G = N_G(P)\Phi(G)$. But in turn by part (a) we know that $N_G(P) = G$. Hence $P \lhd G$. And in particular, we have that $P \lhd \Phi(G)$. Since $\Phi(G)$ is a finite group and p could be any prime, this proves that $\Phi(G)$ is nilpotent.

(d) Prove that G is nilpotent if and only if $G/\Phi(G)$ is nilpotent.

$(\Longrightarrow)$
Let $G = N_1 \supseteq N_2 \supseteq \cdots \supseteq N_{c+1} = \{1\}$ be a central series for G. Then define $\overline{N_i} = \pi(N_i)$ for each i where $\pi : G \to G/\Phi(G)$ is the natural quotient map. Note by the correspondance theorem that $\overline{N_i} = (N_i\Phi(G))/\Phi(G)$ is normal in $G/\Phi(G)$ for all i since $N_i\Phi(G) \lhd G$. So, all we need to show is that $\overline{N_{i-1}}/\overline{N_i} \subseteq Z((G/\Phi(G))/\overline{N_i})$ for all i.

Recall that if $\phi : H \to H'$ is any group homomorphism then $\phi(Z(H)) \subseteq Z(\phi(H))$.

After all, if $b \in Z(H)$ then we know that $f(b) = \phi(aba^{-1}) = \phi(a)\phi(b)\phi(a)^{-1}$ for any $a \in H$. Hence $\phi(b) \in Z(\phi(H))$.

In particular, this means that if $\phi : H \to H'$ is a surjective map then $\phi(Z(H)) \subseteq Z(H')$. And if ϕ is an isomorphism, then we can repeat the same reasoning with ϕ^{-1} to get that $\phi(Z(H)) = Z(H')$.

Consequently, by the third isomorphism theorem we can say that the following two statements are equivalent:

- $\overline{N_{i-1}}/\overline{N_i} = \frac{(N_{i-1}\Phi(G))/\Phi(G)}{(N_i\Phi(G))/\Phi(G)} \subseteq Z(\frac{G/\Phi(G)}{(N_i\Phi(G))/\Phi(G)}) = Z((G/\Phi(G))/\overline{N_i})$
- $\frac{N_{i-1}\Phi(G)}{N_i\Phi(G)} \subseteq Z(\frac{G}{N_i\Phi(G)})$.

Fortunately, by considering the surjective group homomorphism $\tau : G/N_i \to G/(N_i\Phi(G))$ such that $xN_i \mapsto x(N_i\Phi(G))$ and recalling that $N_{i-1}/N_i \subseteq Z(G/N_i)$, we know that:

$$\frac{N_{i-1}\Phi(G)}{N_i\Phi(G)} = \tau(N_{i-1}/N_i) \subseteq Z(\tau(G/N_i)) = Z(G/(N_i\Phi(G))).$$

Note: The prior argument still works if $\Phi(G)$ is replaced by any normal subgroup of G. Thus if G is a nilpotent group we have that G/N is also nilpotent for all $N \lhd G$.

$(\Longleftarrow)$
Let Q be a Sylow p-subgroup of G. Then we know that $P = Q \cap \Phi(G)$ is a Sylow p-subgroup of G. (See problem 4b on the third homework set [*page 321*]). In turn, $|(Q\Phi(G))/\Phi(G)| = p^{\nu_p(|G|)}/p^{\nu_p(|\Phi(G)|)} = p^{\nu_p(|G/\Phi(G)|)}$. So, we can conclude that $(Q\Phi(G))/\Phi(G)$ is a Sylow p-subgroup of $G/\Phi(G)$. And since $G/\Phi(G)$ is nilpotent, we know that $(Q\Phi(G))/\Phi(G) \lhd G/\Phi(G)$.

By the correspondance theorem, we have that $Q\Phi(G) \lhd G$. Furthermore, note that as $Q < Q\Phi(G) < G$ and $\nu_p(|Q|) = \nu_p(|G|)$, we have that $Q \in \mathrm{Syl}_p(Q\Phi(G))$. Hence, by Frattini's trick we have that $N_G(Q)Q\Phi(G) = G$. And by part (a), we can conclude that $N_G(Q)Q = G$. Finally, as $Q \subseteq N_G(Q)$, this is equivalent to saying that $N_G(Q) = G$. Hence, $Q \lhd G$. ■

Set 7 Problem 2: Suppose P is a finite group and $|P| = p^n$ where p is prime and n is a positive integer.

(a) Prove that for all $M \in \mathrm{Max}(P)$, $P/M \cong \mathbb{Z}/p\mathbb{Z}$.

For any $M \in \mathrm{Max}(P)$, since P is nilpotent (cause all finite p-groups are nilpotent) and $M \subsetneqq P$, we know that $M \subsetneqq N_P(M)$. Yet now we must have $N_P(M) = G$ since otherwise would contradict that M is maximal. Hence, we know that $M \lhd P$ for all $M \in \mathrm{Max}(G)$.

But next note that the only subgroups of P/M must be $\{1M\}$ and P/M. After all, if there was some other subgroup, then we would be able to use the correspondance theorem to get a proper subset of P which properly contains M. Consequently, we know that P/M is cyclic since $\langle xM \rangle = P/M$ for all $x \notin M$. And as P is a p-group, we know so is P/M.

Finally, let x be a generator of P/M. Then we know that $o(x) = p^k$ for some $k < n$. Suppose $k > 1$. Then we'd have that $1 < o(x^p) < o(x)$. But that contradicts that P/M has no cyclic subgroups. Hence, we conclude $o(x) = p^1$. And this proves that $P/M \cong \mathbb{Z}/p\mathbb{Z}$.

(b) Prove that $P/\Phi(P)$ is isomorphic to a subgroup of $(\mathbb{Z}/p\mathbb{Z})^{|\mathrm{Max}(P)|} := \prod_{i=1}^{|\mathrm{Max}(P)|}(\mathbb{Z}/p\mathbb{Z})$.

Define $\pi : P \to \prod_{M \in \mathrm{Max}(G)}(P/M)$ by $\pi(x) = (xM)_{M \in \mathrm{Max}(G)}$. Then π is clearly a group homomorphism. Also, if $\pi(x) = (1M)_{M \in \mathrm{Max}(G)}$, we know that:

$$x \in \bigcap_{M \in \mathrm{Max}(P)} M = \Phi(P).$$

Hence, by the first isomorphism theorem plus the fact from part (a) that $P/M \cong \mathbb{Z}/p\mathbb{Z}$, we know that:

$P/\Phi(P)$ is isomorphic to a subgroup of $\prod_{M \in \mathrm{Max}(G)}(P/M) \cong \prod_{i=1}^{|\mathrm{Max}(P)|}(\mathbb{Z}/p\mathbb{Z})$.

(c) Prove that $\Phi(P) = P^p[P,P] := \{x^p y : x \in P, y \in [P,P]\}$.

To start off, every nonidentity element of $\prod_{i=1}^{|\mathrm{Max}(P)|}(\mathbb{Z}/p\mathbb{Z})$ has order p. Therefore, by part (b) we know that $x^p\Phi(P) = (x\Phi(P))^p = \Phi(P)$ for all $x \in P$. And this proves that $P^p := \{x^p : x \in P\} \subseteq \Phi(P)$.

Similarly, note that since $\prod_{i=1}^{|\mathrm{Max}(P)|}(\mathbb{Z}/p\mathbb{Z})$ is abelian, we have by part (b) that:

$$x_1x_2x_1^{-1}x_2^{-1}\Phi(P) = (x_1\Phi(P))(x_2\Phi(P))(x_1\Phi(P))^{-1}(x_2\Phi(P))^{-1} = 1\Phi(P).$$

Thus $[P,P] \subseteq \Phi(P)$. And since both $P^p \subseteq \Phi(P)$ and $[P,P] \subseteq \Phi(P)$, we know that $P^p[P,P] \subseteq \Phi(P)$.

To show the other relation, first note that $P/[P,P]$ is abelian (see *page 378*). It follows like in part (c) of the *fourth problem of the fourth homework set* that $x[P,P] \mapsto x^p[P,P]$ is a group homomorphism from $P/[P,P]$ to itself. And since the image of this group homomorphism is $(P^p[P,P])/[P,P]$, we know that $(P^p[P,P])/[P,P]$ is a group. In turn, by the correspondance theorem we know that $P^p[P,P]$ is also a group.

It's also easy to see that $P^p[P,P]$ is closed under conjugation. After all, if $x \in P$ and $y \in [P,P]$, then for any $a \in P$ we have that $ax^pya^{-1} = (axa^{-1})^p aya^{-1}$. And since $[P,P]$ is normal, we have that $(axa^{-1})^p aya^{-1} \in P^p[P,P]$. Thus we know that $P^p[P,P] \lhd P$.

Now let $V := P/P^p[P,P]$. Once again, since $[P,P] < P^p[P,P]$ we know that V is abelian. Plus, V is a finite group. And since $x^p \in P^p[P,P]$ for all $x \in P$, we know that $o(v) = p$ for every nontrivial element $v \in V$.

Now it is convenient to switch to using additive notation on V.

Define $s : \mathbb{Z}/p\mathbb{Z} \times V \to V$ by $s(k + p\mathbb{Z}, v) = kv$. This formula is well defined because $o(v) = p$ for all $v \neq 0$. Also, we can easily see that s satisfies the axioms of a scalar multiplication function on V. Hence, we may view V as a $(\mathbb{Z}/p\mathbb{Z})$-vector space. (Also note that V is finite-dimensional since V is a finite group.)

By a standard result of linear algebra, for any $v \in V - \{0\}$ we can find a vector subspace $W \subseteq V$ with codimension 1 (meaning it has one less dimension then V) such that $v \notin W$. But then V/W is spanned by $v + W$. So, $|V/W| = p$ and we've shown that $|V/W| \cong \mathbb{Z}/p\mathbb{Z}$. And since the only subgroups of V/W are $\{0 + W\}$ and V/W we in turn know that $W \in \mathrm{Max}(V)$.

By the correspondance theorem, we get that $\pi^{-1}(W)$ is maximal in P where $\pi : P \to V = P/(P^p[P,P])$ is the surjective natural quotient map. But crucially $\pi^{-1}(\{v\}) \notin \pi^{-1}(W)$. And since $\pi(x) \neq 0$ for any $x \notin P^p[P,P]$ and v in our prior construction was arbitrary, this proves that we can always find a maximal subgroup M of P such that $x \notin M$ whenever $x \notin P^p[P,P]$. Hence:

$$\Phi(P) = \textstyle\bigcap_{M \in \mathrm{Max}(P)} M \subseteq P^p[P,P].$$

(d) Suppose $P = \langle S \rangle$ and a proper subset of S does not generate P. Prove that:

$$|S| = \dim_{\mathbb{Z}/p\mathbb{Z}}(P/(P^p[P,P])) = \dim_{\mathbb{Z}/p\mathbb{Z}}(P/(\Phi(P))).$$

We showed in the last part that $P/(P^p[P,P])$ is a finite dimensional $(\mathbb{Z}/p\mathbb{Z})$-vector space. Now let $\pi : P \to P/(P^p[P,P])$ be the natural projective map. Then note that $P = \langle S \rangle \implies P/(P^p[P,P]) = \langle \pi(S) \rangle$. In particular, since scalar multiplication of $v \in P/(P^p[P,P])$ by $k + \mathbb{Z}/p\mathbb{Z}$ is equivalent to just adding v to itself k times, we know that the linear span of $\{\pi(S)\}$ is all of $P/(P^p[P,P])$. It follows that there is a $(\mathbb{Z}/p\mathbb{Z})$-basis $\overline{S'} \subseteq \pi(S)$ for $P/(P^p[P,P])$.

Suppose $S' \subseteq S$ satisfies that $|S'| = \overline{S'}$ and $\pi(S') = \overline{S'}$. By part (b) of problem 1, we know that $\langle \pi(S') \rangle = P/\Phi(P)$ iff $\langle S' \rangle = P$. But then as any proper subset of S doesn't generate P, we must have that $S' = S$.

This proves that $|S| = |\overline{S'}| = \dim_{\mathbb{Z}/p\mathbb{Z}}(P/(\Phi(P)))$. ■

The T.L.D.R. of this proof is that if P is a finite p-group, then all minimal generating sets of P must have the same cardinality. Note that this result generally doesn't hold for non-p-groups.

Set 7 Problem 5: Suppose G is a group such that $a, b \in G$, $a^2 = 1 = b^2$, and $G = \langle a, b \rangle$.

(a) Prove that G is solvable.

Let $N = \langle ab \rangle = \{(ab)^n : n \in \mathbb{Z}\}$. Then as $a = a^{-1}$ and $b = b^{-1}$, we have that:

- $a(ab)^n a^{-1} = b(ab)^{n-1}a = (ba)^n = ((ab)^{-1})^n = (ab)^{-n}$,
- $b(ab)^n b^{-1} = b(ab)^n b = b(ab)^{n-1}a = (ba)^n = ((ab)^{-1})^n = (ab)^{-n}$.

In particular, this means that $a, b \in N_G(N)$. And since a and b generate G, this shows that $N_G(N) = G$. Hence, we've proven that $N \lhd G$. And as N is cyclic, we know that $N/\{1\}$ is abelian.

Furthermore, note that since G is generated by a and b, we have that G/N is generated by aN and bN. But then $b = (ba)a$ and in turn:

$$bN = (ba)N \cdot aN = N \cdot aN = N.$$

Also, as $a^2 = 1$, we know that $(aN)^2 = 1$. Hence, $G/N = \langle aN \rangle$ has order dividing 2. And this proves that G/N is cyclic (and thus abelian).

Since $G \rhd N \rhd \{1\}$ is a normal series such that G/N and $N/\{1\}$ are both abelian, we have that G is solvable.

(b) Show that G is not necessarily finite.

Recall the group D_∞ from *problem 5 of the sixth problem set*.

Set 7 Problem 6: Suppose G is a group, $a, b \in G$, $ab^2a^{-1} = b^3$, and $ba^2b^{-1} = a^3$. Then show that $a = b = 1$.

To start off, since $b^2 = a^{-1}b^3a$ and $a^2 = b^{-1}a^3b$ we have that:

- $a^2b^4a^{-2} = a^2(a^{-1}b^3a)^2a^{-2} = ab^6a^{-1} = a(a^{-1}b^3a)^3a^{-1} = b^9$,
- $a^2b^4a^{-2} = (b^{-1}a^3b)b^4(b^{-1}a^{-3}b) = b^{-1}a^3b^4a^{-3}b$.

But now $b^{-1}a^3b^4a^{-3}b = b^9 \implies a^3b^4a^{-3} = bb^9b^{-1} = b^9 = a^2b^4a^{-2}$. In turn, we can conclude that $ab^4a^{-1} = b^4$. And this lets us show that $b^3 = 1$ and $b^2 = 1$ as follows:

- Note that $b^6 = (b^3)^2 = (ab^2a^{-1})^2 = ab^4a^{-1} = b^4$. Therefore $b^2 = 1$.
- Note that $ab^4a^{-1} = b^4 \implies a^2b^4a^{-2} = ab^4a^{-1}$. But we've already shown that $b^6 = ab^4a^{-1}$ and $b^9 = a^2b^4a^{-2}$. Thus $b^6 = b^9$ and we've proven that $1 = b^3$.

Since $b^3 = 1 = b^2$, we must have that $b = 1$. And now by plugging that back into the identity $ba^2b^{-1} = a^3$, we get that $a^2 = a^3$. Hence, we also have that $1 = a$.

Math 220a Notes:

Given two closed curves $\gamma_0, \gamma_1 : [0,1] \to X$, we shall say that γ_0 and γ_1 are homotopic in X (denoted $\gamma_0 \sim_G \gamma_1$) if there is a continuous function $\Gamma : [0,1]^2 \to G$ such that $\gamma_0(s) = \Gamma(s, 0)$, $\gamma_1(s) = \Gamma(s, 1)$, and $\Gamma(0, t) = \Gamma(1, t)$ for all $t \in [0, 1]$.

Beware, this is a different definition from Munkres definition of being path homotopic. In fact, this would instead correspond to Munkre's definition of homotopy in general applied to functions mapping the circle S^{2-1} into G

If γ is a closed piecewise C^1 curve in G, then we write $\gamma \sim 0$ if γ is path homotopic to a constant curve.

<u>(Conway) Cauchy's Theorem (Third Version)</u> Let f be analytic in a region G and let $\gamma_0, \gamma_1 : [0,1] \to G$ be closed piecewise C^1 paths such that $\gamma_0 \sim_G \gamma_1$. Then $\int_{\gamma_0} f(z)\mathrm{d}z = \int_{\gamma_1} f(z)\mathrm{d}z$.

Proof:

Let $\Gamma : [0,1]^2 \to G$ be the homotopy.

Note that a difficulty with proving this theorem is that $\gamma_t(s) \coloneqq \Gamma(s,t)$ is not guarenteed to be piecewise C^1 for any $t \neq 0, 1$.

Since Γ is continuous and $[0,1]^2$ is compact, we know that $\Gamma([0,1]^2)$ is compact in G. It follows that $\varepsilon = \inf\{|x-y| : x \in \Gamma([0,1]^2), y \in \mathbb{C} - G\} > 0$. It also follows that Γ is uniformly continuous. And by uniform continuity we can find n such that given any square $I_{j,k} \coloneqq [\frac{j}{n}, \frac{j+1}{n}] \times [\frac{k}{n}, \frac{k+1}{n}]$ we have that $\Gamma(I_{j,k}) \subseteq B_\varepsilon(z_{j,k}) \subseteq G$ for all $j, k \in \{0, \ldots, n-1\}$ where $z_{j,k} \coloneqq \Gamma(\frac{j}{n}, \frac{k}{n})$.

Now we approximate $\gamma_t(s) \coloneqq \Gamma(s,t)$ where $t = \frac{k}{n}$ by taking the closed polygonal path $P_k = [z_{0,k}, z_{1,k}] + \ldots + [z_{n-1,k}, z_{n,k}]$. Note that $[z_{j,k}, z_{j,k+1}] \subseteq B_\varepsilon(z_{j,k})$ for all j, k. Hence, $\{P_k\} \subseteq G$ for each k (meaning we can integrate f along these paths). Our claim is that:

$$\int_{\gamma_0} f\mathrm{d}z = \int_{P_0} f\mathrm{d}z = \int_{P_1} f\mathrm{d}z = \cdots = \int_{P_n} f\mathrm{d}z = \int_{\gamma_1} f\mathrm{d}z$$

<u>Part 1:</u> $\int_{\gamma_0} f\mathrm{d}z = \int_{P_0} f\mathrm{d}z$ and $\int_{\gamma_1} f\mathrm{d}z = \int_{P_n} f\mathrm{d}z$.

The proof of both equalities is the same so I'll focus on the first equation. Let $\gamma_0^{(j)}$ be the restriction of γ to $[\frac{j}{n}, \frac{j+1}{n}]$. Then after some rearranging we get that:

$$\int_{\gamma_0} f\mathrm{d}z - \int_{P_0} f\mathrm{d}z = \sum_{j=0}^{n-1} (\int_{\gamma_0^{(j)}} f\mathrm{d}z + \int_{[z_{j+1,0}, z_{j,0}]} f\mathrm{d}z)$$

But note that $\gamma_0^{(j)}$ starts and ends at $z_{j,0}$ and $z_{j+1,0}$ respectively. Thus $\gamma_0^{(j)} + [z_{j+1,0}, z_{j,0}]$ is a closed C^1 path. And as $\{\gamma_0^{(j)}\} \subseteq \Gamma(I_{j,k}) \subseteq B_\varepsilon(z_{j,0})$, we know that the trace of $\gamma_0^{(j)} + [z_{j+1,0}, z_{j,0}]$ is contained in a convex disc contained in G. So by Cauchy's theorem, we have that $(\int_{\gamma_0^{(j)}} f\mathrm{d}z + \int_{[z_{j+1,0}, z_{j,0}]} f\mathrm{d}z) = \int_{\gamma_0^{(j)} + [z_{j+1,0}, z_{j,0}]} f\mathrm{d}z = 0$ for all j.

<u>Part 2:</u> $\int_{P_k} f\mathrm{d}z = \int_{P_{k+1}} f\mathrm{d}z$ for all k.

Note that the polygon $Q_{j,k} \coloneqq [z_{j,k}, z_{j+1,k}, z_{j+1,k+1}, z_{j,k+1}, z_{j,k}] \subseteq B_\varepsilon(z_{j,k}) \subseteq G$ for all j, k. And as $B_\varepsilon(z_{j,k})$ is convex, we thus know by Cauchy's theorem that:

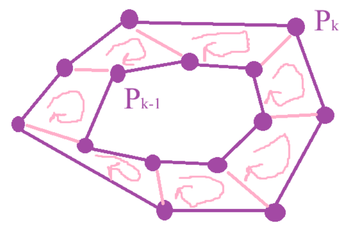
$$\int_{Q_{j,k}} f\mathrm{d}z = 0 \text{ for all } j, k.$$

But now note that after some rearranging we have that:

$$\begin{aligned}\int_{P_k} f\mathrm{d}z - \int_{P_{k+1}} f\mathrm{d}z &= \int_{[z_{n,k}, z_{n,k+1}]} f\mathrm{d}z - \int_{[z_{0,k}, z_{0,k+1}]} f\mathrm{d}z + \sum_{j=1}^{n-1} \int_{Q_{j,k}} f\mathrm{d}z \\ &= \int_{[z_{n,k}, z_{n,k+1}]} f\mathrm{d}z - \int_{[z_{0,k}, z_{0,k+1}]} f\mathrm{d}z + 0\end{aligned}$$

But as $\Gamma(1,t) = \Gamma(0,t)$ for all $t \in [0,1]$ we know that $z_{0,k} = z_{n,k}$ and $z_{0,k+1} = z_{n,k+1}$. Therefore, $\int_{[z_{n,k}, z_{n,k+1}]} f\mathrm{d}z - \int_{[z_{0,k}, z_{0,k+1}]} f\mathrm{d}z = 0$ as well.

Here is a picture to hopefully help describe this part:



Corollary: If $\gamma : [0,1] \to G$ is a closed piecewise C^1 curve and $\gamma \sim_G 0$, then $n(\gamma; a) = 0$ for all $a \in \mathbb{C} - G$.

Proof:

Just apply the previous theorem to the function $f(z) = \frac{1}{2\pi i(z-a)}$. Then as any path integral along a constant curve always evaluates to zero, we are done.

(Conway) Cauchy's Theorem (Second Version) If $f : G \to \mathbb{C}$ is an analytic function and γ is closed C^1 curve in G with $\gamma \sim_G 0$, then $\int_\gamma f = 0$.

Proof:

Apply Cauchy's integral theorem plus the last corollary.

Corollary: If $G \subseteq \mathbb{C}$ is open and simply connected, then $\int_\gamma f\mathrm{d}z = 0$ for any closed piecewise C^1 curve γ in G and analytic function f on G.

Munkres definition of being path homotopic (see *page 117*) is equivalent to Conway's definition of being Fixed-End-Point (F.E.P.) homotopic. Note that if γ_1 and γ_2 are closed curves rooted at the same point, then γ_1, γ_2 being F.E.P. homotopic implies $\gamma_1 \sim_G \gamma_2$. Also note that if γ_1 and γ_2 are F.E.P. homotopic then $\gamma_1 + (-\gamma_2)$ is F.E.P. homotopic to a constant curve. In turn, we get the following theorem:

Independence of Path Theorem: If γ_0, γ_1 are two piecewise C^1 curves in an open set $G \subseteq \mathbb{C}$ from a to b and $\gamma_0 \sim_G \gamma_1$, then $\int_{\gamma_0} f = \int_{\gamma_1} f$ for any analytic function f on G.

Proof:

$\int_{\gamma_0} f\mathrm{d}z - \int_{\gamma_1} f\mathrm{d}z = \int_{\gamma_0 + (-\gamma_1)} f\mathrm{d}z = 0$ by the last corollary.

When G is simply connected (so that all curves in G from a point a to a point b are path homotopic), we thus have that $\int_\gamma f$ depends only on the endpoints of γ and not on the particular path taken. This has the following consequences:

Theorem: If G is simply connected then every analytic function f has a primitive F.

Proof:

Fix $a \in G$ and then for every $z \in G$ define $F(z) = \int_{\gamma_z} f\mathrm{d}w$ where γ_z is any piecewise C^1 curve from a to z.

Recall from *theorem II.2.3* on page 247 that if G is connected then we can always find a polygonal path in G going between any two points of G. Thus, we don't need to worry about if a piecewise C^1 curve from a to z exists.

We claim F is a primitive of f. After all, given any fixed z_0, let $r > 0$ be such that $B_r(z_0) \subseteq G$. Now by the corollary following Cauchy's theorem (second version), since $\gamma_z + [z, z_0] + (-\gamma_{z_0})$ is a closed piecewise C^1 curve in G for any arbitrary piecewise C^1 curves γ_z and γ_{z_0} in G going from a to z and z_0 respectively, we know that:

$$F(z) + \int_{[z,z_0]} f\mathrm{d}w - F(z_0) = 0 \text{ for all } z \in B_r(z_0).$$

In other words, $\frac{F(z)-F(z_0)}{z-z_0} = \frac{1}{z-z_0}\int_{[z_0,z]} f(w)\mathrm{d}w$. Then after subtracting $f(z_0)$ from both sides we get that:

$$\frac{F(z)-F(z_0)}{z-z_0} - f(z_0) = \frac{1}{z-z_0}\int_{[z_0,z]} f(w) - f(z_0)\mathrm{d}w$$

Finally, since f is continuous at z_0, we know for any $\varepsilon > 0$ that there exists $0 < \delta < r$ such that when $|w - z_0| < \delta$ then $|f(w) - f(z_0)| < \varepsilon$. In turn, for all $z \in B_\delta(z_0)$ we have that:

$$|\frac{F(z)-F(z_0)}{z-z_0} - f(z_0)| \leq \frac{1}{|z-z_0|}\int_{[z_0,z]} |f(w) - f(z_0)|\mathrm{d}w < \frac{1}{|z-z_0|} \cdot |z_0 - z|\varepsilon = \varepsilon.$$

This proves that F is differentiable at z_0 with $F'(z_0) = f(z_0)$. ■

Theorem: If $G \subseteq \mathbb{C}$ is simply connected and f is an analytic nowhere vanishing function in G, then there exists a branch of $\log(f)$ on G (i.e. an analytic function g on G such that $e^{g(z)} = f(z)$).

Proof:

Since $f \neq 0$ in G, we know $\frac{f'}{f}$ is analytic on G. Hence by the prior theorem there exists $g : G \to \mathbb{C}$ such that $g' = \frac{f'}{f}$.

Next, pick $z_0 \in G$ and $w_0 \in \mathbb{C}$ such that $f(z_0) = e^{w_0}$. Since g will still be a primitive even after adding a constant, we can without loss of generality assume $g(z_0) = w_0$. That way, $f(z_0) = e^{g(z_0)}$.

Finally, consider $h(z) = e^{g(z)}$. Then:

$$(\frac{h}{f})' = \frac{h'f - hf'}{f^2} = \frac{g'e^g f - hf'}{f^2} = \frac{g'h}{f} - \frac{h}{f}\frac{f'}{f} = \frac{h}{f}(g' - \frac{f'}{f}) = \frac{h}{f}(0) = 0$$

Since G is connected, this shows that $\frac{h}{f}$ is constant on G. And since $\frac{h(z_0)}{f(z_0)} = 1$, we've proven that $h = f$ everywhere on G. ■

Math 200a Notes:

Given any integer $k > 0$, we let F_k denote the free group generated by k elements.

Theorem: $\langle \left[\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right] \rangle \cong F_2$.

Proof:

Let $G = \langle \left[\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right] \rangle$. Then note that $G \curvearrowright \mathbb{R}^2$ by linear transformations. In particular, if ℓ is a line passing through 0, then each element of G sends ℓ to another line. So, we can actually say that $G \curvearrowright X := \mathbb{RP}$.

Next, let $G_1 = \langle \left[\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right] \rangle$ and $G_2 = \langle \left[\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right] \rangle$. Then note that $\left[\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right]^n = \left[\begin{smallmatrix} 1 & 2n \\ 0 & 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right]^n = \left[\begin{smallmatrix} 1 & 0 \\ 2n & 1 \end{smallmatrix}\right]$ (you can easily show this via induction).

Thus, $G_1 = \{\left[\begin{smallmatrix}1 & 2n\\ 0 & 1\end{smallmatrix}\right] : n \in \mathbb{Z}\}$ and $G_2 = \{\left[\begin{smallmatrix}1 & 0\\ 2n & 1\end{smallmatrix}\right] : n \in \mathbb{Z}\}$. In particular, this means $G_1 \cong \mathbb{Z}$, $G_2 \cong \mathbb{Z}$.

Recall from *page 336* that any line in $\mathbb{RP}$ passing through (x, y) can be uniquely represented by the homogeneous coordinates $[x : y] = x/y$. Then as $\left[\begin{smallmatrix}1 & 2n\\ 0 & 1\end{smallmatrix}\right]\left[\begin{smallmatrix}x\\ y\end{smallmatrix}\right] = \left[\begin{smallmatrix}x+2ny\\ y\end{smallmatrix}\right]$, we have that $\left[\begin{smallmatrix}1 & 2n\\ 0 & 1\end{smallmatrix}\right][1 : 0] = [1 : 0]$ and $\left[\begin{smallmatrix}1 & 2n\\ 0 & 1\end{smallmatrix}\right][k : 1] = [k + 2n : 1]$.

Similarly, we have that $\left[\begin{smallmatrix}1 & 0\\ 2n & 1\end{smallmatrix}\right][0 : 1] = [0 : 1]$ and $\left[\begin{smallmatrix}1 & 2n\\ 0 & 1\end{smallmatrix}\right][1 : k] = [1 : k + 2n]$. So finally, let $X_1 = \{[1 : 0]\} \cup \{[k : 1] : |k| \geq 1\}$ and $X_2 = \{[0 : 1]\} \cup \{[1 : k] : |k| \geq 1\}$.

If $g \in G_1 - \{\left[\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right]\}$, then $g \cdot X_2 \subseteq X_1$. (After all, $[1 : k] = [1/k : 1]$ and $|x + 2n| \geq 1$ for all $n \in \mathbb{Z}$ if $|x| \leq 1$). Similarly, $(G_2 - \{\left[\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right]\}) \cdot X_1 \subseteq X_2$.

By the ping pong lemma we conclude:

$$G = \langle \left[\begin{smallmatrix}1 & 2\\ 0 & 1\end{smallmatrix}\right], \left[\begin{smallmatrix}1 & 0\\ 2 & 1\end{smallmatrix}\right] \rangle = \langle G_1, G_2 \rangle \cong G_1 * G_2 \cong \mathbb{Z} * \mathbb{Z} = F_2. \blacksquare$$

$\mathrm{SL}_n(\mathbb{Z})$ refers to the collection of nxn matrices with determinant 1 and integer coefficients. At least for $\mathrm{SL}_2(\mathbb{Z})$ I already know how to show that $\mathrm{SL}_2(\mathbb{Z})$ is a group with respect to matrix multiplication.

In slightly more generallity, given any commutative ring R, the formula for matrix multiplication and the determinant of a matrix can still be carried out in R and the formula for the determinant of a matrix in R still makes sense. It follows that we can define $\mathrm{SL}_n(R)$ to be the collection of nxn matrices with determinant $1 \in R$ and coefficients in R.

Again, I don't know enough linear algebra to prove $\mathrm{SL}_n(R)$ is a group for arbitrary n. That said, if $n = 2$ then it is easy to see that $\mathrm{SL}_2(R)$ is a group.

- $$\begin{aligned}\det(\left[\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right]\left[\begin{smallmatrix}e & f\\ g & h\end{smallmatrix}\right]) &= (ae + bg)(cf + dh) - (af + bh)(ce + dg)\\ &= aecf + adeh + bgcf + bgdh - afce - afdg - bhce - bhdg\\ &= adeh + bgcf - afdg - bhce\\ &= ad(eh - fg) - bc(eh - fg) = \det(\left[\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right])\det(\left[\begin{smallmatrix}e & f\\ g & h\end{smallmatrix}\right]) = 1.\end{aligned}$$
- $\left[\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right]\left[\begin{smallmatrix}d & -b\\ -c & a\end{smallmatrix}\right] = \left[\begin{smallmatrix}ad-bc & -ba+ab\\ cd-dc & ad-bc\end{smallmatrix}\right] = \left[\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right]$.

Next we define $\mathrm{PSL}_2(\mathbb{Z}) \coloneqq SL_2(\mathbb{Z})/\{\pm I\}$. Note that $\{\pm I\} = Z(SL_2(\mathbb{Z}))$ and is thus a normal subgroup. Hence, $\mathrm{PSL}_2(\mathbb{Z})$ is well-defined. Also we denote $\overline{A} = A\{\pm I\} \in \mathrm{PSL}_2(\mathbb{Z})$. Note that $\overline{A} = \{A, -A\}$.

<u>Theorem:</u> $\langle \overline{\left[\begin{smallmatrix}1 & 2\\ 0 & 1\end{smallmatrix}\right]}, \overline{\left[\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}\right]} \rangle \cong \mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$.

Proof:

Let $G_1 = \langle \overline{\left[\begin{smallmatrix}1 & 2\\ 0 & 1\end{smallmatrix}\right]} \rangle$ and $G_2 = \langle \overline{\left[\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}\right]} \rangle$. We already know from the last proof that:

$$G_1 = \left\{\overline{\left[\begin{smallmatrix}1 & 2n\\ 0 & 1\end{smallmatrix}\right]} : n \in \mathbb{Z}\right\} \cong \mathbb{Z}.$$

Meanwhile, $(\overline{\left[\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}\right]})^2 = \overline{\left[\begin{smallmatrix}-1 & 0\\ 0 & -1\end{smallmatrix}\right]} = \overline{I}$. Thus $G_2 \cong \mathbb{Z}/2\mathbb{Z}$.

Next, note that $\mathrm{PSL}_2(\mathbb{R}) \curvearrowright H := \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ by Möbius transformations (recall *problem 3 on the second set*).

In particular, $\overline{\left[\begin{smallmatrix} 1 & 2n \\ 0 & 1 \end{smallmatrix}\right]} \cdot z = z + 2n$ and $\overline{\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]} \cdot z = \frac{-1}{z}$. Thus $(G_1 - \{\overline{I}\}) \cdot X_2 \subseteq X_1$ and $(G_2 - \{\overline{I}\}) \cdot X_1 \subseteq X_2$ where $X_1 = \{z \in H : |z| > 1\}$ and $X_2 = \{z \in H : |z| < 1\}$.

By the ping pong lemma we are done. ■

We say a group Γ is residually $\mathcal{C}$ if for all $x \in \Gamma - \{1\}$ there exists a finite group G which satisfies $\mathcal{C}$ and a group homomorphism $\phi : \Gamma \to G$ such that $\phi(x) \neq 1$.

We say Γ is residually finite if $\forall x \in G - \{1\}$ there exists a finite group G and a group homomorphism $\phi : \Gamma \to G$ such that $\phi(x) \neq 1$.

(By first isomorphism theorem, this is equivalent ot saying that for all $x \in \Gamma - \{1\}$ there exists a group $N \lhd \Gamma$ of finite index such that $x \notin N$.)

Theorem: F_2 is residually finite.

Proof:

Recall $F_2 \cong \langle \left[\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right] \rangle \subseteq \mathrm{SL}_2(\mathbb{Z})$. Thus, we can define a group homomorphism $\phi_n : F_2 \to \mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$ such that $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \mapsto \left[\begin{smallmatrix} \overline{a} & \overline{b} \\ \overline{c} & \overline{d} \end{smallmatrix}\right]$ where $\overline{x} = x + n\mathbb{Z}$.

Since the mod map preserves addition and multiplication, it's clear that $\phi_n(AB) = \phi_n(A)\phi_n(B)$ and that:

$$\det(A) = 1 \in \mathbb{Z} \Longrightarrow \det(\phi_n(A)) = 1 \in \mathbb{Z}/n\mathbb{Z}.$$

Hence ϕ_n is a well-defined group homomorphism into $\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$.

But now $\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$ has less than n^4 elements. Also, if $x \in F_2 - \{I\}$ then we can choose n large enough so that $\phi_n(x) \neq \left[\begin{smallmatrix} \overline{1} & \overline{0} \\ \overline{0} & \overline{1} \end{smallmatrix}\right]$. ■

A group Γ is virtually $\mathcal{C}$ if there exists $\Lambda < \Gamma$ with finite index such that Λ satisfies $\mathcal{C}$.

Γ is virtually solvable if there exists $\Lambda \lhd \Gamma$ such that $[\Gamma : \Lambda] < \infty$ and Λ is solvable.

Note that it is not a restriction to assume Λ is a normal subgroup. After all, suppose $\Lambda < \Gamma$ and Λ is solvable. Then if we consider the group action $\Gamma \curvearrowright \Gamma/\Lambda$ by left translation, we get a group homomorphism $\phi : \Gamma \to S_{\Gamma/\Lambda}$. In turn, $\mathrm{core}_\Gamma(\Lambda) = \ker(\phi)$ is a normal subgroup of Γ whose index is finite as $|\mathrm{im}(\phi)|$ divides $[\Gamma : \Lambda]! < \infty$. And as $\mathrm{core}_\Gamma(\Lambda) < \Lambda$ we know that $\mathrm{core}_\Gamma(\Lambda)$ is solvable.

One other observation: If Γ is virtually solvable then so is any quotient of Γ.

Why?

Consider any subgroup $N \lhd \Gamma$. Then $\Lambda N/N \cong \Lambda/N \cap \Lambda$, and the latter is solvable. Hence $\Lambda N/N$ is solvable (see *problem 3 on the sixth set*). At the same time, $(\Lambda N)/N \lhd \Gamma/N$ as Λ and N are both normal subgroups of Γ. So, $(\Gamma/N)/(\Lambda N/N) \cong \Gamma/(\Lambda N)$ and the latter clearly has less elements than the finitely many in Γ/Λ. So, $(\Lambda N)/N$ satisfies the requirements for Γ/N to be virtually solvable.

Lemma: F_2 is not virtually solvable.

Proof:

Recall that $N \triangleleft S_n$ with N solvable implies that $N = \{1\}$ when $n \geq 5$. That said, we also have that $S_n = \langle (1\ \ 2), (1\ \ 2\ \ \cdots\ \ n) \rangle$. (For a proof of this see page 53 of my math 100c notes.) So, by the universal property of free groups we know there exists a surjective group homomorphism $\Phi : F_{\{a,b\}} \to S_n$ such that $\phi(a) = (1\ \ 2)$ and $\phi(b) = (1\ \ 2\ \ \cdots\ \ n)$.

($F_{\{a,b\}}$ is just notation for F_2 that makes it explicit what the generators of F_2 are...)

Suppose there exists $\Lambda \triangleleft F_{\{a,b\}}$ such that $[F_{\{a,b\}} : \Lambda] = m$ and Λ is solvable. Then we'd have that $\Phi(\Lambda)$ is solvable and $\Phi(\Lambda) \triangleleft S_n$. And when $n \geq 5$, this means that $\Phi(\Lambda) = \{\mathrm{Id}\}$. So, $\Lambda < \ker(\Phi)$ and in turn there is a surjective mapping:

$$F_{\{a,b\}}/\Lambda \twoheadrightarrow F_{\{a,b\}}/\ker(\Phi) \cong \mathrm{im}(\Phi) = S_n.$$

Consequently, we must have that $m \geq n!$ for any $n \geq 5$. This is a contradiction. ■

If G is a group and $R \subseteq G$, then we say $\langle\!\langle R \rangle\!\rangle$ is the smallest normal subgroup of G containing R. Next, given the sets S and $R \subseteq \mathcal{F}(S)$ (where $\mathcal{F}(S)$ is the free group of S), we define $\langle S | R \rangle \coloneqq \mathcal{F}(S)/\langle\!\langle R \rangle\!\rangle$. Also, we call $\langle S|R\rangle$ a presentation.

In other words, $\langle\!\langle R \rangle\!\rangle$ is the set of all words in $\mathcal{F}(S)$ identified with 1. Also, note that a common abuse of notation is to list an element of R as "word 1" = "word 2" as opposed to ("word 1")("word 2")$^{-1}$. Given this abuse of notation, it shouldn't be surprising that we call R the set of defining relations of $\langle S|R\rangle$.

When trying to prove what group a presentation is isomorphic to, there is a general procedure that works.

1. Already have an idea that $\langle S|R\rangle \cong G$. (Unfortunately, this procedure can only verify hunches one already has).
2. Let $S' \subseteq G$ be a generating set for G such that there exists a bijection $f : S \to S'$. Then using the universal property of free groups, let $\Phi : \mathcal{F}(S) \to G$ be a group homomorphism such that $\Phi(x) = f(x)$ for all $x \in S$. This group homomorphism is a surjection.
3. Check the relations to make sure that $R < \ker(\Phi)$. That way, we know that $\langle\!\langle R \rangle\!\rangle < \ker(\Phi)$. And in turn, there is a well-defined surjective group homomorphism $\overline{\Phi} : \langle S|R\rangle \to G$ such that $\overline{\Phi}(x) = \Phi(x) = f(x)$ for all $x \in S$.
4. Finally, find a trick to show that $\overline{\Phi}$ is injective.

Example 1: $\langle x | x^n = 1\rangle \cong C_n$.

Let a be a generator for C_n. Then there is surjective homomorphism $\Phi : \mathcal{F}(\{x\}) \to C_n$ given by $\Phi(x) = a$. Also, it is clear that $\Phi(x^n) = a^n = 1$. So $\langle\!\langle x^n \rangle\!\rangle \subseteq \ker(\Phi)$ and we can define a surjective group homomorphism $\overline{\Phi} : \langle x | x^n = 1\rangle \to C_n$ such that $\overline{\Phi}(x) = a$. Finally, note that $|\langle x|x^n = 1\rangle| = n = |C_n|$. So by pigeonhole we know $\overline{\Phi}$ is a bijection.

Example 2: $\langle x, y \mid x^n = 1, y^2 = 1, yxy = x^{-1} \rangle \cong D_{2n}$.

Show this yourself. The proof is mostly identical to the prior example. :p

---

Set 8 Problem 1: Prove that $\langle a, b \mid [a, b] \rangle \cong \mathbb{Z} \times \mathbb{Z}$.

By the universal property of free groups, we know there is a group homomorphism $f : F_{\{a,b\}} \to \mathbb{Z} \times \mathbb{Z}$ such that $f(a) = (1, 0)$ and $f(b) = (0, 1)$. Furthermore, we then have that $f([a, b]) = f(a)f(b)f(a^{-1})f(b^{-1}) = (1, 0) + (0, 1) - (1, 0) - (0, 1) = 0$. Hence, by quotienting out $\ll [a, b] \gg$ we can get a well-defined group homomorphism:

$\widetilde{f} : \langle a, b \mid [a, b] \rangle \to \mathbb{Z} \times \mathbb{Z}$ such that $f(a) = (1, 0)$ and $f(b) = (0, 1)$.

Also note that as $\langle (1, 0), (0, 1) \rangle = \mathbb{Z} \times \mathbb{Z}$, we know that f and in turn $\widetilde{f}$ are surjective.

What's left to show is that $\widetilde{f}$ is a bijection. So first we note that the following relevant commutators are in $\ll [a, b] \gg$.

$[b, a] = ([a, b])^{-1}$, $a^{-1}[a, b]a = [b, a^{-1}]$, $b^{-1}[a, b]b = [b^{-1}, a]$, and $(a^{-1}b^{-1})[b, a]ba = [b^{-1}, a^{-1}]$.

This shows that $a^{e_1}b^{e_2} = b^{e_2}a^{e_1}$ where $e_1, e_2 \in \{\pm 1\}$. Then by induction on k we can conclude that for all $k \in \mathbb{N}$:

- $b^k a = b^{k-1}ba(b^{-1}a^{-1}ab) = b^{k-1}ab = ab^{k-1}b = ab^k$,
- $b^{-k}a = b^{-k+1}b^{-1}a(ba^{-1}ab^{-1}) = b^{-k+1}ab^{-1} = ab^{-k+1}b^{-1} = ab^{-k}$,
- $b^k a^{-1} = b^{k-1}ba^{-1}(b^{-1}aa^{-1}b) = b^{k-1}a^{-1}b = a^{-1}b^{k-1}b = a^{-1}b^k$,
- $b^{-k}a^{-1} = b^{-k+1}b^{-1}a^{-1}(baa^{-1}b^{-1}) = b^{-k+1}a^{-1}b^{-1} = a^{-1}b^{-k+1}b^{-1} = a^{-1}b^{-k}$.

Another round of induction then shows that $a^m b^n = b^n a^m$ for all $m, n \in \mathbb{Z}$. And finally, this lets us show (again through induction) that every element of $\langle a, b \mid [a, b] \rangle$ can be represented by a word of the form $a^m b^n$ where $m, n \in \mathbb{Z}$.

We also claim that $a^m b^n = 1$ iff $m = 0 = n$. To see this, note that we can define the "a-power" and "b-power" of any word in $F_{\{a,b\}}$ by adding up the powers of all the a terms and b terms respectively.

Technically I'm overlooking the fact that the elements of $F_{\{a,b\}}$ are equivalence classes of words. That said, the two manipulations that let you go between any two words in the same equivalence class preserve "a-power" and "b-power". So, this technicality doesn't really matter.

But now if we let $N \subseteq F_{\{a,b\}}$ be the collection of all words with an a-power and b-power of 0, then we have that N is closed under word concatenation, inversing, and conjugation. Also $[a, b] \in N$. So $\ll [a, b] \gg < N \lhd F_{\{a,b\}}$. And in turn, we know that if $a^m b^n = 1$ in $\langle a, b \mid [a, b] \rangle$ then we must have that $a^m b^n \in N$ when considered as an element of $F_{\{a,b\}}$. But that implies that $m = 0 = n$.

Consequently, we know that if $a^{m_1}b^{n_1} = a^{m_2}b^{n_2}$ then $m_1 = m_2$ and $n_1 = n_2$. Hence by all the prior reasoning, if we define $g : \mathbb{Z} \times \mathbb{Z} \to \langle a, b \mid [a, b] \rangle$ by $g(m, n) = a^m b^n$ then we know g is an injective and surjective function satisfying that $\widetilde{f} \circ g = \mathrm{Id}_{\mathbb{Z} \times \mathbb{Z}}$. In turn, $\widetilde{f} = g^{-1}$ and this proves that $\widetilde{f}$ is a bijection. ■

Set 8 Problem 2: Suppose X_1 and X_2 are two disjoint sets. Prove that:
$$\langle X_1 \mid R_1\rangle * \langle X_2 \mid R_2\rangle \cong \langle X_1 \cup X_2 \mid R_1 \cup R_2\rangle$$

I shall start by proving something I think my professor meant for me to take as obvious. Consider the natural inclusion maps $j_i : \mathcal{F}(X_i) \hookrightarrow \mathcal{F}(X_1 \cup X_2)$ for each i.

To see that each j_i is an injection, note that adding symbols to an alphabet X does not change the reduced form of words already in $\mathcal{F}(X)$. And since each word equivalence class in $\mathcal{F}(X_i)$ has a unique reduced form (see *page*___ for more on this), we know that j_i is an embedding. That said, the fact that j_i is an injection isn't important to the proof.

Then we know that $j_i^{-1}(\ll j_i(R_i)\gg)$ is a normal subgroup of $\mathcal{F}(X_i)$ containing R_i. Hence, there is a well-defined map $\overline{j_i} : \langle X_i \mid R_i\rangle \to \langle X_1 \cup X_2 \mid j_i(R_i)\rangle$ such that:
$$\overline{j_i}(\omega \ll R_i \gg) = j_i(\omega) \ll j_i(R_i) \gg \text{ for all words } \omega.$$

Furthermore, since $\ll j_i(R_i)\gg \subseteq \ll j_1(R_1) \cup j_2(R_2)\gg$ for both i, we know that there are well defined maps $k_i : \langle X_1 \cup X_2 \mid j_i(R_i)\rangle \to \langle X_1 \cup X_2 \mid j_1(R_1) \cup j_2(R_2)\rangle$ with $k_i(\omega \ll j_i(R_i)\gg) = \omega \ll j_1(R_1) \cup j_2(R_2)\gg$ for all words ω.

Now by setting $\theta_i = k_i \circ \overline{j_i}$ for both i, we now have shown that the obvious inclusion function $\langle X_i \mid R_i\rangle \to \langle X_1 \cup X_2 \mid j_1(R_1) \cup j_2(R_2)\rangle$ given by $\theta_i(\omega) = j_i(\omega) \ll j_1(R_1) \cup j_2(R_2)\gg$ is a well-defined group homomorphism.

With that out of the way I'm now going to identify $j_i(\omega)$ with ω for all $\omega \in \mathcal{F}(X_i)$. Also, I'll just write $\theta_i(\omega)$ as ω.

By the universal property of free products there exists a group homomorphism $\theta : \langle X_1 \mid R_1\rangle * \langle X_2 \mid R_2\rangle \to \langle X_1 \cup X_2 \mid R_1 \cup R_2\rangle$ such that $\theta(x_1) = x_1$ and $\theta(x_2) = x_2$ in $\langle X_1 \cup X_2 \mid R_1 \cup R_2\rangle$ for all $x_1 \in X_1$ and $x_2 \in X_2$.

Meanwhile, by the universal property of free groups there exists a group homomorphism $\phi : \mathcal{F}(X_1 \cup X_2) \to \langle X_1 \mid R_1\rangle * \langle X_2 \mid R_2\rangle$ such that $\phi(x_1) = x_1$ and $\phi(x_2) = x_2$ in $\langle X_1 \mid R_1\rangle * \langle X_2 \mid R_2\rangle$ for all $x_1 \in X_1$ and $x_2 \in X_2$. Also, note that if $\omega \in R_1 \cup R_2$ then $\phi(\omega) = 1$. Hence, by quotienting out $\ll R_1 \cup R_2\gg$ we get a well-defined map
$$\widetilde{\phi} : \langle X_1 \cup X_2 \mid R_1 \cup R_2\rangle \to \langle X_1 \mid R_1\rangle * \langle X_2 \mid R_2\rangle$$

...with $\widetilde{\phi}(x_1) = x_1$ and $\widetilde{\phi}(x_2) = x_2$ in $\langle X_1 \mid R_1\rangle * \langle X_2 \mid R_2\rangle$ for all $x_1 \in X_1$ and $x_2 \in X_2$.

Finally, $\widetilde{\phi} \circ \theta(x) = x$ and $\theta \circ \widetilde{\phi}(x) = x$ for all $x \in X_1 \cup X_2$. And as $X_1 \cup X_2$ is a generating subset of both $\langle X_1 \mid R_1\rangle * \langle X_2 \mid R_2\rangle$ and $\langle X_1 \cup X_2 \mid R_1 \cup R_2\rangle$, we can extrapolate that $\widetilde{\phi} \circ \theta = \mathrm{Id}$ and $\theta \circ \widetilde{\phi} = \mathrm{Id}$. So, θ and $\widetilde{\phi}$ are isomorphisms. ■

Set 8 Problem 3: Prove that the subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ which is generated by $\overline{\left[\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right]}$ and $\overline{\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]}$ has the presentation $\langle a, b \mid b^2\rangle$.

Recall from *page 417* that $\langle \overline{\left[\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right]}$ and $\overline{\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]}\rangle = \mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$. Then as $\mathbb{Z} \cong \langle a \mid \emptyset\rangle$ and $\mathbb{Z}/2\mathbb{Z} \cong \langle b \mid b^2\rangle$, we have by the prior problem that $\mathbb{Z} * \mathbb{Z}/2\mathbb{Z} \cong \langle a, b \mid b^2\rangle$. ■

Before moving on to the next problem, I want to show that $\mathrm{PSL}_2(\mathbb{Z})$ is generated by the matrices $\sigma := \overline{\left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]}$ and $\tau := \overline{\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]}$.

Why?

Note that $\sigma^n = \overline{\left[\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right]}$. In turn, given any matrix $\overline{\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]} \in \mathrm{PSL}_2(\mathbb{Z})$ we have that:
$$\sigma^n \overline{\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]} = \overline{\left[\begin{smallmatrix} a+nc & b+nd \\ c & d \end{smallmatrix}\right]} \text{ and } \tau \overline{\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]} = \overline{\left[\begin{smallmatrix} c & d \\ -a & -b \end{smallmatrix}\right]}.$$

This suggests the following construction. Suppose $\overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}$ is any matrix in $\mathrm{PSL}_2(\mathbb{Z})$ such that $|a| \geq |c| > 0$. Then we know there exists $n \in \mathbb{Z}$ such that $\sigma^n \overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} = \overline{\begin{bmatrix} a+nc & b+nd \\ c & d \end{bmatrix}} = \overline{\begin{bmatrix} a' & b' \\ c & d \end{bmatrix}}$ where $|a'| < |c'|$. (This is a consequence of the division algorithm). In turn:

$$\tau\sigma^n \overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} = \overline{\begin{bmatrix} c & d \\ -a' & -b' \end{bmatrix}}, \text{ where } |-a'| < c \leq |a|.$$

As for the case that $|a| < |c|$ initially, then we can just apply the prior reasoning to $\tau\overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}$. Either way, if $G := \langle \sigma, \tau \rangle \subseteq \mathrm{PSL}_2(\mathbb{Z})$ and $\overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} \in \mathrm{PSL}_2(\mathbb{Z})$, then we've proven that there is a matrix $g \in G$ such that $g\overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} = \overline{\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}}$ satisfies that $|c'| < c$.

By induction on $|c|$, we can thus conclude for any $\overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} \in \mathrm{PSL}_2(\mathbb{Z})$ that there exists $g_1, \ldots, g_n \in G$ such that $g_n \cdots g_1 \overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} = \overline{\begin{bmatrix} a' & b' \\ 0 & d' \end{bmatrix}}$.

But as $a'd' - 0b' = a'd' = 1$ and both a' and d' are integers, we may assume $a' = d' = 1$. Hence, we actually have that:

$$g_n \cdots g_1 \overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} = \overline{\begin{bmatrix} 1 & b' \\ 0 & 1 \end{bmatrix}} = \sigma^{b'}$$

And finally $\overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} = g_1^{-1} \cdots g_n^{-1} \sigma^{b'} \in G$. This proves that $\mathrm{PSL}_2(\mathbb{Z}) = \langle \overline{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}, \overline{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}} \rangle$.

Set 8 Problem 4: Prove that $\mathrm{PSL}_2(\mathbb{Z}) = \langle a, b \mid a^2, b^3 \rangle$.

Let $\sigma := \overline{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}$ and $\tau := \overline{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}$ like before. Then set $\omega = \sigma\tau$ and define $G_1 := \langle \tau \rangle$ and $G_2 := \langle \omega \rangle$.

Claim 1: $\langle G_1, G_2 \rangle = \langle \tau, \omega \rangle = \mathrm{PSL}_2(\mathbb{Z})$.

Why? We already know $\mathrm{PSL}_2(\mathbb{Z}) = \langle \tau, \sigma \rangle$. Also, $\tau = \tau^{-1}$. Therefore, $\sigma = \omega\tau$ is in $\langle \tau, \omega \rangle$. And this proves that:

$$\mathrm{PSL}_2(\mathbb{Z}) = \langle \tau, \sigma \rangle \subseteq \langle \tau, \omega \rangle \subseteq \mathrm{PSL}_2(\mathbb{Z})$$

Claim 2: $G_1 \cong C_2$ and $G_2 \cong C_3$.

Why? We already know from class that $\tau^2 = \overline{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}$. Meanwhile $\omega = \sigma\tau = \overline{\begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}}$. In turn, $\omega^3 = \overline{\begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}}\,\overline{\begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}}\,\overline{\begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}} = \overline{\begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}}\,\overline{\begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}} = \overline{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}$. And as $o(\omega)$ divides 3 and doesn't equal 1, we know $o(\omega) = 3$.

Now consider the action $\mathrm{PSL}_2(\mathbb{Z}) \curvearrowright \mathbb{R} \cup \{\infty\}$ by Möbius transformations.

In other words, $\overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} \cdot x = \frac{ax+b}{cx+d}$ for all $x \in \mathbb{R}$ and $\overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} \cdot \infty = \frac{a}{c}$ (and if any of the right-hand expressions are undefined, then $\overline{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}$ sends the element of $\mathbb{R} \cup \{\infty\}$ to ∞.)

Recall from *page 335* that if $T(x) = \frac{a_1x+b_1}{c_1x+d_1}$ and $S(x) = \frac{a_2x+b_2}{c_2x+d_2}$, then:

$$(T \circ S)(x) = \frac{(a_1a_2+b_1c_2)x+(a_1b_2+b_1d_2)}{(c_1a_2+d_1c_2)x+(c_1b_2+d_1d_2)}.$$

So, we do have that $\overline{\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix}} \cdot (\overline{\begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix}} \cdot x) = (\overline{\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix}}\,\overline{\begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix}}) \cdot x$ and $\overline{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} \cdot x = x$.

Let $X_1 = (-\infty, 0]$ and $X_2 = (0, \infty) \cup \{\infty\}$. Then $\tau \cdot x = \frac{1}{-x}$ and so $(G_1 - \{\overline{I}\}) \cdot X_2 \subseteq X_1$. Meanwhile, $\omega \cdot x = \frac{x-1}{x}$ and $\omega^2 \cdot x = \frac{-1}{x-1}$ and so $(G_2 - \{\overline{I}\}) \cdot X_1 \subseteq X_2$. Thus by ping pong lemma, we have that:

$$C_2 * C_3 \cong G_1 * G_2 = \langle G_1, G_2 \rangle = \mathrm{PSL}_2(\mathbb{Z}).$$

Finally, by problem 2 we know that $C_2 * C_3 \cong \langle a \mid a^2 \rangle * \langle b \mid b^3 \rangle \cong \langle a, b \mid a^2, b^3 \rangle$. ■

Interestingly, this and the last problem shows that $\langle a, b \mid a^2 \rangle$ is isomorphic to a subgroup of $\langle a, b \mid a^2, b^3 \rangle$. So that's cool.

Set 8 Problem 5: Prove that the group of Euclidean symmetries of the integers is isomorphic to $\langle a, b \mid a^2, b^2 \rangle$.

To start off, a Euclidean symmetry of the integers is an isometry $\theta : \mathbb{R} \to \mathbb{R}$ satisfying that $\theta(\mathbb{Z}) = \mathbb{Z}$. Note that all isometries are equal to an orthogonal linear function followed by translation by constant. So, we must have that $\theta(x) = ax + b$ where $a \in \{\pm 1\}$ and $b \in \mathbb{R}$. Also, as $\theta(0) \in \mathbb{Z}$ we must have that $b \in \mathbb{Z}$. Hence, the group the problem is asking us about is D_∞ from *problem 5 of the sixth problem set*.

Now, we already know from that prior homework set that D_∞ is generated by the maps $r(x) = -x$ and $s(x) = -x + 1$ where both r and s have order 2. So using the universal property of free groups, let $\Phi : F_{\{a,b\}} \to D_\infty$ be a group homomorphism such that $\Phi(a) = r$ and $\Phi(b) = s$. This map is surjective because r and s generate D_∞. Also, since $\Phi(a^2) = \mathrm{Id} = \Phi(b^2)$ we know there is a well-defined surjective group homomorphism $\overline{\Phi} : \langle a, b \mid a^2, b^2 \rangle \to D_\infty$ with $\overline{\Phi}(a) = r$ and $\overline{\Phi}(b) = s$.

But now because $a^2 = b^2 = 1$, all words in $\langle a, b \mid a^2, b^2 \rangle$ can be reduced to the form $(ab)^k a$, $(ab)^k$, $(ba)^k b$ or $(ba)^k$ where k is a nonnegative integer. Also, note that:

$$(ab)^{-k} = ((ab)^k)^{-1} = (ba)^k \text{ and } (ab)^{-k}a = (ba)^k a = (ba)^{k-1}b$$

Therefore, we can actually write that:

$$\langle a, b \mid a^2, b^2 \rangle = \{(ab)^n : n \in \mathbb{Z}\} \cup \{(ab)^n : n \in \mathbb{Z}\}a$$

And finally, $\overline{\Phi}$ sends each of the elements in the above two cosets to different isometries in D_∞. Specifically, $(\overline{\Phi}((ab)^n))(x) = x - n$ while $(\overline{\Phi}((ab)^n a))(x) = -x - n$. (This was shown in my write up for that prior homework set).

So, $\overline{\Phi}$ is an injection and we are done.

Set 8 Problem 6: Let $G_n := \langle s_1, \ldots, s_{n-1} \mid s_i^2, (s_i s_j)^2 \text{ if } |i - j| > 1; (s_i s_{i+1})^3 \rangle$ (where $n \geq 2$). Prove that $G_n \cong S_n$.

Let $\tau_i := (i \;\; i+1)$ for all $1 \leq i < n$. Then using the universal property of free groups, let $\Phi : F_{\{s_1, \ldots, s_{n-1}\}} \to S_n$ be the unique group homomorphism such that $\Phi(s_i) = \tau_i$ for each i. Note that Φ is surjective since the τ_i generate all of S_n.

If you really doubt that, recall that $\tau_1 = (1 \;\; 2)$ and $\tau_1 \tau_2 \cdots \tau_n = (1 \;\; 2 \;\; \cdots \;\; n)$ generate all of S_n.

Also note that $\tau_i^2 = \mathrm{Id}$. And if $|i - j| > 1$ then $\tau_i \tau_j$ has cycle type $(2 \geq 2 \geq 1 \geq \cdots \geq 1)$. So, $(\tau_i \tau_j)^2 = \mathrm{Id}$. And finally, $\tau_i \tau_{i+1}$ is a three cycle so $(\tau_i \tau_{i+1})^3 = \mathrm{Id}$ for each i. All in all, this shows that all the defining relations of the proposed presentation are in the kernel of Φ. Hence, after quotienting out the normal subgroup generated by them we get a well defined surjective group homomorphism $\overline{\Phi} : G_n \to S_n$ such that $\overline{\Phi}(s_i) = \tau_i$ for each $1 \leq i < n$.

Now to prove that $\overline{\Phi}$ is injective, we proceed by induction on n to show that $|G_n| \leq n!$. That way the only way for $\overline{\Phi}$ to also be surjective is if $|G_n| = n!$ and $\overline{\Phi}$ is one-to-one. For our base case, note that $G_2 = \langle s_1 \mid s_1^2 \rangle \cong C_2$ and $|C_2| = 2 = 2!$

Meanwhile for the inductive step, let H_{n-1} be the subgroup of G_n generated by $s_1, \ldots, s_{n-2}$. Then using the universal property of free groups, let $\Psi : F_{\{s_1, \ldots, s_{n-2}\}} \to H_{n-1}$ be the unique group homomorphism such that $\Psi(s_i) = s_i$ for each i. Again, Ψ is surjective.

It's clear that all the relations defining G_{n-1} are in the kernel of Ψ. Thus, after quotienting them out we get a well-defined surjective group homomorphism $\overline{\Psi} : G_{n-1} \to H_{n-1}$ such that $\overline{\Psi}(s_i) = s_i$ for all i. And by induction, this proves that $|H_{n-1}| \leq (n-1)!$.

Next let $H_{n-1}^{(n-j)}$ be the coset $s_{n-j} \cdots s_{n-1} H_{n-1}$ and also denote $H_{n-1}^{(n)} = H_{n-1}$. Then set $X_n := \{H_{n-1}^{(1)}, \ldots, H_{n-1}^{(n)}\} \subseteq G_n/H_{n-1}$. We can easily see that $s_i H_{n-1}^{(i+1)} = H_{n-1}^{(i)}$. And as $s_i^2 = 1$ we can also see that $s_i H_{n-1}^{(i)} = H_{n-1}^{(i+1)}$.

To show the other cases, note that if $j \leq i-2$, then $s_j s_i = s_i s_j$. Thus since $s_j \in H_{n-1}$ for all $j \leq n-2$, we know that:

$$\begin{aligned} s_j H_{n-1}^{(i)} = s_j s_i s_{i+1} \cdots s_{n-1} H_{n-1} &= s_i s_{i+1} \cdots s_{n-1} s_j H_{n-1} \\ &= s_i s_{i+1} \cdots s_{n-1} H_{n-1} = H_{n-1}^{(i)} \text{ when } j \leq i-2. \end{aligned}$$

As for if $j > i$, then we can write $s_j s_i s_{i+1} \cdots s_{n-1} = s_i \cdots s_{j-2} s_j s_{j-1} s_j \cdots s_{n-1}$ using the identity from the previous paragraph. After that, as $(s_{j-1} s_j)^3 = 1$, we know that $s_j s_{j-1} s_j = s_{j-1} s_j s_{j-1}$. Hence:

$$\begin{aligned} s_i \cdots s_{j-2} s_j s_{j-1} s_j s_{j+1} \cdots s_{n-1} &= s_i \cdots s_{j-2} s_{j-1} s_j s_{j-1} s_{j+1} \cdots s_{n-1} \\ &= s_i \cdots s_{j-2} s_{j-1} s_j s_{j+1} \cdots s_{n-1} s_{j-1} \end{aligned}$$

And as $s_{j-1} \in H_{n-1}$, this shows that $s_j H_{n-1}^{(i)} = H_{n-1}^{(i)}$ when $j > i$.

All in all, this proves that $s_j X_n = X_n$ for all $1 \leq j < n$. And since the s_j generate all of G_n, we in turn know that $\omega X_n = X_n$ for all words $\omega \in G_n$. In particular, this means $\omega H_{n-1} \in X_n$ for all $\omega \in G_n$. So, $[G_n : H_{n-1}] \leq |X_n| \leq n$.

Thus $|G_n| = |H_{n-1}|[G_n : H_{n-1}] \leq (n-1)! \cdot n = n!$. ■

Math 200a notes:

In this class, we define a ring to be a set A equipped with operations $+, \cdot$ such that $(A, +)$ is an abelian group and $(A, \cdot)$ is a semigroup (i.e. a set with an associative operation) such that $0 \cdot a = 0 = a \cdot 0$, $c \cdot (a+b) = ca + cb$, and $(a+b) \cdot c = ac + bc$.

Note, that we shall make a distinction between unital rings and non-unital rings (also called rng's). Specifically, a unital ring has a multiplicative identity element 1 whereas a non-unital ring doesn't. (So in other words we won't take it by definition that a ring has an element 1.)

Usually, we shall assume we are working with commutative unital rings. That said, there are cases where we sometimes want to drop those assumptions.

- Given any ring A, we can define a ring $M_n(A)$ of $n \times n$ matrices of A using standard matrix addition and multiplication. In other words, $[a_{i,j}] + [b_{i,j}] = [(a_{i,j} + b_{i,j})]$ and $[a_{i,j}] \cdot [b_{i,j}] = [(\sum_{k=1}^{n} a_{i,k} b_{k,j})]$. Note that $M_n(A)$ is usually not a commutative even if A is.

    I'm not gonna show these operations satisfy the ring axioms.

- A common counter example is the rng where multiplication sends all pairs of elements to 0.

If G is a group or M is a monoid, then given a ring A we call $A[M]$ or $A[G]$ the monoid ring or group ring where $A[M]$ (resp. $A[G]$) is the collection of formal sums $\sum_{m\in M} a_m m$ (resp. $\sum_{g\in G} a_g g$) where each $a_m \in A$ and $a_m = 0$ for all but finitely many $m \in M$. To turn $A[M]$ (resp. $A[G]$) into a ring, we define:

- $\sum\limits_{m\in M} a_m m + \sum\limits_{m\in M} a'_m m \coloneqq \sum\limits_{m\in M} (a_m + a'_m)m$,
- $(\sum\limits_{m\in M} a_m m)(\sum\limits_{m\in M} a'_m m) = \sum\limits_{m\in M}\left(\sum\limits_{m_1\cdot m_2 = m} a_{m_1}a'_{m_2}\right) m$. (This is called a convolution...)

  I'm not gonna show these operations satisfy the ring axioms.

  (Also note that if A is a commutative ring and M (or G) is abelian, then $A[M]$ (resp. $A[G]$) is a commutative ring.)

If $M = (\mathbb{Z}_{\geq 0})^k \cong \{x_1^{i_1}x_2^{i_2}\cdots x_k^{i_k} : (i_1, \ldots, i_k) \in (\mathbb{Z}_{\geq 0})^k\}$, then $A[(\mathbb{Z}_{\geq 0})^k] \cong A[x_1, \ldots, x_k]$ is the polynomial ring.

Given two rings A_1, A_2, we say $\phi : A_1 \to A_2$ is a ring homomorphism if $\phi(x+y) = \phi(x) + \phi(y)$ and $\phi(xy) = \phi(x)\phi(y)$ for all $x, y \in A_1$. Also, we say ϕ is a unital ring homomorphism if A_1, A_2 are unital rings and $\phi(1_{A_1}) = 1_{A_2}$.

In other words, unlike in math 100b we are not assuming by default that ring homomorphisms are unital.

For example: if B is a commutative ring and $A \subseteq B$ is a subring, then for all $b \in B$ we have that the map $e_b : A[x] \to B$ given by $e_b(f) \coloneqq f(b)$ is a ring homomorphism. And if B and A share a multiplicative identity, then e_b is also unital.

You can see my 100b notes on why this is a homomorphism.

Also, something that's not at all clear is how the professor defines a subring since we've loosened our definition of a ring and ring homomorphism. In this class we say $A \subseteq B$ is a subring if A is closed under multiplication and a subgroup of B with respect to addition.

We say $\mathfrak{a}$ is an ideal of A (also written as $\mathfrak{a} \lhd A$) if $(\mathfrak{a}, +)$ is a subgroup of $(A, +)$ and $ax, xa \in \mathfrak{a}$ for all $x \in \mathfrak{a}$ and $a \in A$.

Note that if A is a unital ring then it suffices to show A is closed under addition and has the mentioned multiplication property. After all, we then have that $-x = (-1)\cdot x \in \mathfrak{a}$ for all $x \in \mathfrak{a}$.

Lemma: If $\phi : A_1 \to A_2$ is a ring homomorphism then $\mathrm{im}(\phi)$ is a subring of A_2 and $\ker(\phi)$ is an ideal of A.

Proof:
Since $\phi : (A_1, +) \to (A_2, +)$ is a group homomorphism, we know that $\mathrm{im}(\phi)$ and $\ker(\phi)$ are subgroups of A_1 and A_2 respectively with respect to $+$. To show that $\ker(\phi)$ is an ideal, note that $\phi(ax) = \phi(a)\phi(x) = 0_{A_2} = \phi(x)\phi(a) = \phi(xa)$ for all $x \in A_1$ and $a \in \ker(\phi)$. Meanwhile, to show that $\mathrm{im}(\phi)$ is a subring, note that if $\phi(x) = a$ and $\phi(y) = b$ then $\phi(xy) = ab$. ■

Switching our perspective, note that if $\mathfrak{a}$ is an ideal of a ring A, then we can define a quotient ring $A/\mathfrak{a}$ by defining $(x+\mathfrak{a})\cdot(y+\mathfrak{a}) = xy+\mathfrak{a}$ on the abelian quotient group $(A/\mathfrak{a}, +)$.

See my math 100b notes for why this is well-defined.

Then the natural projection map $j : A \twoheadrightarrow A/\mathfrak{a}$ satisfies that $\ker(j) = \mathfrak{a}$. Hence, all ideals are kernels of some ring homomorphism.

Returning to the evaluation map $e_b : A[x] \to B$ where B is a commutative ring and $A \subseteq B$ is a subring, one can fairly easy see that $\mathrm{im}(e_b)$ is the smallest subring of B containing A and b. We denote $\mathrm{im}(e_b)$ as $A[b]$ (not to be confused with a monoid ring or polynomial ring).

11/24/2025

Math 220a Notes:

If G is an open set, then we say γ is homologous to zero (denoted $\gamma \approx_G 0$) iff $n(\gamma; w) = 0$ for all $w \in \mathbb{C} - G$.

Note that by the first corollary on *page 415*, we have that $\gamma \sim_G 0 \implies \gamma \approx_G 0$.

Suppose $G \subseteq \mathbb{C}$ is a region and $f : G \to \mathbb{C}$ is analytic on G with the zeros $a_1, \ldots, a_n$ (where the a_k are allowed to be repeated). As noted on *page 382* of my journal as well as my spring notes, we can then find an analytic function $g : G \to \mathbb{C}$ with no zeros such that $f(z) = (z-a_1)\cdots(z-a_n)g(z)$. Then by product rule, we get that:

$$f'(z) = \sum_{k=1}^{n} \Big(\prod_{i\neq k}(z-a_i)\Big)g(z) + g'(z)\prod_{k=1}^{n}(z-a_k)$$

And dividing both sides by $f(z)$ we get that:

$$\frac{f'(z)}{f(z)} = \frac{g'(z)}{g(z)} + \sum_{k=1}^{n}(z-a_k)^{-1} \text{ when } z \neq a_1, \ldots, a_n$$

(Conway) Theorem IV.7.2: Let G be a region and let f be an analytic function on G with zeros $a_1, \ldots, a_n$ (repeated according to multiplicity) like above. If γ is a closed piecewise C^1 curve in G which does not pass through any point a_k and if $\gamma \approx_G 0$ then:

$$\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)}\mathrm{d}z = \sum_{k=1}^{m} n(\gamma; a_k).$$

Proof:
Letting g be as above, we know that:

$$\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)}\mathrm{d}z = \frac{1}{2\pi i}\int_\gamma \frac{g'(z)}{g(z)}\mathrm{d}z + \sum_{k=1}^{n}\frac{1}{2\pi i}\int_\gamma \frac{1}{z-a_k}\mathrm{d}z = \frac{1}{2\pi i}\int_\gamma \frac{g'(z)}{g(z)}\mathrm{d}z + \sum_{k=1}^{m} n(\gamma; a_k)$$

Then since $g(z) \neq 0$ for any $z \in G$, we know that $\frac{g'}{g}$ is analytic on G. Hence, we have that $\int_\gamma \frac{g'(z)}{g(z)}\mathrm{d}z = 0$. ■

<u>(Conway) Corollary IV.7.3:</u> Let f, G, γ be as in the last theorem but let $a_1, \ldots, a_n$ (repeated according to multiplicity) be all the points where f equals α. In other words, $a_1, \ldots, a_n$ are the zeros of $f(z) - \alpha$. Then:

$$\frac{1}{2\pi i} \int_\gamma \frac{f'(z)}{f(z)-\alpha} \mathrm{d}z = \sum_{k=1}^{m} n(\gamma; a_k).$$

Note that if $f : G \to \mathbb{C}$ is an analytic non-constant function on G, it is possible for f to have infinitely many zeros in G. That said, because the set of zeros can't have a limit point in G, we know that if $K \subseteq G$ is compact then f can only have finitely many limit points in K. Consequently, if $\gamma \approx_G 0$ we can show that $f(z) = \alpha$ must have only finitely many solutions in G such that $n(\gamma; z) \neq 0$.

Exercise IV.7.2: Let $G \subseteq \mathbb{C}$ be open and suppose γ is a closed piecewise C^1 curve in G such that $\gamma \approx_G 0$. Set $H := \{z \in \mathbb{C} : n(\gamma; z) = 0\}$.

(a) Suppose G is a proper subset of $\mathbb{C}$ and define $r := \inf(\{|z - w| : z \in \{\gamma\}, w \in \partial G\})$. Note that r exists and is positive because $\partial G, \{\gamma\}$ are closed disjoint nonempty sets with $\{\gamma\}$ compact. Now show that $\{z \in \overline{G} : \inf\{|z - w| : w \in \partial G\} < \frac{r}{2}\} \subseteq H$

It suffices to show that if $\inf\{|z-w| : w \in \partial G\} < \frac{r}{2}$ then z is in the same component of $G - \{\gamma\}$ as some $w \in \partial G$. After all, as $w \in G^{\mathsf{C}}$ and $f \approx_G 0$ we know that $n(\gamma; w) = 0$. Also, as $n(\gamma; z)$ is constant on each component of $\mathbb{C} - G$, we would thus have that $n(\gamma; z) = n(\gamma; w)$. Fortunately, we can just pick $w \in \partial G$ such that $|z - w| < \frac{1}{2}r$. Next, we note that the line segment $[z, w]$ can't intercept $\{\gamma\}$ as that would contradict how we defined r. So, z, w must be in the same component of $\mathbb{C} - \{\gamma\}$.

(b) Use part (a) to show that if $f : G \to \mathbb{C}$ is analytic and non-constant then $f(z) = \alpha$ has at most a finite number of solutions z such that $n(\gamma; z) \neq 0$.

Since γ is bounded, we can find an open ball $B \subseteq \mathbb{C}$ of finite radius with $\{\gamma\} \subseteq B$. Then as B is convex, we know that $\gamma \sim_B 0$. Hence $G - B \subseteq H$.

Meanwhile, let $r := \inf(\{|z - w|\} : z \in \{\gamma\}, w \in \partial(B \cap G))$. Then by part (a) we know that $V := \{z \in \overline{B \cap G} : \inf\{|z-w| \in \overline{B \cap G} : w \in \partial(B \cap G)\} < \frac{r}{2}\} \subseteq H$. Hence $K := \overline{B \cap G} - V$ must contain H^{C}. But also note that V is an open subset of $\overline{B \cap G}$. Hence, K is a closed subset relative to the compact set $\overline{B \cap G}$. In turn, K is compact. Also as $\partial(B \cap G) \subseteq V$ we know that $K \subseteq B \cap G$.

With that, we've proven there is a compact set $K \subseteq G$ with $n(\gamma; z) = 0$ outside of K. And as noted before, $f(z) = \alpha$ can only have finitely many solution on K as any infinite subset of a compact set has a limit point. ■

A simple root of $f(z) = \xi$ is a zero of $f(z) - \xi$ with multiplicity 1.

(Conway) Theorem IV.7.4: Suppose $f : G \to \mathbb{C}$ is analytic and $z_0 \in G$ is such that $f(z) - w_0$ has a zero of multiplicity m at z_0. Then there exists $\varepsilon, \delta > 0$ such that for all $w \in B_\varepsilon(w_0)$ the equation $f(z) - w$ has exactly m zeros in $B_\delta(z_0)$ which furthermore are all simple if $w \neq w_0$.

Proof:

To start off, we may pick $\delta > 0$ such that $f(z) - w_0 \neq 0$ for all $z \in \overline{B_\delta(z_0)} - \{z_0\} \subseteq G$. Then let $\gamma(s) = z_0 + \delta e^{is}$ and note that $\sigma := f \circ \gamma$ is a closed piecewise C^1 curve not passing through w_0. Hence, $\varepsilon := \inf\{|w - w_0| : w \in \{\sigma\}\} > 0$ and in turn $g(w) := \frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)-w}\mathrm{d}z$ is continuous (by Leibniz's rule) as w ranges over $B_\varepsilon(w_0)$. Yet also recall from our prior theorems that $g(w)$ is integer valued. Hence, we know g is constant on $B_\varepsilon(w_0)$.

But now note that $n(\gamma; z) = 1$ for all $z \in B_\delta(z_0)$ and $n(\gamma; z) = 0$ for all other $z \in \mathbb{C} - \{\gamma\}$. Therefore, we can calculate that $g(w_0) = \sum_{k=1}^{m} n(\gamma; z_0) = m \cdot 1$. And this proves that $g(w) = \frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)-w}\mathrm{d}z = m$ for all $w \in B_\varepsilon(w_0)$.

Next note that if $a_1, \ldots, a_n$ are the zeros in $B_\delta(z_0)$ (repeated according to multiplicity) of $f(z) - w$, then since $n(\gamma; a_k) = 1$ for each k, we have for all $w \in B_\varepsilon(w_0)$ that:

$$m = g(w) = \sum_{k=1}^{n} n(\gamma; a_k) = n$$

So, there are exactly m solutions in $B_\delta(z_0)$ to the equation $f(z) = w$ for all $w \in B_\varepsilon(w_0)$.

Finally, if $m = 1$ then there is nothing to prove. Meanwhile, if $m > 1$ then we can easily show that $f'(z_0) = 0$ (see the exercise below). In turn, as f' is analytic we can say that if we had initially started with a small enough δ then we'd have that $f'(z) \neq 0$ for all $z \in \overline{B}_\delta(z_0) - \{z_0\}$. In turn, each root of $f(z) - w$ must be simple when $w \neq w_0$. ■

Exercise IV.7.3: Let f be analytic in $B_R(a)$ and suppose that $f(a) = 0$. Show that a is a zero of multiplicity m iff $f^{(m-1)}(a) = \cdots = f^{(1)}(a) = f(a) = 0$ and $f^{(m)}(a) \neq 0$.

$(\Longleftarrow)$

Write f as a power series $\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(z-a)^n$. If $f^{(m-1)}(a) = \cdots = f^{(1)}(a) = f(a) = 0$ and $f^{(m)}(a) \neq 0$ then we can factor out $(z-a)^m$ and get a power series which is nonzero at a.

$(\Longrightarrow)$

Write $f(z) = (z-a)^m g(z)$ where $g(a) \neq 0$ and both f and g are analytic. Then we can express g as a power series $\sum_{n=0}^{\infty} \frac{g^{(n)}(a)}{n!}(z-a)^n$. In turn:

$$f(z) = \sum_{n=0}^{m-1} \frac{0}{n!}(z-a)^n + \sum_{n=0}^{\infty} \frac{g^{(n)}(a)}{(n)!}(z-a)^{n+m}.$$

And by looking at each coefficient in the power series, we see that $f^{(n)}(a) = 0$ for all $n < m$ and $f^{(m)}(a) = m!g(a) \neq 0$. ■

Open Mapping Theorem: Let G be a region and suppose that $f : G \to \mathbb{C}$ is a non-constant analytic function on G. Then for any open set $U \subseteq G$ we have that $f(U)$ is open in $\mathbb{C}$.

Proof:

By the last theorem, for all $z \in G$ we can find $\varepsilon, \delta > 0$ such that:

$$B_\varepsilon(f(z)) \subseteq f(B_\delta(z)) \subseteq f(G).$$

One more comment before I start with the 220 homework. Conway finally proves that being complex differentiable a single time on an open set makes a function holomoprhic on that open set. The proof he uses doesn't have any new ideas from my notes from last Spring though.

Math 220a Homework:

Exercise IV.6.1: Let G be a region and let $\sigma_1, \sigma_2 : [0,1] \to G$ be the constant curves at a and b in G. Show that if γ is a closed piecewise C^1 curve and $\gamma \sim_G \omega_1$ then $\gamma \sim_G \omega_2$.

Proof:

Since G is a connected open subset of $\mathbb{C}$, we know G is path connected. Then letting $\omega : [0,1] \to G$ be any path going from a to b, we have that $\Gamma(s,t) = \omega(t)$ is a homotopy from σ_1 to σ_2. Hence, $\sigma_1 \sim_G \sigma_2$.

Then as $\sim_G$ is an equivalence relation and $\gamma \sim_G \sigma_1 \sim_G \sigma_2$, we are done. ■

Exercise IV.6.4: Let $G = \mathbb{C} - \{0\}$ and show that every closed curve in G is homotopic to a closed curve whose trace is contained in $\{z : |z| = 1\}$.

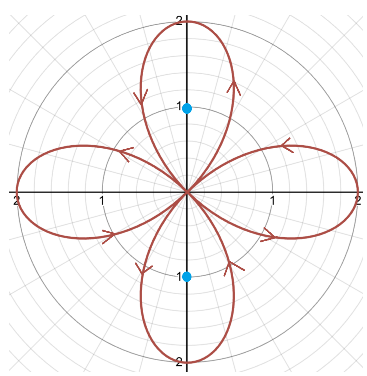
Define $\Gamma(s,t) := (1-t)\gamma(s) + t\frac{\gamma(s)}{|\gamma(s)|}$. Since $\gamma(s) \neq 0$ ever, we know that Γ is continuous. Also, $\Gamma(s,0) = \gamma(s)$ and $|\Gamma(s,1)| = |0 + 1\frac{\gamma(s)}{|\gamma(s)|}| = 1$. So, the curve $\gamma_1(s) = \Gamma(s,1)$ is a continuous curve whose trace is contained in $\{z : |z| = 1\}$. Finally, as $\gamma(0) = \gamma(1)$ we know that $\Gamma(0,t) = \Gamma(1,t)$ for all t. Hence, Γ is a homotopy.

Exercise IV.6.5: Evaluate the integral $\int_\gamma \frac{\mathrm{d}z}{z^2+1}$ where $\gamma(\theta) = 2|\cos(2\theta)|e^{i\theta}$ for $0 \leq \theta \leq 2\pi$.

Note that $\frac{1}{z^2+1} = \frac{1}{(z+i)(z-i)}$. Then if $a(z+i) + b(z-i) = 1$, we must have that $a + b = 0$ and $i(a-b) = 1$. In particular $a = -b$ and so $i(2a) = 1$. In turn $a = \frac{1}{2i}$ and $b = \frac{-1}{2i}$. And this lets us conclude that:

$$\int_\gamma \frac{\mathrm{d}z}{z^2+1} = \frac{1}{2i}\int_\gamma \frac{\mathrm{d}z}{z-i} - \frac{1}{2i}\int_\gamma \frac{\mathrm{d}z}{z+i} = \pi n(\gamma; i) - \pi n(\gamma; -i)$$

But now note that $\gamma(\theta) = 2|\cos(2\theta)|e^{i\theta}$ traces out the same curve as the polar graph $r(\theta) = 2|cos(2\theta)|$. Specifically, that curve has 4 equally spaced petals flower drawn counter-clockwise about the origin as shown to the right.



From here it is clear that $n(\gamma; i) = 1$ and $n(\gamma; -i) = 1$. So:

$$\pi n(\gamma; i) - \pi n(\gamma; -i) = \pi - \pi = 0.$$

Exercise IV.7.4: Suppose that $f : G \to \mathbb{C}$ is analytic and injective. Then $f'(z) \neq 0$ for any $z \in G$.

Proof:

Suppose to the contrary that $f'(z_0) = 0$ for some $z_0 \in G$ and let $w_0 = f(z_0)$. Then we'd know that $f(z) - w_0$ has a zero of multiplicity $m > 1$ at $z = z_0$. So in turn, there exists $\varepsilon, \delta > 0$ such that $f(z) - w$ has two simple roots for in $B_\delta(z_0)$ for all $w \in B_\varepsilon(w_0)$. But that also means that f isn't injective on $B_\delta(z_0) \subseteq G$.

This proves that $f'(z) \neq 0$ anywhere on G is a necessary condition for an analytic function $f : G \to \mathbb{C}$ to be injective.

Exercise IV.7.5: Let X and Ω be metric spaces and suppose that $f : X \to \Omega$ is a bijection. Then f is an open map iff f is a closed map.

To start off, for any set $E \subseteq X$ we have that $f(E^{\mathsf{C}}) = f(X) - f(E)$ since f is injective. Then as f is surjective we have that $f(X) - f(E) = (f(E))^{\mathsf{C}}$. Hence, we've shown that set complements commute in and out of the function.

$(\Longrightarrow)$

Suppose $f(U)$ is open for all open U. Then given any closed set C, we know that $f(C^{\mathsf{C}})$ is open. But we also know that $f(C^{\mathsf{C}}) = (f(C))^{\mathsf{C}}$. So, $f(C)$ is closed.

$(\Longleftarrow)$

Literally do the same reasoning but swap the words open and closed.

I want to finish taking notes on Haar measures now. See *page 364* for where I'm starting from. As a reminder, I'm following Folland's real analysis book. Also, if G is a topological group then e is the identity element of G.

Theorem 11.9: If μ and ν are left Haar measures on a locally compact group G then there exists $c > 0$ such that $\mu = c\nu$.

(Proof for when μ is both left- and right-invariant [which will for example happen if G is abelian]):

Pick $h \in C_c^+(G)$ such that $h(x) = h(x^{-1})$. (One way of doing this would be to just define $h(x) = g(x) + g(x^{-1})$ where $g \in C_c^+(G)$). Then for any $f \in C_c(G)$, we have that:

$$\begin{aligned}
\int h \mathrm{d}\nu \cdot \int f \mathrm{d}\mu &= \iint h(y) f(x) \mathrm{d}\mu(x) \mathrm{d}\nu(y) \\
&= \iint h(y) f(xy) \mathrm{d}\mu(x) \mathrm{d}\nu(y) \text{ (by right-invariance of } \mu) \\
&= \iint h(y) f(xy) \mathrm{d}\nu(y) \mathrm{d}\mu(x) \text{ (by Fubini's theorem)} \\
&= \iint h(x^{-1}y) f(x^{-1}xy) \mathrm{d}\nu(y) \mathrm{d}\mu(x) \text{ (by left invariance of } \nu) \\
&= \iint h(y^{-1}x) f(y) \mathrm{d}\nu(y) \mathrm{d}\mu(x) \text{ (by how we chose } h) \\
&= \iint h(y^{-1}x) f(y) \mathrm{d}\mu(x) \mathrm{d}\nu(y) \text{ (by Fubini's theorem)} \\
&= \iint h(yy^{-1}x) f(y) \mathrm{d}\mu(x) \mathrm{d}\nu(y) \text{ (by left-invariance of } \mu) \\
&= \iint h(x) f(y) \mathrm{d}\mu(x) \mathrm{d}\nu(y) = \int h \mathrm{d}\mu \cdot \int f \mathrm{d}\nu
\end{aligned}$$

Hence $\int f \mathrm{d}\mu = c \int f \mathrm{d}\nu$ for all $f \in C_c^+(G)$ where $c = (\int h \mathrm{d}\mu)/(\int h \mathrm{d}\nu)$. (Recall that $\int h \mathrm{d}\nu > 0$ by *proposition 11.4(c) on page 353*). In turn, this implies that $\mu = c\nu$ (since μ and ν are Radon measures).

(General Proof:)

Note that $\mu = c\nu$ iff the ratio $r_f := (\int f\mathrm{d}\mu)/(\int f\mathrm{d}\nu)$ is independent of $f \in C_c^+(G)$.

The $(\Longleftarrow)$ implication is obvious. Meanwhile, to see the other direction, note that for any nonempty open set we can find a sequence of functions such that $(\int f_n\mathrm{d}\mu)/(\int f_n\mathrm{d}\nu) \to \mu(U)/\nu(U)$ as $n \to \infty$ (again, U has nonzero ν measure by *proposition 11.4(c)*). So the right side statement would imply $\mu(U) = r_f\nu(U)$ for all open sets U. Then by the outer regularity of μ and ν we'd have that $\mu = r_f\nu$.

So, suppose $f, g \in C_c^+(G)$. Then fix a compact symmetric neighborhood V_0 of e and set $A := (\mathrm{supp}(f))V_0 \cup V_0(\mathrm{supp}(f))$ and $B := (\mathrm{supp}(g))V_0 \cup V_0(\mathrm{supp}(g))$.

Note by the continuity of $x \mapsto x^{-1}$ that if N is a compact neighborhood of e then so is N^{-1}. So, we have no issue defining $V_0 = N \cap N^{-1}$ like in *proposition 11.1(b)*. Similarly, A and B are compact by proposition 11.1(f).

Now for any $y \in V_0$ the functions $x \mapsto f(xy) - f(yx)$ and $x \mapsto g(xy) - g(yx)$ are supported in A and B. Also by *proposition 11.2*, given any $\varepsilon > 0$ we can get a symmetric compact neighborhood $V \subseteq V_0$ of e such that:

$\sup_{x\in G} |f(xy) - f(yx)| < \varepsilon$ and $\sup_{x\in G} |g(xy) - g(yx)| < \varepsilon$ for all $y \in V$.

To get V, first just take the intersection of V_0 with four different neighborhoods gotten by proposition 11.2. Then use LCH space properties to get compact neighborhood of e contained in that intersection. And finally, use proposition 11.1(b) to get a compact symmetric neighborhood.

Pick $h \in C_C^+(G)$ with $\mathrm{supp}(h) \subseteq V$ and $h(x) = h(x^{-1})$. Similarly to the last page, you can do this by defining $h(x) = g(x) + g(x^{-1})$ where $g \in C_c^+(G)$ satisfies that $\mathrm{supp}(g) \subseteq V$. Then:

$$\int h\mathrm{d}\nu \int f\mathrm{d}\mu = \iint h(y)f(x)\mathrm{d}\mu(x)\mathrm{d}\nu(y)$$
$$= \iint h(y)f(yx)\mathrm{d}\mu(x)\mathrm{d}\nu(y) \text{ (by the left-invariance of } \mu)$$

But also note that:

$$\int h\mathrm{d}\mu \int f\mathrm{d}\nu = \iint h(x)f(y)\mathrm{d}\mu(x)\mathrm{d}\nu(y)$$
$$= \iint h(y^{-1}x)f(y)\mathrm{d}\mu(x)\mathrm{d}\nu(y) \text{ (by the left-invariance of } \mu)$$
$$= \iint h(y^{-1}x)f(y)\mathrm{d}\nu(y)\mathrm{d}\mu(x) \text{ (by Fubini's theorem)}$$
$$= \iint h(x^{-1}y)f(y)\mathrm{d}\nu(y)\mathrm{d}\mu(x) \text{ (by how we chose } h)$$
$$= \iint h(xx^{-1}y)f(xy)\mathrm{d}\nu(y)\mathrm{d}\mu(x) \text{ (by left-invariance of } \nu)$$
$$= \iint h(y)f(xy)\mathrm{d}\mu(x)\mathrm{d}\nu(y) \text{ (by Fubini's theorem)}$$

Therefore, we have that:

$$|\int h\mathrm{d}\mu \int f\mathrm{d}\nu - \int h\mathrm{d}\nu \int f\mathrm{d}\mu| = |\iint h(y)\cdot(f(xy) - f(yx))\mathrm{d}\mu(x)\mathrm{d}\nu(y)|$$
$$\leq \varepsilon\mu(A)\int h\mathrm{d}\nu$$

By identical reasoning we can also conclude that:

$$|\int h\mathrm{d}\mu \int g\mathrm{d}\nu - \int h\mathrm{d}\nu \int g\mathrm{d}\mu| \leq \varepsilon\mu(B)\int h\mathrm{d}\nu.$$

So, divide these inequalities by $(\int h\mathrm{d}\nu)(\int f\mathrm{d}\nu)$ and $(\int h\mathrm{d}\nu)(\int g\mathrm{d}\nu)$ respectively to get that:

$$\left|\frac{\int h\mathrm{d}\mu}{\int h\mathrm{d}\nu} - \frac{\int f\mathrm{d}\mu}{\int f\mathrm{d}\nu}\right| \leq \frac{\varepsilon\mu(A)}{\int f\mathrm{d}\nu} \text{ and } \left|\frac{\int h\mathrm{d}\mu}{\int h\mathrm{d}\nu} - \frac{\int g\mathrm{d}\mu}{\int g\mathrm{d}\nu}\right| \leq \frac{\varepsilon\mu(B)}{\int g\mathrm{d}\nu}$$

In turn, by triangle inequality we know that $\left|\frac{\int f\mathrm{d}\mu}{\int f\mathrm{d}\nu} - \frac{\int g\mathrm{d}\mu}{\int g\mathrm{d}\nu}\right| \leq \varepsilon(\frac{\mu(A)}{\int f\mathrm{d}\nu} + \frac{\mu(B)}{\int g\mathrm{d}\nu})$. And to finish the proof we take $\varepsilon \to 0$ (which we can do because A and B were chosen before we considered ε). ■

If μ is a left Haar measure on G and $x \in G$, then the measure $\mu_x(E) = \mu(Ex)$ is another left Haar measure. Hence by the prior theorem there exists a number $\Delta(x)$ such that $\mu_x = \Delta(x)\mu$. Also by the prior theorem, $\Delta(x)$ is independent of our choice of left Haar measure μ.

We call $\Delta : G \to (0, \infty)$ the modular function of G.

Proposition 11.10: Δ is a continuous homomorphism from G to the multiplicative group of positive real numbers. Moreover, if μ is a left Haar measure on G, for any $f \in L^1(\mu)$ and $y \in G$ we have that $\int(R_y f)\mathrm{d}\mu = \Delta(y^{-1})\int f\mathrm{d}\mu$.

Proof:

For any $x, y \in G$ and $E \in \mathcal{B}_G$ we have that:

$$\Delta(xy)\mu(E) = \mu(Exy) = \Delta(y)\mu(Ex) = \Delta(y)\Delta(x)\mu(E) = \Delta(x)\Delta(y)\mu(E).$$

Hence, Δ is a group homomorphism from G to $(0, \infty)$.

Next note that $\mu_{y^{-1}}$ is just the image (or pushfoward) measure of the function $x \mapsto xy$. Hence by *proposition 10.1 on page 193*:

$$\int(R_y f)\mathrm{d}\mu = \int f\mathrm{d}\mu_{y^{-1}} = \Delta(y^{-1})\int f\mathrm{d}\mu$$

Finally, the below exercise plus the above formula shows that the map $y \mapsto \Delta(y^{-1})\int f\mathrm{d}\mu$ is continuous for any $f \in L^1(\mu)$. After fixing f so that $\int f\mathrm{d}\mu = 1$ and composing this map from the inside with the continuous inversion map, we get that Δ is continuous.

Exercise 11.2: If μ is a Radon measure on the locally compact group G and $f \in C_c(G)$ then the functions $x \mapsto \int(L_x f)\mathrm{d}\mu$ and $x \mapsto \int(R_x f)\mathrm{d}\mu$ are continuous.

The proof is analogous for the left translation and right translation cases. So I'll just focus on the map $x \mapsto \int(R_x f)\mathrm{d}\mu$.

Given $f \in C_c(G)$, consider any fixed $x_0 \in G$ and $\varepsilon > 0$. By *proposition 11.2* we can find a neighborhood V of e such that for all $y \in V$:

$$\|R_y(R_{x_0}f) - (R_{x_0}f)\|_u = \|R_{yx_0}f - (R_{x_0}f)\|_u < \frac{\varepsilon}{\mu(\mathrm{supp}(R_{x_0}f))}.$$

In particular, this means for any x in the neighborhood Vx_0 of x_0 that

$$|\int(R_x f)\mathrm{d}\mu - \int(R_{x_0}f)\mathrm{d}\mu| \leq \frac{\varepsilon}{\mu(\mathrm{supp}(R_{x_0}f))} \cdot \mu(\mathrm{supp}(R_{x_0}f)) = \varepsilon.$$

And this proves that $x \mapsto \int(R_x f)\mathrm{d}\mu$ is continuous at $x = x_0$. ■

Any left Haar measure of a locally compact group G is also a right Haar measure iff $\mathrm{im}(\Delta) = 1$, in which case G is called unimodular. Now it's obvious that all abelian locally compact groups are unimodular. But interestingly enough, we can also show that if a group becomes not abelian enough, then it's also guarenteed to be unimodular.

Proposition 11.12: Let G be a locally compact group. If $G/[G, G]$ is finite then G is unimodular.

> Since Δ is a homomorphism from G to an abelian group, we must have that the commutator subgroup $[G, G]$ is contained in the kernel of Δ. Hence, by quotienting out $[G, G]$ we get a well-defined homomorphism $\widetilde{\Delta} : G/[G, G] \to (0, \infty)$. But now as $G/[G, G]$ is finite, we must have that $\mathrm{im}(\widetilde{\Delta}) = \mathrm{im}(\Delta)$ is a finite subgroup of $(0, \infty)$. Yet, the only finite subgroup of the multiplicative group of positive real numbers is $\{1\}$. So, $\Delta(g) = 1$ for all $g \in G$. ■

Another useful case is as follows:

Proposition 11.13: If G is a compact group then G is unimodular.

> Proof:
> Let μ be a left Haar measure. Then for any $x \in G$ we have that $\mu(G) = \mu(Gx^{-1}) = \Delta(x)\mu(G)$. And since $0 < \mu(G) < \infty$, this means that $\Delta(x) = 1$ for all $x \in G$. ■

This is where I'm going to stop covering Folland again and instead switch over to the math 241 class (which I'm still in by the way).

11/26/2025

Math 241a Notes:

In this class we'll assume topological groups are always Hausdorff. Recall *page 351* for why this isn't must of a restriction.

(Example 1.3.4:) Here are some relevant examples of topological groups.

- Note that $\mathrm{GL}_n(\mathbb{R})$ is a group with an obvious embedding into $\mathbb{R}^{n^2}$. Furthermore, matrix multiplication and inversion can be written such that each component of the resulting matrix is a rational function of the components of the input matrices. Hence, giving $\mathrm{GL}_n(\mathbb{R})$ the Euclidean topology induced by $\mathbb{R}^{n^2}$ turns $\mathrm{GL}_n(\mathbb{R})$ into a topological group.
- If G is a topological group and $H < G$, then H equipped with the subspace topology will be a topological group. In particular, this means any subgroup of $\mathrm{GL}_n(\mathbb{R})$ is a topological group.

  > Side note, on *page 92* I showed that the set of all orthogonal nxn matrices $O_n(\mathbb{R})$ is a smooth compact manifold in $\mathbb{R}^{n^2}$. And since the group operations on $O_n(\mathbb{R})$ are smooth, we say $O_n(\mathbb{R})$ is a lie group.

- If $\mathcal{X}$ is a normed vector space, then $\mathrm{Iso}(\mathcal{X})$ is a topological group when equipped with the strong operator topology.

Proof:
Let $\langle (T_i, S_i)\rangle_{i\in I}$ be a net in $\mathrm{Iso}(\mathcal{X}) \times \mathrm{Iso}(\mathcal{X})$ converging to (T, S) operator strongly. Then we claim that $T_iS_i \to TS$ operator strongly. After all, fix any $x \in \mathcal{X}$ and $\varepsilon > 0$. Then as T_i is an isometry for each i, we have that:

$$\begin{aligned}\|T_i(S_i(x)) - T(S(x))\| &\leq \|T_i(S_i(x)) - T_i(S(x))\| + \|T_i(S(x)) - T(S(x))\\ &= \|S_i(x) - S(x)\| + \|T_i(S(x)) - T(S(x))\|\end{aligned}$$

Then because $S_i \to S$ and $T_i \to T$ operator strongly, we know that $\|S_i(x) - S(x)\| \to 0$ and $\|T_i(S(x)) - T(S(x))\| \to 0$.

Next, let $\langle T_i\rangle_{i\in I}$ be a net in $\mathrm{Iso}(\mathcal{X})$ converging to T operator strongly. Then since T_i is an isometry, for any fixed $x \in \mathcal{X}$ we have that:

$$\|T_i^{-1}(x) - T^{-1}(x)\| = \|x - T_i(T^{-1}(x))\| = \|T(T^{-1}(x)) - T_i(T^{-1}(x))\|$$

And since $T_i \to T$ operator strongly, $\|T(T^{-1}(x)) - T_i(T^{-1}(x))\| \to 0$. Hence $T_i^{-1} \to T^{-1}$ operator strongly. ■

(Zimmer) Exercise 1.21: Let $\mathcal{H}$ be a Hilbert space and let $U(\mathcal{H})$ be the group of unitary linear operators on $\mathcal{H}$. Then the strong and weak operator topologies are the same on $U(\mathcal{H})$.

Proof:
We already know the strong operator topology is finer than the weak operator topology. Meanwhile, to show the other direction it suffices to show by the corollary on *page 229* that weak operator convergence in $U(\mathcal{H})$ implies strong operator convergence in $U(\mathcal{H})$.

Let $\{e_i\}_{i\in I}$ be an orthonormal basis for $\mathcal{H}$. Then consider any net $\langle T_\alpha\rangle_{\alpha\in A}$ in $U(\mathcal{H})$ converging to T. Since $\|T_\alpha\|, \|T\| = 1$ for all $\alpha \in A$, we know by example 1.3.2 on *pages 303-304* that if $\|T_\alpha e_i - Te_i\| \to 0$ for all i then $T_\alpha \to T$ operator strongly.

As a side note, none of the reasoning I wrote on pages 303 and 304 breaks down if you use nets instead of sequences. I just used sequences because the professor prefers them.

Fortunately, note that:

$$\begin{aligned}\|T_\alpha e_i - Te_i\|^2 &= \langle T_\alpha e_i - Te_i, T_\alpha e_i - Te_i\rangle\\ &= \langle T_\alpha e_i, T_\alpha e_i - Te_i\rangle - \langle Te_i, T_\alpha e_i - Te_i\rangle\\ &= \langle T_\alpha e_i, T_\alpha e_i\rangle - \langle T_\alpha e_i, Te_i\rangle - \langle Te_i, T_\alpha e_i\rangle + \langle Te_i, Te_i\rangle\\ &= \langle e_i, e_i\rangle - \langle T_\alpha e_i, Te_i\rangle - \langle Te_i, T_\alpha e_i\rangle + \langle e_i, e_i\rangle\\ &= 2 - (\langle T_\alpha e_i, Te_i\rangle - \langle Te_i, T_\alpha e_i\rangle)\\ &= 2 - (\langle T_\alpha e_i, Te_i\rangle - \overline{\langle T_\alpha e_i, Te_i\rangle})\end{aligned}$$

Since $T_\alpha \to T$ operator weakly, we know that $\langle T_\alpha x, y\rangle \to \langle Tx, y\rangle$ for all $x, y \in \mathcal{H}$. In particular, setting $x = e_i$ and $y = Te_i$ we have that $\langle T_\alpha e_i, Te_i\rangle \to \langle Te_i, Te_i\rangle$. And as T is unitary, the latter is equal to $\langle e_i, e_i\rangle = 1$. This shows that:

$$2 - (\langle T_\alpha e_i, Te_i\rangle - \overline{\langle T_\alpha e_i, Te_i\rangle}) \to 2 - (1 + 1) = 0. \blacksquare$$

Given a topological group G and a topological space X, we say an action $G \curvearrowright X$ is continuous if it's corresponding induced map $G \times X \to X$ is continuous. Note in that case that the map $\varphi_g(x) := g \cdot x$ is a homeomorphism on X with inverse $\varphi_{g^{-1}}$.

If G is a group and V is a vector space, a representation of G on V is a homomorphism $G \to \mathrm{GL}(V)$ where $\mathrm{GL}(V)$ is the group of invertible linear maps on V. If $\mathcal{X}$ is a topological vector space (which is always assumed to be over $\mathbb{R}$ or $\mathbb{C}$ in this class), a representation of G on $\mathcal{X}$ is a homomorphism $\pi : G \to \mathrm{Aut}(\mathcal{X})$. When G is also a topological group, we can talk about π as being continuous with respect to an operator topology on $\mathrm{Aut}(\mathcal{X})$.

If $\mathcal{X}$ is a normed vector space, a representation $\pi : G \to \mathcal{X}$ is called an isometric representation if $\pi(G) \subseteq \mathrm{Iso}(\mathcal{X})$. We similarly define unitary representations into Hilbert spaces.

(Zimmer) Exercise 1.12: If G is a topological group and $\mathcal{X}$ is a normed space, show that a representation $\pi : G \to \mathrm{Aut}(\mathcal{X})$ is continuous iff it is continuous at the identity e of G.

The ($\Longleftarrow$) direction with respect to each of the three settings below is trivial. Meanwhile, given any net $\langle g_i \rangle_{i \in I}$ in G converging to some element g, note that:

- $\|\pi(g_i) - \pi(g)\|_{\mathrm{op}} = \|\pi(g_i g^{-1})\pi(g) - \pi(g)\|_{\mathrm{op}} \leq \|\pi(g_i g^{-1}) - \mathrm{Id}\|_{\mathrm{op}} \cdot \|\pi(g)\|_{\mathrm{op}}$,
- $\|(\pi(g_i))(x) - (\pi(g))(x)\| = \|(\pi(g_i g^{-1}))((\pi(g))(x)) - (\pi(g))(x)\|$
  $= \|(\pi(g_i g^{-1}) - \mathrm{Id})((\pi(g))(x))\|$ for all $x \in \mathcal{X}$
- $|f((\pi(g_i))(x) - (\pi(g))(x))| = |f((\pi(g_i g^{-1}))((\pi(g))(x)) - (\pi(g))(x))|$
  $= |f\left((\pi(g_i g^{-1}) - \mathrm{Id})((\pi(g))(x))\right)|$ for all $x \in \mathcal{X}$ and $f \in \mathcal{X}^*$

Thus, $\pi(g_i) \to \pi(g)$ in norm, operator strongly, or operator weakly if $\pi(g_i g^{-1}) \to \pi(e)$ in norm, operator strongly, or operator weakly respectively. Fortunately, the latter happens if π is continuous at e. ■

Proposition 1.3.9: Let G be a topological group acting continuously on an LCH space X. Then let $\pi : G \to \mathrm{Iso}(C_c(X))$ be given by $(\pi(g))(f) := f(g^{-1} \cdot x)$. Now π is a continuous representation when $\mathrm{Iso}(C_c(X))$ has the strong operator topology.

Proof:

To start off, recall *example 1.2.4* on page 284 for why $\pi(g) \in \mathrm{Iso}(C_c(X))$ for each g.

Technically, on page 284 I showed that $\pi(g)$ would be an isometric isomorphism on $BC(X)$. That said, as $x \mapsto g \cdot x$ and $x \mapsto g^{-1} \cdot x$ are continuous maps, we know that $\mathrm{supp}(f)$ is compact iff $g \cdot \mathrm{supp}(f)$ is compact. Hence, $\pi(g)$ maps $C_c(X)$ bijectively into $C_c(X)$.

Meanwhile, it's easy to see π is a group homomorphism. So, all that's left to show is that π is continuous, and to do that it suffices by the prior exercise to show π is continuous at $e \in G$. Thus, we want to show that if $f \in C_c(X)$ and $\varepsilon > 0$ then there is a neighborhood V of e with $\|(\pi(g))(f) - f\|_u < \varepsilon$ for all $g \in V$.

Fortunately, since $\mathrm{supp}(f)$ is compact and X is locally compact, we can find a precompact open set $U \subseteq X$ containing $\mathrm{supp}(f)$. Then for each $x \in \mathrm{supp}(f)$, continuity of the group action implies there is an open neighborhood U_x of x in X and an open neighborhood W_x of e in G such that $W_x \cdot U_x \subseteq U$.

$W_x \times U_x$ is an open neighborhood of (e, x) which is in the preimage of U with respect to the group action.

Next, by the compactness of $\mathrm{supp}(f)$ there exists $x_1, \ldots, x_n \in \mathrm{supp}(f)$ such that $\mathrm{supp}(f) \subseteq \bigcup_{i=1}^n U_{x_i}$. In turn, $W := \bigcap_{i=1}^n W_{x_i}$ is an open neighborhood of e such that $W \cdot \mathrm{supp}(f) \subseteq U$. And in particular, after making W symmetric (remember *proposition 11.1(b)* from Folland), we can say that $\mathrm{supp}((\pi(g))(f)) \subseteq \overline{U}$ for all $g \in W$.

Now we just need to find an open neighborhood $V \subseteq W$ of e such that:

$$|f(g^{-1} \cdot x) - f(x)| < \varepsilon \text{ for all } x \in \overline{U}.$$

To do that, note by the continuity of f that for each $x \in \overline{U}$ we can choose an open neighborhood U'_x of x such that $|f(y) - f(x)| < \varepsilon/2$ for all $y \in U'_x$. Then by the continuity of the group action, we can find open neighborhoods Z_x of e in G and Y_x of x in X such that $Z_x \cdot Y_x \subseteq U'_x$.

Using the compactness of $\overline{U}$, choose a new finite set $x_1, \ldots, x_m \in \overline{U}$ such that $\overline{U} \subseteq \bigcup_{i=1}^m Y_{x_i}$. Then set $W' = W \cap \bigcap_{i=1}^m Z_{x_i}$ and define $V := W' \cap (W')^{-1}$. Now V is an open neighborhood of e in G. Also if $g \in V$ and $y \in \overline{U}$, then because $y \in Y_{x_i} \subseteq U_{x_i}$ for some i (which also means $g^{-1} \cdot y \in Y_{X_i} \subseteq U_{x_i}$), we know that:

$$|f(g^{-1} \cdot y) - f(y)| \leq |f(g^{-1}y) - f(x_i)| + |f(x_i) - f(y)| < \varepsilon/2 + \varepsilon/2. \blacksquare$$

A basic corollary of the above proposition is that every $f \in C_c(\mathbb{R}^n)$ is uniformly continuous. After all, we can apply the above proposition to the action $\mathbb{R}^n \curvearrowright \mathbb{R}^n$ by translation. That said, I already proved this corollary in my notes from Spring 2025.

In a similar vein, the next two results will prove a generalization of Folland proposition 8.5 from my math 240c notes from last spring.

(Zimmer) exercise 1.13: Suppose $\mathcal{X}$ is a normed topology and $\langle T_i \rangle_{i \in I}$ is a net in $B(\mathcal{X})$. Also suppose that $T \in B(\mathcal{X})$ and there exists $C > 0$ with $\|T_i\|, \|T\| < C$ for all $i \in I$. Then $T_i \to T$ operator strongly if and only if there is a dense set $\mathcal{X}_0 \subseteq \mathcal{X}$ such that $T_i x \to Tx$ for all $x \in \mathcal{X}$.

Proof:

The $(\Longrightarrow)$ direction is trivial. As for the other direction, consider any $x \in \mathcal{X}$ and let $\{x_n\}_{n \in \mathbb{N}}$ be a sequence in $\mathcal{X}_0$ converging to x. Then, we know that $\|x_n - x\| < \varepsilon/2C$ for some $n \in \mathbb{N}$. In turn:

$$\begin{aligned}
\|T_i x - Tx\| &\leq \|T_i x - T_i x_n\| + \|T_i x_n - T x_n\| + \|T x_n - Tx\| \\
&\leq \|T_i\| \|x - x_n\| + \|T_i x_n - T x_n\| + \|T\| \|x_n - x\| \\
&< C \tfrac{\varepsilon}{2C} + \|T_i x_n - T x_n\| + C \tfrac{\varepsilon}{2C} = \|T_i x_n - T x_n\| + \varepsilon.
\end{aligned}$$

And since $T_i x_n \to T x_n$, we thus know that $\|T_i x - Tx\| < 2\varepsilon$ eventually for all $\varepsilon > 0$. $\blacksquare$

<u>Proposition 1.3.10:</u> Let G be a topological group acting continuously on an LCH space X. Suppose μ is a Radon measure on X which is G-invariant (i.e φ_g is measure preserving for each $g \in G$ [recall *page 263*]). Then for $1 \leq p < \infty$, the representation $\pi : G \to \mathrm{Iso}(L^p(X))$ given by $(\pi(g)(f))(x) := f(g^{-1} \cdot x)$ is continuous for the strong operator topology.

Proof:

To see that π really does map G into $\mathrm{Iso}(L^p(X))$ just apply lemma 2.6 on *page 264* to $|f(x)|^p$ and $|f(g^{-1} \cdot x)|^p$. Also, π is seen to be a group homomorphism identically as in the last proposition. So, we just need to show that π is continuous for the strong operator topology. This is equivalent to saying that $g \mapsto (\pi(g))(f)$ is a continuous map from G to $L^p(X)$ for all $f \in L^p(X)$.

Fortunately, like in the last proposition it suffices to show that $\|(\pi(g_i))(f) - f\|_p \to 0$ for any net $\langle g_i \rangle_{i \in I}$ converging to e in G. Also, by the prior exercise plus the fact that $C_c(X)$ is dense in $L^p(X)$ for $1 \leq p < \infty$ (see my math 240c notes), it suffices to assume that $f \in C_c(X)$.

But now we already know from the proof of the last proposition that $(\pi(g_i))(f) \to f$ uniformly and that we can find a compact set $\overline{U}$ such that $\mathrm{supp}(f) \subseteq \overline{U}$ and $\mathrm{supp}((\pi(g_i))(f)) \subseteq \overline{U}$ eventually. This implies L^p convergence. ■

Note that every representation $\pi : G \to \mathrm{Aut}(\mathcal{X})$ is a group action $G \times \mathcal{X} \to \mathcal{X}$ by linear automorphisms (i.e. $g \cdot x = (\pi(g))x$) and vice versa. Thus, when we talk about fixed points of representations we are really talking about the fixed points of their induced group action.

<u>Kakutani-Markov Fixed Point Theorem:</u> Let $\mathcal{X}$ be a topological vector space whose topology is defined by a sufficient family of seminorms. Suppose G is an abelian group and $\pi : G \to \mathrm{Aut}(\mathcal{X})$ is a representation. Let $A \subseteq \mathcal{X}$ be a compact convex set that is G-invariant (i.e. $(\pi(g))(A) \subseteq A$ for all $g \in G$). Then there is a G-fixed point in A.

Proof:

For each $g \in G$ and $n \geq 0$, define $M_{n,g} \in B(\mathcal{X})$ by $M_{n,g} = \frac{1}{n}\sum_{i=0}^{n-1} \pi(g^i)$. Since A is convex and G-invariant, we have that $M_{n,g}(A) \subseteq A$ for all n, g. Next, let G^* be the semigroup of operators generated by $\{M_{n,g} : n \geq 0, g \in G\}$ (i.e. the collection of all finite compositions of such operators).

(Technically G^* would be a monoid as $M_{1,e} = \mathrm{Id}$. That said, we don't necessarily know if the operators in G^* have inverses, or if they do have inverses which aren't in G^*).

Since G is abelian we know that G^* is commutative. After all:

$$\begin{aligned} M_{n_1,g_1}M_{n_2,g_2} &= \left(\frac{1}{n_1}\sum_{i=0}^{n_1-1} \pi(g_1^i)\right)\left(\frac{1}{n_2}\sum_{j=0}^{n_2-1} \pi(g_2^j)\right) \\ &= \frac{1}{n_1 n_2}\sum_{i=0}^{n_1-1}\sum_{j=0}^{n_2-1} \pi(g_1^i)\pi(g_2^j) = \frac{1}{n_2 n_1}\sum_{i=0}^{n_1-1}\sum_{j=0}^{n_2-1} \pi(g_2^j)\pi(g_1^i) \end{aligned}$$

$$= \left(\frac{1}{n_2}\sum_{j=0}^{n_2-1}\pi(g_2^j)\right)\left(\frac{1}{n_1}\sum_{i=0}^{n_1-1}\pi(g_1^i)\right)$$

$$= M_{n_2,g_2}M_{n_1,g_1}$$

Now we claim that $F := \bigcap_{T\in G^*} T(A) \neq \emptyset$ and that every element of F is a G-fixed point. To see this, note that as A is compact and each $T \in G^*$ is continuous, we know that $T(A)$ is compact for each $T \in G^*$. Then since $\mathcal{X}$ is Hausdorff, we know that $T(A)$ is closed for each $T \in G^*$. And since $T(A) \subseteq A$ for each $T \in G^*$ and A is compact, we will be able to use the finite intersection property to prove F is nonempty if for any finite set $T_1, \ldots, T_n \in G^*$ we have that $\bigcap_{i=1}^n T_i(A) \neq \emptyset$.

Fortunately, if we let $S = T_1 \circ \cdots \circ T_n \in G^*$ then we know that:

$$S(A) = T_1(T_2 \circ \cdots \circ T_n(A)) \subseteq T_1(A).$$

Also note that as G^* is commutative, we can rewrite $S = T_i \circ T_1 \circ \cdots \circ T_{i-1} \circ T_{i+1} \circ \cdots \circ T_n$ and apply the prior line's reasoning to get that $S(A) \subseteq T_i(A)$ for any i. Therefore:

$$\emptyset \neq S(A) \subseteq \bigcap_{i=1}^n T_i(A).$$

With that we know F is nonempty. So, we now consider any $y \in F$. By the definition of F, for each $n \geq 0$ and $g \in G$ there is some $x \in A$ such that $y = \frac{1}{n}(x + \cdots + \pi(g^{n-1})x)$. In turn:

$$\pi(g)y - y = \frac{\pi(g)x+\cdots+\pi(g^n)x}{n} - \frac{x+\cdots+\pi(g^{n-1})x}{n} = \frac{\pi(g^n)x - x}{n}$$

Now if p is any of the seminorms defining the topology, then for each n we have that $p(\pi(g)y - y) \leq 2B_p/n$ where $B_p = \sup\{p(a) : a \in A\}$. (Note that B_p is well-defined (and finite) since A is compact and p is continuous.) As n was arbitrary, we can take $n \to \infty$ to get that $p(\pi(g)y - y) = 0$. And since p was an arbitrary seminorm in the sufficient family defining the topology on $\mathcal{X}$, we have that $\pi(g)y = y$. Finally, as g was arbitrary we know y is a G-fixed point. ■

As a side note, I proved a similar fact on the second problem set for math 200a (see *page 287*). To briefly compare the two results, this result doesn't assume G is a finite group like the homework problem did, or that we are working in $\mathbb{R}^n$. That said, the homework problem didn't require A to be compact and G to be abelian. Also, the homework problem assumed G acted by affine transformations as opposed to linear transformations.

Although, it is worth noting that none of the reasoning in this proof breaks down if we assume $\pi(g)$ is a continuous affine transformation on $\mathcal{X}$ instead of a linear transformation.

As shown on page 286, every affine transformation is a linear transformation minus a constant and vice versa. So, we know that summing, scaling, and composing continuous affine transformations always yields a continuous affine transformation. Furthermore, our evaluation of $\pi(g)y - y$ doesn't change and we still have that each $M_{n,g}$ maps A into A.

One particular compact convex subset of a topological vector space we care about is the set of probability measures $M(X) \subseteq C(X)^*$ (where X is a compact metric space). (It is easy to see that $M(X)$ is convex and we know from *corollary 1.1.29 (see page 283)* that $M(X)$ is compact in the weak* topology.)

Before continuing on, I want to actually prove that all probability measures on X are regular (and thus Radon) since I forgot to show that back on *page 283*.

Claim: If X is a compact metric space, then X is second countable.

Proof:

Since X is totally bounded, we know that there is a finite collection $\mathcal{U}_n$of open balls of radius $1/n$ covering X for all $n \in \mathbb{N}$. Then we claim that $\mathcal{U} = \bigcup_{n\in\mathbb{N}} \mathcal{U}_n$ is a countable base for X. After all, we already know each $x \in X$ is contained in some $B \in \mathcal{U}$. On the other hand, suppose $x \in X$ and $V \subseteq X$ is an open set containing x. Then we know $B_r(x) \subseteq V$ for some $r > 0$. In turn, if we fix $k \in \mathbb{N}$ such that $1/k < r/2$ then we know there is some ball B of radius $1/k$ in $\mathcal{U}$ satisfying that $x \in B \subseteq B_r(x) \subseteq V$.

Consequently, we know by Folland theorem 7.8 and the comment right after it in my math 240c notes that any Borel measure on a compact metric space X that is finite on compact sets is regular. In particular, this means all finite Borel measures on compact metric spacess are regular.

(Definition 1.3.12:) If $\pi : G \to \mathrm{GL}(V)$ is a representation, we define the adjoint of π by $\pi^* : G \to \mathrm{GL}(V^*)$ where $\pi^*(g) := \pi(g^{-1})^*$. ($\pi(g^{-1})^*$ denotes the transpose of $\pi(g^{-1})$ [see *page 102* and *page 300*.])

Note that π^* is a representation. After all, recalling *proposition 1.3.18 on page 103*, we know that:

$$\pi^*(gh) = \pi(h^{-1}g^{-1})^* = (\pi(h^{-1})\pi(g^{-1}))^* = \pi(g^{-1})^*\pi(h^{-1})^* = \pi^*(g)\pi^*(h)$$

When $V = \mathcal{X}$ is a topological vector space (so that our definition of a representation requires π to map G into $\mathrm{Aut}(\mathcal{X})$), then π^* still defines a representation if either:

- $\mathcal{X}^*$ has the weak* topology;
- $\mathcal{X}$ is normed and $\mathcal{X}^*$ has the operator norm topology.

To see this just recall *page 301* and also note that $\pi^*(g)$ has the continuous inverse $\pi^*(g^{-1})$ for all $g \in G$.

Now let's return to considering a compact metric space X and suppose G is a topological group acting continuously on X. By using *proposition 1.3.9* we get a representation $\pi : G \to \mathrm{Iso}(C(X))$ given by $(\pi(g))(f) := f(g^{-1} \cdot x)$. Then, by taking the adjoint of π we get a representation $\pi^* : G \to \mathrm{Aut}(C(X)^*)$ given by:

$$(\pi^*(g))(\lambda) = (\pi(g^{-1}))^*(\lambda) = \lambda \circ \pi(g^{-1}).$$ (FYI this construction continues on the next page…)

As a bit of notation, if μ is a measure on X and $\varphi : X \to Y$ is a measurable function, then I'll follow Zimmer and denote $\varphi_*\mu(E) \coloneqq \mu(\varphi^{-1}(E))$ to be the pushforward / image measure of the function φ. Also, as G is acting continuously on X I'll let $g_*\mu$ denote the pushforward of $\varphi_g(x) \coloneqq g \cdot x$ (in other words $g_*\mu(E) = \mu(g^{-1} \cdot E)$).

Now recall by the Riesz-representation theorem that every linear functional in $C(X)^*$ corresponds to integration by a complex Radon measure and vice versa. (Also as mentioned on the prior page, all complex Borel measures on X are Radon.) Therefore, we now ask what does the action of π^* look like on a given measure μ?

Note for any $f \in C(X)$ that:

$$((\pi^*(g))(\int \cdot \, \mathrm{d}\mu))(f) = \int (\pi(g^{-1}))(f)\mathrm{d}\mu = \int f(g \cdot x)\mathrm{d}\mu = \int f \mathrm{d}g_*\mu$$

Thus $(\pi^*(g))(\mu) = g_*\mu$ for all $\mu \in C(X)^*$.

Importantly, recall that pushforwards preserve the positivity of a measure and the total measure of the space. Thus $M(X)$ is invariant under π^* and we arrive at the following result:

Corollary 2.1.6: Let G be an abelian group acting continuously on a compact metric space X. Then there is a G-invariant probability measure on X.

Proof:

Apply the Kakutani-Markov fixed point theorem to $M(X) \subseteq C(X)^*$ (equipped with the weak* topology) and the representation $\pi^* : G \to \mathrm{Aut}(C(X)^*)$ described above. Then the resulting probability measure satisfies that $g_*\mu(E) \coloneqq \mu(g^{-1} \cdot E) = \mu(E)$ for all $g \in G$. ■

11/27/2025

It's Thanksgiving so let's do some more representation theory before I have to go back to taking notes for the classes with actual due dates. (In other words I'm still taking math 241a notes…)

To start off, I already proved in math math 240b homework that there is only one norm topology for any finite dimensioned real or complex vector space. Now, I would like to prove the more general statement that there is only one Hausdorff topology on a finite dimensioned real or complex vector space $\mathcal{X}$ such that $\mathcal{X}$ is a topological vector space.

First, I'd like to go over some quick facts about topological vector spaces now that I've learned some topological group theory.

Lemma 1: If G, G' are topological groups and $\phi : G \to G'$ is a group homomorphism, then ϕ is continuous iff ϕ is continuous at $e \in G$.

Proof:

One direction is trivial. Meanwhile, to show the other direction suppose $\langle g_i \rangle_{i\in I}$ is any net in G converging to g. Then $\phi(g_i) \to \phi(g)$ if and only if $\phi(g_i)\phi(g)^{-1} = \phi(g_i g^{-1}) \to e' \in G'$. But note that $g_i \to g$ implies that $g_i g^{-1} \to e \in G$. So, if ϕ is continuous at $e \in G$ then we know that $\phi(g_i g^{-1}) \to \phi(e) = e'$. ■

A common application of this fact is that if $T : \mathcal{X} \to \mathcal{Y}$ is a linear map between topological vector spaces, then T is continuous iff T is continuous at 0. After all, topological vector spaces are just abelian topological groups which also satisfy that scalar multiplication by numbers other than ± 1 are continuous.

<u>Lemma 2:</u> A topological vector space $\mathcal{X}$ is Hausdorff iff $\{0\}$ is closed.

Proof:

By (Folland) proposition 11.3 on *page 351*, we know that $\mathcal{X}$ is Hausdorff iff $\mathcal{X}$ is T_1. And the latter is true iff every singleton is closed. Finally, note that as translation on $\mathcal{X}$ is a homeomorphism, we have that all singletons in $\mathcal{X}$ are closed iff $\{0\}$ is closed. ■

<u>Lemma 3:</u> Hausdorff topological groups (and therefore Hausdorff topological vector spaces) are regular (i.e. T_3). Specifically, given any closed subset F and compact subset K of a topological group G such that $F \cap K = \emptyset$, we claim there is an open neighborhood V of e such that $KV \cap FV = \emptyset$.

Proof:

Consider any $x \in K$ and note that $U = F^{\mathsf{C}}$ is an open neighborhood of x. In turn, $x^{-1}U$ is an open neighborhood of e. So, by applying (Folland) proposition 11.1(b) and (c) on *page 333*, we can find an open set V_x such that $e \in V_x$, $V_x V_x V_x \subseteq x^{-1}U$, and V_x is symmetric.

Now xV_x is an open neighborhood of x in G. Also we claim that xV_xV_x is disjoint from FV_x. After all, suppose $y \in xV_xV_x \cap FV_x$. Then we can write $y = xv_1v_2 = gv_3$ where $g \in F$ and $v_1, v_2, v_3 \in V$. But now $xv_1v_2v_3^{-1} = g$. This is a contradiction as V_x is symmetric and $V_xV_xV_x \subseteq x^{-1}U$ implies $xV_xV_xV_x \subseteq U = F^{\mathsf{C}}$.

Finally, since K is compact, there are finitely many $x_1, \ldots, x_n \in K$ with $K \subseteq \bigcup_{i=1}^n x_i V_{x_i}$. So, put $V = \bigcap_{i=1}^n V_i$. Then:

- $KV \subseteq (\bigcup_{i=1}^n x_i V_{x_i})V = (\bigcup_{i=1}^n x_i V_{x_i} V) \subseteq (\bigcup_{i=1}^n x_i V_{x_i} V_{x_i})$;
- $FV = F(\bigcap_{i=1}^n V_{x_i}) = \bigcap_{i=1}^n FV_{x_i}$.

Since $x_i V_{x_i} V_{x_i} \cap FV_{x_i} = \emptyset$ for all i, we know $KV \cap FV = \emptyset$. ■

Next, I'm going to follow grandpa Rudin for a bit. (see *citation 19* in the bibliography).

<u>Lemma (unnumbered):</u> Suppose V and W are K-vector spaces (where $K = \mathbb{R}$ or $\mathbb{C}$). Then suppose $T : V \to W$ is a linear map, $A \subseteq V$, and $B \subseteq W$. Then:

- If A is a subspace, convex set, or balanced set then so is $T(A)$.
- If B is a subspace, convex set, or balanced set then so is $T^{-1}(B)$.

Proof:

I think the only nontrivial claim to see here is that the preimage of a balanced set with respect to T is still balanced. (See *page 230* for a reminder of what balanced means.) Suppose $c \in K$ With $|c| \leq 1$. Then if $x \in cT^{-1}(B)$ we know that $T(c^{-1}x) \in B$. In turn, $T(x) \in cB \subseteq B$ and hence $x \in T^{-1}(B)$. Thus, $cT^{-1}(B) \subseteq T^{-1}(B)$. ■

Theorem 1.18 Let λ be a linear functional on a Hausdorff topological vector space $\mathcal{X}$ such that $\lambda(x) \neq 0$ for some $x \in \mathcal{X}$. Then the following are equivalent:

(a) λ is continuous,

(b) $\ker(\lambda)$ is closed,

(c) $\ker(\lambda)$ is not dense in $\mathcal{X}$,

(d) λ is bounded on some neighborhood V of 0 in $\mathcal{X}$ (meaning $\sup_{x \in V} \lambda(x) < \infty$).

$(a \implies b)$
Since λ is continuous and $\{0\}$ is closed, we know that $\lambda^{-1}(\{0\}) = \ker(\lambda)$ is closed.

$(b \implies c)$
Suppose $\ker(\lambda)$ is closed. Then since there exists $x \in \mathcal{X}$ such that $x \notin \ker(\lambda)$, we know that $\ker(\lambda)$ is not dense in $\mathcal{X}$.

$(c \implies d)$
Suppose $x \notin \overline{\ker(\lambda)}$. Then by lemma 3 on the prior page plus the reasoning on *pages 230-232*, we can find a balanced neighborhood V of 0 such that $x + V$ and $\overline{\ker(\lambda)} + V$ are disjoint open sets containing x and $\overline{\ker(\lambda)}$ respectively. But also note that as V is balanced, we know that $\lambda(V)$ is balanced. The only way this is possible is if $\lambda(V)$ is an open or closed ball in $\mathbb{R}$ or $\mathbb{C}$ about 0 or if $\lambda(V)$ is all of $\mathbb{R}$ or $\mathbb{C}$.

Suppose the latter is true (that $\lambda(V) = \mathbb{R}$ or $\mathbb{C}$). Then there exists $y \in V$ such that $\lambda(y) = -\lambda(x)$. But then $x + y \in \overline{\ker(\lambda)}$ and that contradicts that $x + V$ is disjoint from $\overline{\ker(\lambda)} + V$. Hence, we conclude that λ is bounded on V.

$(d \implies a)$
Let $M > 0$ be such that $|\lambda(x)| \leq M$ for all $x \in V$. Then for any $r > 0$ we know that $W := (r/M)V$ is a neighborhood of 0 satisfying that $|\lambda(x)| < r$ for all $x \in W$. Hence, λ is continuous at 0. ■

Lemma 1.20: If $\mathcal{X}$ is a topological K-vector space (where $K = \mathbb{R}$ or $\mathbb{C}$) and $f : K^n \to \mathcal{X}$ is linear then f is continuous.

Proof:
Let $e_1, \ldots, e_n$ be the standard basis for K^n. Then put $u_k = f(e_k)$ for each k. Now $f(z_1, \ldots, z_n) = z_1 u_1 + \ldots + z_n u_n$ and the latter expression is continuous since $\mathcal{X}$ is a topological vector space. ■

Theorem 1.21: If n is a positive integer and $\mathcal{Y}$ is an n-dimensional subspace of a Hausdorff topological K-vector space $\mathcal{X}$, then:

(a) every bijective linear map f from K^n to $\mathcal{Y}$ is a homeomorphism,

(b) $\mathcal{Y}$ is closed.

Proof:
Consider $S^{n-1} = \{z \in K^n : |z| = 1\}$ and $B^n = \{x \in K^n : |z| < 1\}$. Then put $F := f(S^{n-1})$. Since S^{n-1} is closed and f is continuous by the last lemma, we know F is a closed set. (In fact F is compact.)

Next, as f is injective, $f(0) = 0$ and $0 \notin S^{n-1}$, we know $0 \notin F$. So, using the regularity of $\mathcal{X}$ we can find an open neighborhood V of 0 in $\mathcal{X}$ that is disjoint from F. Also, we'll choose F to be balanced. Then, $E := f^{-1}(V) = f^{-1}(V \cap \mathcal{Y})$ is an open set disjoint from S^{n-1} since f is injective. Also, E is balanced by the prior unnumbered lemma.

Claim: A balanced set A in a topological vector space is path-connected.

> Why? Given any $x, y \in A$, we know that the functions $\gamma_1(t) = (1-t)x$ and $\gamma_2(t) = ty$ (where $0 \leq t \leq 1$) are continuous paths going from x to 0 and 0 to y respectively. Also, since A is balanced, we know that $\gamma_1([0,1])$ and $\gamma_2([0,1])$ are contained in A. So, just concatenate those two paths.

Thus E is connected. And as $0 \in B^n \cap E$ and E is disjoint from $\partial B^n = S^{n-1}$, we must have that E is entirely contained in B^n. This proves that there is an open neighborhood $V \cap \mathcal{Y}$ of 0 in $\mathcal{Y}$ such that $\|f^{-1}(x)\| < 1$ for all $x \in V \cap \mathcal{Y}$. Finally, note that f^{-1} is an n-tuple of linear functionals on $\mathcal{Y}$. By applying theorem 1.18 on the prior page to each linear functional in that n-tuple, we can thus show that f^{-1} is continuous. Hence, f is a homeomorphism.

As for showing (b), consider any $p \in \overline{\mathcal{Y}}$ and let f and V be as in the prior reasoning.

> Note that it is a basic fact of linear algebra that if $\mathcal{Y}$ is n-dimensional K-vector space then there always exists a bijective linear map $f : K^n \to \mathcal{Y}$.

For some $t > 0$ we have that $p \in tV$. In turn, we claim that $p \in \overline{\mathcal{Y} \cap tV}$.

> To see why, observe the following result.
>
> Lemma: If X is a topological space and $A, B \subseteq X$ satisfy that B is open, then $\overline{A} \cap B \subseteq \overline{A \cap B}$.
>
> > Proof:
> > Suppose $x \in \overline{A} \cap B$. If $x \in A$ then we trivially have that $x \in \overline{A \cap B}$. Meanwhile, suppose $x \notin A$ and let U be any neighborhood containing x. Since $U \cap B$ is a neighborhood of x and $x \in \overline{A}$, we must have that $(U \cap B) \cap A \neq \emptyset$. Or in other words, $U \cap (A \cap B) \neq \emptyset$. As U was arbitrary we've proven that $x \in \mathrm{Acc}(A \cap B)$.

But now $\overline{\mathcal{Y} \cap tV} \subseteq \overline{f(tB^n)} \subseteq \overline{f(t\overline{B^n})} = f(t\overline{B^n})$ (the final manipulation works because f is continuous and $t\overline{B^n}$ is closed). Hence, $p \in f(t\overline{B^n}) \subseteq \mathcal{Y}$. This proves that $\overline{\mathcal{Y}} = \mathcal{Y}$. ■

Well I'm on the topic of classifying topological vector spaces, something that's been bugging me in the back of my mind is whether there are non-locally convex topological vector spaces. As it turns out, yes there are.

Let $(X, \mathcal{M}, \mu)$ be a finite measure space. Then recall (Folland) exercise 2.32 that I did for homework in math 240a [see pages 58-59 of my math 240a latex notes]. There I proved that:

> If (X, μ) is a finite measure space and $L^0(X)$ is the space of all complex-valued measurable functions on X (where f and g are identified with each other if $f = g$ a.e.), then $\rho(f, g) = \int \frac{|f-g|}{1+|f-g|}\mathrm{d}\mu$ is a metric on $L^0(X)$ such that $f_n \to f$ with respect to ρ iff $f_n \to f$ in measure.

Also, by applying exercise 2.38 in that same homework assignment [see pages 60-61 of my math 240a latex notes], we know that the topology of convergence in measure does turn $L^0(X)$ into a topological vector space.

Given a measure space $(X, \mathcal{M}, \mu)$ we say a set $A \in \mathcal{M}$ is an atom if $\mu(A) > 0$ and for any $B \in \mathcal{M}$ with $B \subseteq A$, either $\mu(B) = 0$ or $\mu(B) = \mu(A)$. We say a measure space is non-atomic or atomless if it has no atoms.

Claim: The Lebesgue measure m on $\mathbb{R}^n$ is atomless.

Proof:

Suppose for the sake of contradiction that A is an atom of the Lebesgue measure. Then by the inner regularity of the Lebesgue measure we can find a compact subset $K \subseteq A$ such that $m(K) = m(A)$ and K is an atom. But importantly, since K is compact we know that K is bounded. Hence, there is some $R > 0$ such that $K \subseteq B_R(0)$.

Now define the function $f : [0, \infty) \to [0, \infty)$ by $f(r) = m(B_r(0) \cap K)$. Then f is continuous. After all, if $r > s$ then there is a constant $\sigma(S^{n-1})$ such that:

$$|f(r) - f(s)| \leq m(\{x : s \leq |x| < r\}) = (r^n - s^n)\sigma(S^{n-1})$$

Clearly the latter expression goes to 0 as $r \to s$ for any $s \in [0, \infty)$.

But now since f is continuous, $f(0) = 0$ and $f(R) = m(K)$, we must have by the intermediate value theorem that f attains every value between 0 and $m(K)$ on the interval $[0, R]$. This contradicts that K is an atom. ■

More generally, if X is an LCH space and μ is a regular measure on X, then any atom $A \subseteq X$ contains a point $\{a\}$ such that $\mu(\{a\}) = \mu(A)$

Proof:

By the inner regularity of μ we can find a compact set $K \subseteq A$ such that $\mu(K) = \mu(A)$ and K is an atom. Now, suppose for the sake of contradiction that $\mu(\{x\}) = 0$ for all $x \in K$. Then, for each $x \in K$ we can find an open set V_x containing $x \in K$ such that $\mu(V_x) < \mu(K)/2$. In turn, since $V_x \cap K$ is a subset of K with measure less than $\mu(K)$ and K is an atom, we know that $\mu(V_x \cap K) = 0$. But now as $\{V_x \cap K\}_{x \in K}$ is an open cover (relative to K) of the compact set K, we know that there is a finite set $x_1, \ldots, x_n$ with $K \subseteq \bigcup_{i=1}^n V_{x_i} \cap K$. This is a contradiction as it implies that:

$$0 < \mu(K) \leq \sum_{i=1}^n \mu(V_{x_i} \cap K) = 0.$$

So, we conclude that there must exist $x \in K \subseteq A$ with $\mu(\{x\}) > 0$. Since, A is an atom, we must have that $\mu(\{x\}) = \mu(A)$. ■

Also note that then $\mu(A - \{x\}) = 0$.

Thus as a corollary, if X is an LCH space and μ is a nonzero regular measure on X such that every singleton is a null set, we know that μ is atomless.

An obvious property of atomless measures is as follows:

Suppose $(X, \mathcal{M}, \mu)$ is an atomless measure space and $E \in \mathcal{M}$ satisfies that $\mu(E) > 0$. Then there is a decreasing sequence of sets $E = E_1 \supseteq E_2 \supseteq \cdots$ such that $\mu(E_1) > \mu(E_2) > \cdots > 0$.

Using the above property, we can also show that if $(X, \mathcal{M}, \mu)$ is an atomless measure space and $E \in \mathcal{M}$ satisfies that $\mu(E) > 0$, then there is a countabley infinite collection of disjoint sets $\{F_n\}_{n\in\mathbb{N}}$ with positive measure such that $E = \bigcup_{n\in\mathbb{N}} F_n$.

Proof:

Let $E = E_1 \supseteq E_2 \supseteq \cdots$ be a decreasing sequence of sets such that $\mu(E_1) > \mu(E_2) > \cdots > 0$. Then define $F_1 \coloneqq (E_1 - E_2) \cup (\bigcap_{n\in\mathbb{N}} E_n)$ and $F_k \coloneqq E_k - E_{k+1}$ when $k \geq 2$. ■

<u>Proposition:</u> Let $(X, \mathcal{M}, \mu)$ be a finite atomless measure space. Then $L^0(X)$ contains no open convex sets.

<u>Lemma:</u> A topological vector space $\mathcal{X}$ has no open convex sets if and only if $\mathcal{X}^* = \{0\}$.

I already went over the $(\Longrightarrow)$ direction in an observation on *page 252*. To show the other direction, note that if $\mathcal{X}$ had a convex open set then we'd able to find a nontrivial continuous seminorm p on $\mathcal{X}$ (via a Minkowski functional). Next, let $x \in \mathcal{X}$ satisfy that $p(x) \neq 0$ and define a linear functional f on the span of x by $f(cx) = c \cdot p(x)$. By the Hahn-Banach theorem we are then able to get a nonzero linear functional F on $\mathcal{X}$ such that $|F(y)| \leq p(y)$ for all $y \in \mathcal{X}$.

Finally, note that F is continuous at 0. After all, suppose $\langle y_i \rangle_{i\in I}$ is any net converging to 0. Since p is continuous, we know that $p(y_i) \to p(0) = 0$. In turn, $F(y_i) \to 0$ as well. But now this means that $F \in \mathcal{X}^*$. ■

Now we go to prove our claim about $L^0(X)$ by showing that $L^0(X)$ has no nonzero continuous linear functionals. Suppose $\lambda \in L^0(X)^*$ and let $\pi : L^1(X) \hookrightarrow L^0(X)$ be the natural inclusion map. Then note that $\lambda \circ \pi \in L^1(X)^*$.

To see why just note that π is a linear map from $L^1(X)$ to $L^0(X)$ that is continuous because L^1-convergence implies convergence in measure.

By (Folland) theorem 6.15 in my math 240b notes plus the fact that μ is finite, we know that $L^1(X)^* \cong L^\infty(X)$. Hence, there is a function $g \in L^\infty(X)$ such that $\lambda \circ \pi(f) = \int g f \mathrm{d}\mu$ for all $f \in L^1(X)$.

Now suppose for the sake of contradiction that $g \neq 0$. Then there is some $\varepsilon > 0$ such that $E \coloneqq \{x \in X : |g(x)| > \varepsilon\}$ has positive measure. In turn, as μ is atomless we can a sequence of disjoint sets $\{E_n\}_{n\in\mathbb{N}}$ with positive measure such that $E = \bigcup_{n\in\mathbb{N}} E_n$.

Let $f_n = \sum_{k=1}^{n} \frac{\chi_{E_k}}{\mu(E_k)g}$ and then set $f = \lim_{n\to\infty} f_n$. Importantly each f_n is in $L^1(X)$. Therefore $\lambda(f_n) = \int g f_n \mathrm{d}\mu = \sum_{k=1}^{n} \int_{E_k} \frac{g}{\mu(E_n)g} \mathrm{d}\mu = n$. But also note that since $f_n \to f$ pointwise and μ is finite, we know that $f_n \to f$ in measure.

(See *pages 213-214*...)

Now we have a contradiction since $f \in L^0(X)$ and yet $\lambda(f) = \lim_{n\to\infty} \lambda(f_n) = \infty$. Hence, we conclude that we must have that $g = 0$ a.e. But in turn, we must have that $\lambda(f) = 0$ for all $f \in L^1(X) \subseteq L^0(X)$.

Finally, note that any measurable function f in X is a pointwise limit of a sequence of simple functions $\{\phi_n\}_{n\in\mathbb{N}}$ (which are all in $L^1(X)$.)

(See the theorem on page 44 of my latex math 240a notes...)

By the reasoning on *pages 213-214* again, we know that that $\phi_n \to f$ in measure. So, $L^1(X)$ is dense in $L^0(X)$. And since λ is continuous and $\lambda = 0$ on a dense subset of it's domain, we know λ is the zero functional. ■

By the way, I'm not smart enough to have come up with this proof by myself. Instead, I adapted a proof from math stack exchange by the user David Gao (see *citation 20 in the bibliography*).

One more note I want to make before stopping for the day is that $L^0(X)$ is complete. After all, we proved in math 240a that if a sequence $\{f_n\}_{n\in\mathbb{N}}$ is Cauchy in measure (meaning $f_n - f_m \to 0$ in measure as $m, n \to \infty$), then there exists $f \in L^0(X)$ such that $f_n \to f$ in measure. (Recall Folland real analysis theorem 2.30).

All of this is to say that there does exist complete Hausdorff topological vector spaces with no continuous seminorms. Furthermore, those spaces aren't entirely useless to study either.

11/30/2025

Unfortunately, since Thanksgiving break is almost over I now need to return to looking at math 220 and math 200 stuff now.

Math 220a Notes:

f has an isolated singularity at $z = a$ if f is analytic in $B_R(a) - \{a\}$ for some $R > 0$. Recall from my notes from last Spring that an isolated singularity at a will either be:

- a removable singularity, meaning there exists an analytic function g on $B_R(a)$ with $g = f$ on $B_R(a) - \{a\}$;
- a pole of order m, meaning that there are complex numbers $A_1, \ldots, A_m$ and an analytic function g on $B_R(a)$ such that $f(z) = \sum_{k=1}^{m} \frac{A_k}{(z-a)^k} + g(z)$.
- an essential singularity, meaning that for any $0 < r \leq R$, $f(B_r(a) - \{a\})$ is dense in the complex plane.

(Conway) Proposition V.1.2: If f has an isolated singularity at $z = a$, then the singularity is removable iff $\lim_{z\to a}(z-a)f(z) = 0$.

$(\Longrightarrow)$

Suppose g is analytic on $B_R(a)$ and $g = f$ on $B_R(a) - \{a\}$ for some $R > 0$. Then fix $r < R$. Since $\overline{B_r(a)}$ is compact and g is continuous, we know that g and in turn f is bounded on $\overline{B_r(a)} - \{a\}$. Therefore, it must be the case that $\lim_{z\to a}(z-a)f(z) = 0$.

$(\Longleftarrow)$

Define $g(z) = (z-a)f(z)$ for $z \neq a$ and $g(a) = 0$. Then by assumption, we know that g is a continuous function on $B_R(a)$ for some $R > 0$. Also g is analytic on $B_R(a) - \{a\}$. So, by either following the reasoning in *exercise IV.5.9* or just citing a proposition in my complex analysis notes from Spring 2025, we also know that g is analytic at $\{a\}$. Thus, $g(z) = (z-a)h(z)$ for some analytic function h defined on $B_R(a)$ since a is a zero of g. Yet as $(z-a)h(z) = (z-a)f(z)$ for all $z \in B_R(a) - \{a\}$, this must mean h satisfies all the requirements to prove that a is a removable singularity of f. ■

If an analytic function f on a region G has a pole of order m at $z = a$ (meaning $f(z) = \sum_{k=1}^{m} \frac{A_k}{(z-a)^k} + g(z)$ where $A_1, \ldots, A_m \in \mathbb{C}$ and $g : G \to \mathbb{C}$ is analytic), then we call $\sum_{k=1}^{m} \frac{A_k}{(z-a)^k}$ the singular part of f at $z = a$.

By the way if my notes feel disjoint, it's cause I'm trying not to take notes on things I already took notes on last Spring.

Let $A(a; R_1, R_2)$ denote the annulus $\{z : R_1 < |z-a| < R_2\}$.

Theorem: Let f be analytic in $A(a; R_1, R_2)$. Then $f(z) = \sum_{n\in\mathbb{Z}} a_n(z-a)^n$ with absolute and uniform convergence in $\overline{A(a; r_1, r_2)}$ for every $R_1 < r_1 < r_2 < R_2$.

As a reminder, summing on $\mathbb{Z}$ is defined via integration against the counting measure. Therefore:

- absolute convergence just means $a_n(z-a)^n$ is integrable on $\mathbb{Z}$ for all $r_1 < |z| < r_2$ (technically non-absolute convergence isn't well-defined);
- uniform convergence means that for any $\varepsilon > 0$ there is a finite set $B \subseteq \mathbb{Z}$ such that $|\sum_{n\in B} a_n(z-a)^n - f(z)| < \varepsilon$ for all $r_1 < |z| < r_2$.

Furthermore, the coefficients a_n are given by the formula $a_n = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{(w-a)^{n+1}}\mathrm{d}w$ where γ traces out the circle $|w-a| = r$ counterclockwise for any $R_1 < r < R_2$.

Proof:

If $R_1 < \rho_1 < \rho_2 < R_2$ and γ_1, γ_2 are paths traveling counterclockwise around the circles $|w-a| = \rho_1$ and $|w-a| = \rho_2$ respectively, then $\gamma_1 \sim_{A(a;R_1,R_2)} \gamma_2$. Thus by Cauchy's theorem (third version) on *page 414*, we know that $\int_{\gamma_1} g = \int_{\gamma_2} g$ for any function g that is analytic on $A(a; R_1, R_2)$. This proves that the formula $\frac{1}{2\pi i}\int_\gamma \frac{f(w)}{(w-a)^{n+1}}\mathrm{d}w$ in the theorem is independent of our choice of $r \in (R_1, R_2)$ for all $n \in \mathbb{Z}$.

Next, note by Cauchy's integral formula (see *page 404*) that for any $z \in A(a;\rho_1,\rho_2)$ we have that:

$$f(z) = \frac{1}{2\pi i}\int_{\gamma_2}\frac{f(w)}{w-z}\mathrm{d}w - \frac{1}{2\pi i}\int_{\gamma_1}\frac{f(w)}{w-z}\mathrm{d}w$$

In particular, letting $f_k(z) = \frac{1}{2\pi i}\int_{\gamma_k}\frac{f(w)}{w-z}\mathrm{d}w$ for all $z \in A(a;\rho_1,\rho_2)$, we have that $f(z) = f_2(z) - f_1(z)$ where each f_i is analytic by Leibniz's rule for path integrals (see *page 380-381*).

Let $\rho_1 < r_1 < r_2 < \rho_2$. That way $\overline{A(a;r_1,r_2)} \subseteq A(a;\rho_1,\rho_2)$. Also, at the end we can just take $\rho_i, r_i \to R_i$ to prove the desired theorem.

For f_1 note that when $w \in \{\gamma_1\}$ and $z \in \overline{A(a;r_1,r_2)}$ then $|\frac{w-a}{z-a}| = \frac{\rho_1}{|z-a|} \leq \frac{\rho_1}{r_1} < 1$. Thus $\frac{1}{w-z} = \frac{1}{w-a-(z-a)} = \frac{-1}{z-a}\cdot\frac{1}{1-\frac{w-a}{z-a}} = \frac{-1}{z-a}\sum_{n=0}^{\infty}(\frac{w-a}{z-a})^n$ where convergence is absolute and uniform as w ranges over $\{\gamma_1\}$. Integrating term by term (which we can do since f is bounded on $\{\gamma_1\}$ and the series converges uniformly with respect to w) we get that:

$$\begin{aligned} f_1(z) = \frac{1}{2\pi i}\int_{\gamma_1}\frac{f(w)}{w-z}\mathrm{d}w &= \frac{1}{2\pi i}\int_{\gamma_1}\frac{-f(w)}{z-a}\cdot\sum_{n=0}^{\infty}(\tfrac{w-a}{z-a})^n\mathrm{d}w \\ &= -\sum_{n=0}^{\infty}\frac{1}{2\pi i(z-a)^{n+1}}\int_{\gamma_1}f(w)(w-a)^n\mathrm{d}w \\ &= -\sum_{n=0}^{\infty}\left(\frac{1}{2\pi i}\int_{\gamma_1}\frac{f(w)}{(w-a)^{-(n+1)+1}}\mathrm{d}w\right)(z-a)^{-(n+1)} \\ &= -\sum_{n=1}^{\infty}a_{-n}(z-a)^{-n} \text{ (where } a_n \text{ is defined in the theorem statement).}\end{aligned}$$

Also this sum converges absolutely and uniformly as z ranges $\overline{A(a;r_1,r_2)}$. After all, letting $C > 0$ be such that $|f(w)| \leq C$ for all $w \in \{\gamma_1\}$, we have for any $z \in \overline{A(a;r_1,r_2)}$ that:

$$|a_{-n}(z-a)^{-n}| = \frac{1}{2\pi}|z-a|^{-n}\cdot\left|\int_{\gamma_1}\frac{f(w)}{(w-a)^{-n+1}}\mathrm{d}w\right| \leq \frac{1}{2\pi}|z-a|^{-n}\cdot\frac{C}{\rho_1^{-n+1}}(2\pi\rho_1) \leq C\left(\frac{\rho_1}{r_1}\right)^n$$

Since $\frac{\rho_1}{r_1} < 1$, we can apply standard comparison tests to get the result we want.

Meanwhile for f_2, note that f_2 actually defines an analytic function on $B_{\rho_2}(a)$. So, we can apply identical reasoning as that of *(Conway) Theorem IV.2.8* to get that:

$$f_2(z) = \sum_{n=0}^{\infty}a_n(z-a)^n,$$

...where the above sum converges absolutely and uniformly as z ranges over $\overline{B_{r_2}(a)}$.

Now, it is obvious that $f(z) = f_2(z) - f_1(z) = \sum_{n\in\mathbb{Z}}a_n(z-a)^n$ where the sum converges uniformly as z ranges over $\overline{A(a;r_1,r_2)}$. ■

We call $\sum_{n\in\mathbb{Z}}a_n(z-a)^n$ the Laurent series of f about a. To see why I'm using the definite article "the" here, here is one more relevant theorem.

Theorem: Suppose $f(z) = \sum_{n\in\mathbb{Z}} b_n(z-a)^n = \sum_{n\in\mathbb{Z}} c_n(z-a)^n$ for all $z \in A(a; R_1, R_2)$. Then $b_n = c_n$ for all $n \in \mathbb{Z}$.

Proof:

For hopefully obvious reasons, it suffices to show that if $0 = \sum_{n\in\mathbb{Z}} b_n(z-a)^n$ for all $z \in A(a; R_1, R_2)$ then $b_n = 0$ for all $n \in \mathbb{Z}$. To accomplish this, define:

- $g(z) = \sum_{n=0}^{\infty} b_n(z-a)^n$ when $|z-a| < R_2$
- $g(z) = -\sum_{n=1}^{\infty} b_{-n}(z-a)^{-n}$ when $|z-a| > R_1$.

Imporantly, note that since $0 = \sum_{n\in\mathbb{Z}} b_n(z-a)^n$ when $R_1 < |z-a| < R_2$, we must have that:

$$\sum_{n=0}^{\infty} b_n(z-a)^n = -\sum_{n=1}^{\infty} b_{-n}(z-a)^{-n} \text{ when } R_1 < |z-a| < R_2$$

It follows that g is a well-defined function on $\mathbb{C}$. Also, it is easy to see that g is analytic by noting that that g is the composition of a power series and another analytic function on its entire domain and then applying chain rule. So g is entire. Thirdly, note that if $R_1 < r < R_2$ then we must have that the power series $\sum_{n=0}^{\infty} b_n z^n$ and $\sum_{n=1}^{\infty} b_{-n} z^n$ converge absolutely when $|z| \leq r$. Then since those power series are continuous functions on the compact set $\overline{B_r}(0)$, we know they are bounded on $\overline{B_r}(0)$. Yet this proves that g is also bounded.

Since g is an bounded entire function, we thus must have that g is a constant function by Liouville's Theorem. Also, since $\sum_{n=0}^{\infty} b_n(z-a)^n$ is a power series for g centered at $z = a$, we must have that $b_n = g^{(n)}(a)/n!$ for all $n \geq 0$. Hence, $b_n = 0$ when $n > 0$.

Next, because $-\sum_{n=1}^{\infty} b_{-n}(z-a)^{-n} = b_0$ for all z with $|z-a| > R_1$, we must have that the power series $\sum_{n=1}^{\infty} b_{-n}z^n$ is constant on $B_{R_1}(0) - \{0\}$. Since this set has a limit point, we thus know that $\sum_{n=1}^{\infty} b_{-n}z^n$ is constant everywhere. And by similar reasoning to before, this proves that $b_{-n} = 0$ for all $n \geq 1$.

At last, we've proven $b_n = 0$ for all $n \neq 0$. Then we are done because:

$$0 = \sum_{n\in\mathbb{Z}} b_n(z-a)^n = b_0. \blacksquare$$

12/1/2025

Math 200a Notes:

Suppose $(A, +, \cdot)$ is a unital ring.

- Let $A^\times := \{a \in A : \exists a' \in A \; s.t. \; aa' = a'a = 1\}$. Then $(A^x, \cdot)$ is a group and each $a \in A^\times$ is called a unit. A^x is called the group of units of A.

  To see that this is a group, note that we trivially have that $1 \in A^x$ and A^x is closed under inverses. Also, if $a, b \in A^x$ then $ab(b^{-1}a^{-1}) = (b^{-1}a^{-1})ab = 1$. So $ab \in A^x$.

- $a \in A$ is called a zero divisor if $a \neq 0$ and $\exists b \in A - \{0\}$ such that $ab = 0$ or $ba = 0$.

  You can also imagine how we would define left and right zero divisors. Also I realize this definition didn't require A to be unital but fuck you.

The set of zero-divisors unioned with $\{0\}$ is denoted as $\mathcal{D}(A)$.

Lemma: $0 \neq 1 \Longrightarrow A^x \cap \mathcal{D}(A) = \emptyset$.

Proof:
Suppose $a \in A^x \cap \mathcal{D}(A)$. Then there exists $b \in A$ such that $ab = ba = 1$. But also if $a \neq 0$ then there exists $c \in A - \{0\}$ such that either $ac = 0$ or $ca = 0$. Yet in the former case, $c = 1c = bac = b(0) = 0$. Similarly in the latter case, $c = c1 = cab = 0b = 0$. So we have a contradiction.

Meanwhile, if $a = 0$ then we must have that $0 = 0b = 1$. ■

D is a integral domain if it is a unital commutative ring, $0 \neq 1$, and $\mathcal{D}(D) = \{0\}$.

Corollary: A field F (which is a unital commutative ring such that $F^\times = F - \{0\}$) is automatically an integral domain.

Let $A[x]$ be a polynomial ring. Then any nonzero $f \in A[x]$ can be written as $f(x) = a_n x^n + \ldots + a_1 x + a_0$ where $a_n \neq 0$. We define $\mathrm{lt}(f) := a_n$ and $\deg(f) = n$. Meanwhile if $f = 0$ Then we define $\mathrm{lt}(f) = 0$ and $\deg(f) = -\infty$.

Long Division Theorem: Suppose A is a unital commutative ring, $f, g \in A[x]$, and $\mathrm{lt}(g) \in A^\times$. Then there exists unique $q, r \in A[x]$ such that $f(x) = g(x)q(x) + r(x)$ and $\deg(r) < \deg(g)$.

For a proof see my math 100b notes...

Corollary: If A is a unital commutative ring, $f \in A[x]$, and $a \in A$, then $f(x) = (x - a)q(x) + f(a)$ where $\deg(q) < \deg(f)$.

Proof:
By the long division theorem, we have that $f(x) = (x - a)q(x) + c$ where $c \in A$. Now plug in $x = a$ to get that $f(a) = 0 + c$. ■

Note that $f(a) = 0$ iff $f(x) = (x - a)q(x)$ for some $q \in A[x]$. So, if e_a is the evaluation map then $\ker(e_a) = \langle x - a \rangle$.

As a side note, if $S \subseteq A$ where A is a ring, we write $\langle S \rangle$ to denote the smallest ideal of A containing S. (This is well-defined because the intersection of an arbitrary family of ideals is another ideal). Look at my math 100b notes for more information.

One more note to make is that if A is a ring and $f, g \in A[x]$ then:

- $\deg(f + g) \leq \max(\deg(f), \deg(g))$ with equality iff $\deg(f) \neq \deg(g)$ or $\deg(f) = \deg(g)$ with $\mathrm{lt}(f) \neq -\mathrm{lt}(g)$.
- $\deg(fg) \leq \deg(f) + \deg(g)$ with equality iff $\mathrm{lt}(f) \cdot \mathrm{lt}(g) \neq 0$.

Corollary: D is an integral domain iff $D[x]$ is an integral domain, in which case $(D[x])^\times = D^\times$.

The $\Longleftarrow$ direction is obvious. The other direction is because $\deg(fg) = \deg(f) + \deg(g)$ in $D[x]$.

If A is a ring then $a \in A$ is called nilpotent if $a^n = 0$ for some positive integer n. The set of all nilpotent elements is $\mathrm{Nil}(A)$ (called the nilradical of A).

Lemma: If A is a commutative unital ring, then:

1. $\mathrm{Nil}(A) \lhd A$,
2. $\mathrm{Nil}(A/\mathrm{Nil}(A)) = \{0\}$,
3. $1 + \mathrm{Nil}(A) \subseteq A^\times$ (In fact even more generally $A^\times + \mathrm{Nil}(A) \subseteq A^\times$).

Proof:
To show part 1, suppose $x, y \in \mathrm{Nil}(A)$. Then we know that $x^n = 0$ and $y^m = 0$ for some positive integers n, m. Now suppose $a, b \in A$. Then for any positive integer k we have that:

$$(ax+by)^k = \sum_{i=0}^{k} \binom{k}{i} (ax)^i (by)^{k-i} = \sum_{i=0}^{k} \binom{k}{i} a^i b^{k-i} x^i y^{k-i}$$

If $k \geq m + n$, then either $i \geq n$ or $k - i \geq m$. Hence either x^i or y^{k-i} will equal 0 for all i and our sum cancels out.

To show part 2, suppose $x + \mathrm{Nil}(A) \in \mathrm{Nil}(A/\mathrm{Nil}(A))$ and let m be a positive integer such that $(x + \mathrm{Nil}(A))^m = 0$. Then we must have that $x^m \in Nil(A)$. In turn, there exists a positive integer n such that $(x^m)^n = x^{mn} = 0$. And this implies that $x \in \mathrm{Nil}(A)$.

Finally, to show part 3 note that $a \in \mathrm{Nil}(A)$ implies that $a^n = 0$ for some positive integer n. In turn, $1 = 1 - a^n = (1-a)(1 + a + \ldots + a^{n-1})$. Hence $(1-a) \in A^\times$.

Since $a \in \mathrm{Nil}(A)$ iff $-a \in \mathrm{Nil}(A)$, we can replace a with $-a$ to get that $1 + a \in A^\times$ for all $a \in \mathrm{Nil}(A)$. Therefore, $1 + \mathrm{Nil}(A) \subseteq A^\times$.

Also note that if $u \in A^\times$ and $a \in \mathrm{Nil}(A)$ then $u + a = u(1 + u^{-1}a)$. Since $\mathrm{Nil}(A)$ is an ideal, $u^{-1}a \in \mathrm{Nil}(A)$. Hence we know from before that $1 + u^{-1}a \in A^\times$. And finally since $A^\times$ is a group, $u(1 + u^{-1}a) \in A^\times$. ■

Note on etymology:
If some structure is called a "___-radical" then it should satisfy the property that after quotienting it out we are left with a set with a trivial "___-radical".

If A is a unital commutative ring, we say $a \mid b$ in A if $b = ax$ for some $x \in A$. Notably, recall from math 100b that $a \mid b \iff b \in \langle a \rangle \iff \langle b \rangle \subseteq \langle a \rangle$. Also recall that we say an element $p \in A$ is prime if $p \mid ab \implies p \mid a$ or $p \mid b$. This motivates the following definitions (which are well-defined even for noncommutative rngs although in this class we'll typically focus only on commutative unital rings):

- If $\mathfrak{a}, \mathfrak{b} \lhd A$ then we say $\mathfrak{a} \mid \mathfrak{b}$ if $\mathfrak{b} \subseteq \mathfrak{a}$.
- We say $\mathfrak{p}$ is a prime ideal of A if $\mathfrak{p} \neq A$ and $\mathfrak{p} \mid \mathfrak{a}\mathfrak{b} \implies \mathfrak{p} \mid \mathfrak{a}$ or $\mathfrak{p} \mid \mathfrak{b}$ for any ideals $\mathfrak{a}, \mathfrak{b} \lhd A$.

  Here, $\mathfrak{a}\mathfrak{b} := \{\sum_{i=1}^{n} a_i b_i : n \in \mathbb{N}$ and $a_i \in \mathfrak{a}, b_i \in \mathfrak{b}$ for all $i\}$ is the ideal generated by ab as a ranges over $\mathfrak{a}$ and b ranges over $\mathfrak{b}$.

Question? Can we actually prove $\mathfrak{ab}$ is the ideal we claimed before?

To start off, it's clear that if $\mathfrak{c}$ is any ideal containing $\{ab : a \in \mathfrak{a}, b \in \mathfrak{b}\}$, then $\mathfrak{ab} \subseteq \mathfrak{c}$. After all, all finite sums of the ab must be in $\mathfrak{c}$. Hence, all we need to do now is show that $\mathfrak{ab}$ is itself an ideal.

Now it's clear that $\mathfrak{ab}$ is closed under addition. Also, note that for any $\sum_{i=1}^{n} a_i b_i \in \mathfrak{ab}$, because $\mathfrak{a}$ is a subgroup under addition we know that $-a_i \in \mathfrak{a}$ for each i. Hence $\sum_{i=1}^{n}(-a_i)b_i \in \mathfrak{ab}$ and:

$$\sum_{i=1}^{n} a_i b_i + \sum_{i=1}^{n}(-a_i)b_i = \sum_{i=1}^{n}(a_i + -a_i)b_i = \sum_{i=1}^{n} 0 = 0.$$

This proves $\mathfrak{ab}$ is a subgroup of A with respect to addition. Meanwhile, suppose $\sum_{i=1}^{n} a_i b_i \in \mathfrak{ab}$ and $x \in A$. Then as $\mathfrak{a}$ and $\mathfrak{b}$ are ideals, we have that xa_i and $b_i x$ are in $\mathfrak{a}$ and $\mathfrak{b}$ respectively for all i. Hence:

$$x \cdot \sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n}(xa_i)b_i \in \mathfrak{ab} \text{ and } (\sum_{i=1}^{n} a_i b_i) \cdot x = \sum_{i=1}^{n} a_i(b_i x) \in \mathfrak{ab}.$$

I realize I keep trying to generalize things to noncommutative or non-unital rings. That said, the professor noted in lecture that algebraic approaches are typically not helpful when working with noncommutative rings and that the better approach in that scenario is to study operators (for example math 241a stuff).

Í'll also mention that although the definitions I made before do work for noncommutative rings, ideals are much more messy to work with when you don't assume your rings are commutative. For example (and this holds regardless of if A is unital):

- If A is a commutative ring, then $\langle x_1, \ldots, x_n \rangle = x_1 A + \ldots + x_n A$.
- If A is a non-commutative ring, then:

$$\langle x_1, \ldots, x_n \rangle = \{\textstyle\sum_{i=1}^{m} a_i : m \in \mathbb{N} \text{ and } a_i \in Ax_1A + \ldots + Ax_nA \text{ for all } i\}$$

Note that if $\mathfrak{p}$ is a prime ideal, then $\mathfrak{p} \mid \mathfrak{ab}$ if and only if $\mathfrak{ab} \subseteq \mathfrak{p}$. Also, the latter only happens if for all $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$, $ab \in \mathfrak{p}$.

<u>Lemma:</u> If A is a commutative ring, then $\mathfrak{p}$ is a prime ideal iff $\mathfrak{p} \neq A$ and $ab \in \mathfrak{p}$ implies that $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. (I.e. the definition in this class generalizes the one from math 100b)

$(\Longrightarrow)$

Let $\mathfrak{a} = \langle a \rangle$ and $\mathfrak{b} = \langle b \rangle$. Then $\mathfrak{ab} = \langle ab \rangle$.

This is because $\mathfrak{ab}$ is generated by elements of the form $(ax)(by) = ab(xy)$.

Now if $ab \in \mathfrak{p}$ we know that $\langle ab \rangle \subseteq \mathfrak{p}$. Hence, $\mathfrak{p} \mid \langle ab \rangle = \mathfrak{ab}$, and by definition this means that either $\mathfrak{p} \mid \mathfrak{a} = \langle a \rangle$ or $\mathfrak{p} \mid \mathfrak{b} = \langle b \rangle$. In turn we know either $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$.

$(\Longleftarrow)$

Suppose $\mathfrak{p}$ is a proper ideal of A that is not prime. Then there exists $\mathfrak{a}, \mathfrak{b} \lhd A$ such that $\mathfrak{p} \mid \mathfrak{ab}$, $\mathfrak{p} \nmid \mathfrak{a}$ and $\mathfrak{p} \nmid \mathfrak{b}$. In other words, we know that $\mathfrak{ab} \subseteq \mathfrak{p}$, $\mathfrak{a} \not\subseteq \mathfrak{p}$, and $\mathfrak{b} \not\subseteq \mathfrak{p}$. So, there exists $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$ such that $a, b \notin \mathfrak{p}$ but $ab \in \mathfrak{p}$. This contradicts the right side assumption.

Lemma: Let A be a unital commutative ring. Then $\mathfrak{p} \lhd A$ is prime if and only if $A/\mathfrak{p}$ is an integral domain.

$(\Longrightarrow)$

- $A/\mathfrak{p}$ is ring with more than one element as well as a multiplicative identity $1+\mathfrak{p}$. As a result, we know $A/\mathfrak{p}$ is a unital ring where $0 \neq 1$. Also, A being commutative means that $A/\mathfrak{p}$ is commutative.

  As a side note: if A is a unital ring (not necessarily commutative) then $0 = 1$ in A iff $A = \{0\}$. After all, we know that $x = x \cdot 1 = x \cdot 0 = 0$ for any $x \in A$.

- $A/\mathfrak{p}$ has no zero divisors. After all, suppose $\overline{a} \cdot \overline{b} = \overline{0}$ in $A/\mathfrak{p}$ (where $\overline{x} = x + \mathfrak{p}$). Then $ab \in \mathfrak{p}$, and by our prior lemma we know either $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. So, either $\overline{a} = 0$ or $\overline{b} = 0$.

$(\Longleftarrow)$

The fact that $\overline{0} \neq \overline{1}$ in $A/\mathfrak{p}$ means that $\mathfrak{p} \neq A$. Also, note for any $a, b \in A$ that if $ab \in \mathfrak{p}$ then $\overline{a}\overline{b} = 0$. Hence if $A/\mathfrak{p}$ has no zero divisors we must have that $\overline{a} = \overline{0}$ or $\overline{b} = \overline{0}$. Or in other words, $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. This proves by the last lemma that $\mathfrak{p}$ is prime.

Given a ring A, we let $\mathrm{Ideal}(A)$ denote the collection of all ideals of A. Furthermore, we let $\mathrm{Spec}(A)$ denote the collection of all prime ideals of A.

Ideal Correspondance Theorem: Let $\mathfrak{a}$ be an ideal of a ring A. Then there is a bijective correspondance between $\{\mathfrak{b} \in \mathrm{Ideal}(A) : \mathfrak{a} \subseteq \mathfrak{b}\}$ and $\mathrm{Ideal}(A/\mathfrak{a})$ given by $\mathfrak{b} \mapsto \mathfrak{b}/\mathfrak{a}$.

Proof:

We already know from math 100a that there is a correspondance between subgroups of $(A/\mathfrak{a}, +)$ and subgroups of $(A, +)$ containing $\mathfrak{a}$ which is given by our proposed bijection. So, we just need to show that our proposed bijection (call it π) satisfies that $\mathfrak{b}$ is an ideal iff $\pi(\mathfrak{b})$ is an ideal.

In my math 100b notes, I showed the $\Longrightarrow$ direction. As for the other direction (which I forgot to show in 100b oops), suppose $\mathfrak{b}/\mathfrak{a}$ is an ideal. Then we know that $xy + \mathfrak{a} = (x+\mathfrak{a})(y+\mathfrak{a}) \in \mathfrak{b}/\mathfrak{a}$ for all $x \in A$ and $y \in \mathfrak{b}$. In particular, this means $xy \in b + \mathfrak{a}$ for some $b \in \mathfrak{b}$. And since $\mathfrak{a} \subseteq \mathfrak{b}$, this proves $xy \in \mathfrak{b}$.

(Analogous reasoning shows that $yx \in \mathfrak{b}$ for all $x \in A$ and $y \in \mathfrak{b}$.) $\blacksquare$

An ideal $\mathfrak{m} \lhd A$ is called maximal if $\mathfrak{m} \neq A$ and for any $\mathfrak{b} \lhd A$ such that $\mathfrak{m} \subseteq \mathfrak{b}$ we have that either $\mathfrak{m} = \mathfrak{b}$ or $A = \mathfrak{b}$. We denote the set of all maximal ideals of A as $\mathrm{Max}(A)$.

Recall from math 100b that a nontrivial commutative unital ring is a field if and only if it satisfies that $\mathrm{Ideal}(F) = \{\{0\}, F\}$. As a consequence of this fact and the ideal correspondance theorem, we get the following lemma:

If A is a nontrivial unital commutative ring then $\mathfrak{m} \lhd A$ is maximal if and only if $A/\mathfrak{m}$ is a field.

Corollary: If A is a unital commutative ring, then $\mathrm{Max}(A) \subseteq \mathrm{Spec}(A)$.

Proof:

$\mathfrak{m} \in \mathrm{Max}(A) \Longrightarrow A/\mathfrak{m}$ is a field $\Longrightarrow A/\mathfrak{m}$ is an integral domain $\Longrightarrow \mathfrak{m} \in \mathrm{Spec}(A)$. ■

Ring Theory Break: (the following problem is from Hagan)

Let G be finite group of odd order and suppose $N \lhd G$ with $|N| = 5$. Then $N \subseteq Z(G)$.

Proof:

Consider the action $G \curvearrowright N$ by conjugation. This gives us an induced group homomorphism $\Phi : G \to \mathrm{Aut}(N)$. Furthermore, because $N \cong C_5$, we know that $\mathrm{Aut}(N) \cong C_4$. Therefore, we must have that $\mathrm{im}(\Phi) \in \{1, 2, 4\}$.

That said, we know $|\mathrm{im}(\Phi)|$ divides $|G|$ and $|G|$ is odd. The only way this is possible is $|\mathrm{im}(\Phi)|$ is odd. Hence we must have that $\mathrm{im}(\Phi) = \{\mathrm{Id}\}$. In other words, $gng^{-1} = n$ for all $n \in N$ and $g \in G$. This is the same as saying $N \subseteq Z(G)$.

12/2/2025

More Math 200a notes:

To start off, here are two hopefully obvious statements. Let A be a unital commutative ring. Then:

- A is a domain iff $\{0\}$ is a prime ideal (i.e. $\{0\} \in \mathrm{Spec}(A)$)
- A is a field iff $\mathrm{Max}(A) = \{\{0\}\}$ and $0 \neq 1$.

Also here are some other facts:

Lemma: Suppose D is an integral domain and $|D| < \infty$. Then D is a field.

Proof:

If D is finite, then for any $x \in D - \{0\}$ there must exists positive integers $m > n$ such $x^m = x^n$. In turn, we can conclude that $x^{m-n} = 1$ where $m - n$ is a positive integer. So, x^{-1} exists and equals x^{m-n-1}. ■

As a side note, to see why $x^m = x^n \Longrightarrow x^{m-n} = 1$, note that if $a, b, c \in D$ where $c \neq 0$ and D is an integral domain then $ac = bc \Longrightarrow a = b$. This is because $ac = bc \Longrightarrow (a - b)c = 0$. Then, as D is an integral domain and $c \neq 0$, we must have that $a - b = 0$. Or in other words, $a = b$.

(This cancellation property is precisely why we study integral domains).

Corollary: If A is a unital commutative ring and $\mathfrak{p} \in \mathrm{Spec}(A)$ with $|A/\mathfrak{p}| < \infty$ then $\mathfrak{p} \in \mathrm{Max}(A)$.

Proof:

As $\mathfrak{p} \in \mathrm{Spec}(A)$ we know that $A/\mathfrak{p}$ is an integral domain. Then combining that with the fact that $|A/\mathfrak{p}| < \infty$, we know that $A/\mathfrak{p}$ is a field. So, $\mathfrak{p} \in \mathrm{Max}(A)$. ■

Lemma: Suppose $\mathfrak{C}$ is a chain of ideals (simply-ordered by inclusion) in a ring A. Then $\bigcup_{\mathfrak{a}\in\mathfrak{C}} \mathfrak{a} \lhd A$.

Note: A chain is just a simply-ordered subset of a poset (partially ordered set).

Proof:
Suppose $x, y \in \bigcup_{\mathfrak{a}\in\mathfrak{C}} \mathfrak{a}$ and $a_1, a_2 \in A$. Then we know that $x \in \mathfrak{a}_1$ and $y \in \mathfrak{a}_2$ for some $\mathfrak{a}_1, \mathfrak{a}_2 \in \mathfrak{C}$. But, because $\mathfrak{C}$ is a chain, we must have that either $\mathfrak{a}_1 \subseteq \mathfrak{a}_2$ or $\mathfrak{a}_2 \subseteq \mathfrak{a}_1$. Hence we can actually conclude that x, y are in a single ideal $\mathfrak{a}$ in $\mathfrak{C}$. And now it's obvious that $a_1x + a_2y \in \mathfrak{a}$. $\blacksquare$

Here are two quick definitions:

- If A is a unital commutative ring, we say a subset $S \subseteq A$ is multiplicatively closed if $1 \in S$ and $s_1, s_2 \in S \Longrightarrow s_1 s_2 \in S$.
- If A is a ring and $\mathfrak{a} \lhd A$, we denote the set of prime divisors of $\mathfrak{a}$:
$$V(\mathfrak{a}) := \{\mathfrak{p} \in \mathrm{Spec}(A) : \mathfrak{p} \mid \mathfrak{a}\} = \{\mathfrak{p} \in \mathrm{Spec}(A) : \mathfrak{a} \subseteq \mathfrak{p}\}.$$

Tangent:
Let A be an arbitrary ring.

Suppose $\{\mathfrak{a}_i\}_{i\in I}$ is a collection of ideals in a ring A. Then we define their sum to be:
$$\sum_{i\in I} \mathfrak{a}_i := \{a_1 + \ldots + a_n : n \in \mathbb{Z}^+ \text{ and } \exists i_1, \ldots, i_n \in I \ s.t.\ a_k \in \mathfrak{a}_{i_k} \text{ for each } k\}$$

Note that $\sum_{i\in I} \mathfrak{a}_i \lhd A$ and $\langle \bigcup_{i\in I} \mathfrak{a}_i \rangle = \sum_{i\in I} \mathfrak{a}_i$.

To see this, first note that it's obvious $\sum_{i\in I} \mathfrak{a}_i \subseteq \langle \bigcup_{i\in I} \mathfrak{a}_i \rangle$ since the latter must contain all sums of elements in the $\mathfrak{a}_i$. Hence, we will be done if we can show that $\sum_{i\in I} \mathfrak{a}_i$ is an ideal. Fortunately:

- As each $\mathfrak{a}_i$ is a subgroup of A under addition, it is easy to see that $\sum_{i\in I} \mathfrak{a}_i$ is a subgroup of A under addition.
- Suppose $x \in A$ and $a_1 + \ldots + a_n \in \sum_{i\in I} \mathfrak{a}_i$. Then $xa_k, a_kx \in \mathfrak{a}_{i_k}$ for each k since $\mathfrak{a}_{i_k}$ is an ideal. Hence, $x(a_1 + \ldots + a_n) \in \sum_{i\in I} \mathfrak{a}_i$ and $(a_1 + \ldots + a_n)x \in \sum_{i\in I} \mathfrak{a}_i$.

Furthermore, if $\mathfrak{p} \in \mathrm{Spec}(A)$ then $\mathfrak{p} \mid \mathfrak{a}_i$ for all $i \in I$ if and only if $\mathfrak{p} \mid \sum_{i\in I} \mathfrak{a}_i$.

This is because $\mathfrak{p} \mid \mathfrak{a}_i$ for all $i \in I$ if and only if $\bigcup_{i\in I} \mathfrak{a}_i \subseteq \mathfrak{p}$. But then since $\sum_{i\in I} \mathfrak{a}_i$ is the smallest ideal containing $\bigcup_{i\in I} \mathfrak{a}_i$, we have that $\bigcup_{i\in I} \mathfrak{a}_i \subseteq \mathfrak{p}$ if and only if $\sum_{i\in I} \mathfrak{a}_i \subseteq \mathfrak{p}$.

Meanwhile, recall from the *bottom of page 451* how we defined the product of two ideals. By the definition of a prime ideal we know that $V(\mathfrak{a}\mathfrak{b}) \subseteq V(\mathfrak{a}) \cup V(\mathfrak{b})$. On the other hand, note that $\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{a} \cap \mathfrak{b}$.

This is because:

- $\mathfrak{a}\mathfrak{b}$ is the ideal generated by the ab as a ranges over $\mathfrak{a}$ and b ranges over $\mathfrak{b}$,
- $ab \in \mathfrak{a} \cap \mathfrak{b}$ for all $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$.

It therefore follows that $V(\mathfrak{a}) \cup V(\mathfrak{b}) \subseteq V(\mathfrak{a}\mathfrak{b})$.

Interesting side note: we now know $V(\mathfrak{a}\mathfrak{b}) = V(\mathfrak{a}) \cup V(\mathfrak{b}) = V(\mathfrak{b}) \cup V(\mathfrak{a}) = V(\mathfrak{b}\mathfrak{a})$.

Now, we define a topology on $\mathrm{Spec}(A)$ by considering $\{V(\mathfrak{a}) : \mathfrak{a} \lhd A\}$ to be the collection of closed subsets of $\mathrm{Spec}(A)$.

Proof that this defines a topology:
It suffices to show that $\{V(\mathfrak{a}) : \mathfrak{a} \lhd A\}$ is closed under arbitrary intersections and finite unions, and that $\emptyset, \mathrm{Spec}(A) \in \{V(\mathfrak{a}) : \mathfrak{a} \lhd A\}$. That way the collection of complements of the $V(\mathfrak{a})$ satisfy the topology axioms. Fortunately:

- Note that $\emptyset = V(A)$ and $\mathrm{Spec}(A) = V(\{0\})$.
- Suppose $\{\mathfrak{a}_i\}_{i \in I}$ is a collection of ideals in A. Then we know from before that $\bigcap_{i \in I} V(\mathfrak{a}_i) = V(\sum_{i \in I} \mathfrak{a}_i)$.
- Let $\mathfrak{a}, \mathfrak{b}$ be ideals of A. Then $V(\mathfrak{a}) \cup V(\mathfrak{b}) = V(\mathfrak{a}\mathfrak{b})$.

This topology on $\mathrm{Spec}(A)$ is called the Zariski topology and it is studied heavily in the field of algebraic geometry.

Theorem: Let A be a unital commutative ring. Suppose $\mathfrak{a} \lhd A$, $S \subseteq A$ is nonempty and multiplicatively closed, and $S \cap \mathfrak{a} = \emptyset$. Then:

1. there exists $\mathfrak{p} \in \mathrm{Spec}(A)$ with $\mathfrak{p} \mid \mathfrak{a}$ such that $\mathfrak{p} \cap S = \emptyset$ (i.e. there exists $\mathfrak{p} \in V(\mathfrak{a})$ such that $\mathfrak{p} \cap S = \emptyset$);
2. $\mathrm{Max}(A) \cap V(\mathfrak{a}) \neq \emptyset$.

Proof:
Let $\Sigma_{\mathfrak{a},S} := \{\mathfrak{b} \lhd A : \mathfrak{b} \mid \mathfrak{a} \text{ and } \mathfrak{b} \cap S = \emptyset\}$. Then note that $\Sigma_{\mathfrak{a},S} \neq \emptyset$ since $\mathfrak{a} \in \Sigma_{\mathfrak{a},S}$. We shall prove the theorem by showing the following three claims. Hopefully it is obvious that the first two claims prove part 1 and the third claim proves part 2.

Claim 1: $\Sigma_{\mathfrak{a},S}$ when partially ordered by inclusion has a maximal element.
By Zorn's lemma it suffices to show that every chain $\mathcal{C}$ in $\Sigma_{\mathfrak{a},S}$ has an upper bound. So, let $\mathfrak{b} = \bigcup_{\mathfrak{a}' \in \mathcal{C}} \mathfrak{a}'$. By the lemma at the top of the last page, we know that $\mathfrak{b} \lhd A$. Also $\mathfrak{a} \mid \mathfrak{b}$ since $\mathfrak{a} \subseteq \mathfrak{a}'$ for all $\mathfrak{a}' \in \mathcal{C}$. Finally, note that $\mathfrak{b} \cap S = \emptyset$ since $\mathfrak{a}' \cap S$ for all $\mathfrak{a}' \in \mathcal{C}$. Therefore, $\mathfrak{b} \in \Sigma_{\mathfrak{a},S}$ and is an upper bound to $\mathcal{C}$.

Claim 2: If $\mathfrak{p}$ is maximal in $\Sigma_{\mathfrak{a},S}$ then $\mathfrak{p} \in \mathrm{Spec}(A)$.
Suppose for the sake of contradiction that $\mathfrak{p}$ is maximal in $\Sigma_{\mathfrak{a},S}$ but not prime. Then there exists $a, b \in A$ with $ab \in \mathfrak{p}$ but $a \notin \mathfrak{p}$ and $b \notin \mathfrak{p}$. In turn, if we consider the ideals $\mathfrak{p} + \langle a \rangle$ and $\mathfrak{p} + \langle b \rangle$ we know that $\mathfrak{p} \subsetneq \mathfrak{p} + \langle a \rangle$ and $\mathfrak{p} \subsetneq \mathfrak{p} + \langle b \rangle$. Hence, by the maximality of $\mathfrak{p}$ in $\Sigma_{\mathfrak{a},S}$ we have that $S \cap (\mathfrak{p} + \langle a \rangle) \neq \emptyset$ and $S \cap (\mathfrak{p} + \langle b \rangle) \neq \emptyset$.

Now pick $s_1 \in S \cap (\mathfrak{p} + \langle a \rangle)$ and $s_2 \in S \cap (\mathfrak{p} + \langle b \rangle)$. Since S is multiplicatively closed we know that $s_1 s_2 \in S$. Also, we know there exists $x_1, x_2 \in A$ and $p_1, p_2 \in \mathfrak{p}$ such that $s_1 = ax_1 + p_1$ and $s_2 = bx_2 + p_2$. But now:

$$(ax_1 + p_1)(bx_2 + p_2) = abx_1x_2 + ax_1p_2 + bx_2p_1 + p_1p_2 \in S.$$

This is a contradiction because $ab \in \mathfrak{p}$. Hence, all four terms being summed are in $\mathfrak{p}$.

Claim 3: If $\mathfrak{m}$ is maximal in $\Sigma_{\mathfrak{a},\{1\}}$ then $\mathfrak{m} \in \mathrm{Max}(A)$.

Suppose $\mathfrak{m} \subseteq \mathfrak{m}'$ where $\mathfrak{m}'$ is a proper ideal of A. Then we know that $\mathfrak{m}' \mid \mathfrak{a}$ since $\mathfrak{m} \mid \mathfrak{a}$ and $\mathfrak{m}' \mid \mathfrak{m}$. Also note that as $\mathfrak{m}'$ is a proper ideal of A we know that $1 \notin \mathfrak{m}'$. But now we've shown that $\mathfrak{m}' \in \Sigma_{\mathfrak{a},\{1\}}$. Hence, as $\mathfrak{m}$ is maximal in $\Sigma_{\mathfrak{a},\{1\}}$ we must have that $\mathfrak{m} = \mathfrak{m}'$. This proves $\mathfrak{m} \in \mathrm{Max}(A)$. ■

Recalling the Zariski topology, note that if $\mathfrak{p} \in \mathrm{Spec}(A)$ then $\{\mathfrak{p}\}$ is closed if and only if $\mathfrak{p} \in \mathrm{Max}(A)$.

$(\Longleftarrow)$
If $\mathfrak{p} \in \mathrm{Max}(A)$ then $V(\mathfrak{p}) = \{\mathfrak{p}\}$.

$(\Longrightarrow)$
If $\mathfrak{p} \notin \mathrm{Max}(A)$ then we know there exists $\mathfrak{m} \in \mathrm{Max}(A)$ such that $\mathfrak{m} \in V(\mathfrak{p})$. In other words, $\mathfrak{p} \subseteq \mathfrak{m}$. Now if $V(\mathfrak{a})$ is any closed set containing $\mathfrak{p}$ then it must also contain $\mathfrak{m}$ as $\mathfrak{a} \subseteq \mathfrak{p} \subseteq \mathfrak{m} \Longrightarrow \mathfrak{a} \subseteq \mathfrak{m}$. Hence, $\{\mathfrak{p}\}$ is not closed.

Notably, this means that unless $\mathrm{Spec}(A) = \mathrm{Max}(A)$ we have that the Zariski topology is not T_1. Although, according to a Google search it is T_0. So, I guess this is the reason why we separately define T_0 versus T_1 spaces in a topology class.

Here are some useful sets to take S to be:

- Note that if A is a unital commutative ring and $\mathfrak{p} \in \mathrm{Spec}(A)$ then $S_\mathfrak{p} := A - \mathfrak{p}$ (where the minus sign denotes a set difference) is a multiplicatively closed set.

  Why?
  Since $1 \notin \mathfrak{p}$ we know $1 \in S_\mathfrak{p}$. Also, if $s_1, s_2 \in S_\mathfrak{p}$ then we know $s_1, s_2 \notin \mathfrak{p}$. In turn, because $\mathfrak{p} \in \mathrm{Spec}(A)$ we can't have that $s_1 s_2 \in \mathfrak{p}$. Hence, $s_1 s_2 \in S_\mathfrak{p}$.

- Given any $a \in A$ let $S_a := \{1, a, a^2, \ldots\}$. This is clearly a multiplicatively closed set. Also, the zero ideal intercepts S_a iff a is nilpotent. As a consequence, we get the following theorem:

  Theorem: If A is a unital commutative ring then $\mathrm{Nil}(A) = \bigcap_{\mathfrak{p} \in \mathrm{Spec}(A)} \mathfrak{p}$

  Proof:
  If $x \in \mathrm{Nil}(A)$ then $x^n = 0$ for some positive integer n. But also note that for all $\mathfrak{p} \in \mathrm{Spec}(A)$ we have that $A/\mathfrak{p}$ is an integral domain and thus $\mathrm{Nil}(A/\mathfrak{p}) = \{0 + \mathfrak{p}\}$. Since $(x + \mathfrak{p})^n = 0$ we must have that $x + \mathfrak{p} = 0 + \mathfrak{p}$. So, $x \in \mathfrak{p}$. This proves that $\mathrm{Nil}(A) \subseteq \bigcap_{\mathfrak{p} \in \mathrm{Spec}(A)} \mathfrak{p}$.

  Meanwhile, suppose $x \notin \mathrm{Nil}(A)$ and consider $S_x = \{1, x, x^2, \ldots\}$. Then $\{0\}$ is an ideal of A not intercepting S_x. Hence, by our last theorem we know there exists $\mathfrak{p} \in \mathrm{Spec}(A)$ such that $\mathfrak{p} \cap S_x = \emptyset$. In particular, $x \notin \mathfrak{p}$. Hence, $x \notin \bigcap_{\mathfrak{p} \in \mathrm{Spec}(A)} \mathfrak{p}$. ■

For the next section (until I say otherwise), it will be assumed that A is an integral domain.

Recall the following definitions:

- We say $a \in A$ is irreducible if $a \notin A^\times \cup \{0\}$ and $a = bc \Longrightarrow b \in A^\times$ or $c \in A^\times$.
- We say $a \in A$ is prime if $a \notin A^\times \cup \{0\}$ and $a \mid bc \Longrightarrow a \mid b$ or $a \mid c$.
- A is a Unique Factorization Domain (U.F.D.) if for all $a \in A$ then either $a \in D^\times \cup \{0\}$ or $a = p_1 \dots p_m$ where all the p_i are irreducible and the factorization is "essentially unique".

    By "essentially unique" the professor means that if we also have that $a = q_1 \dots q_n$ then $n = m$ and after some reordering of the q_i we have that $q_i = u_i p_i$ for all i where each $u_i \in A^\times$.

- $a, b \in A$ are companion to each other (written $a \sim b$) iff $b = ua$ for some $u \in A^\times$.

Lemma: $a \sim b$ if and only if $\langle a \rangle = \langle b \rangle$.

($\Longrightarrow$)
Since $b = ua$ is a multiple of a, we have that $\langle b \rangle \subseteq \langle a \rangle$. But also note that $bu^{-1} = a$. So by the same reasoning we know $\langle b \rangle \supseteq \langle a \rangle$.

($\Longleftarrow$)
This implies $b = ax$ and $a = by$ for some $x, y \in A$. Hence, $b = ax = b(yx)$. If $b = 0$ then $a = 0$ and trivially $a \sim b$. Meanwhile, if $b \neq 0$ then $1 = yx$. it follows $y, x \in A^\times$ So, $a \sim b$. ■

Corollary: $\sim$ is an equivalence relation.

Lemma: $a \in A$ is prime iff $\langle a \rangle \neq 0$, $\langle a \rangle \neq A$, and $\langle a \rangle$ is a prime ideal.

Proof:
$a \neq 0$ iff $\langle a \rangle \neq \{0\}$ and $a \notin A^\times$ iff $\langle a \rangle \neq A$. As for the other property we need to show:

($\Longrightarrow$)
Suppose $bc \in \langle a \rangle$. Then $a \mid bc$. So if a is prime we know either $a \mid b$ or $a \mid c$, and that implies that either $b \in \langle a \rangle$ or $c \in \langle a \rangle$.

($\Longleftarrow$)
Suppose $a \mid bc$. Then $bc \in \langle a \rangle$. So, as $\langle a \rangle$ is prime either $b \in \langle a \rangle$ or $c \in \langle a \rangle$. And that happens iff either $a \mid b$ or $a \mid c$. ■

Lemma: $a \in A$ is irreducible if and only if $\langle a \rangle \neq \{0\}$, $\langle a \rangle \neq A$, and $\langle a \rangle$ is maximal among principal proper ideals.

Note, a principal ideal is just an ideal generated by a single element of a ring.

($\Longrightarrow$)
Suppose $\langle a \rangle \subseteq \langle a' \rangle \subsetneq A$. Then as $a \in \langle a' \rangle$, we know that $a = a'x$ for some $x \in A$. Next, since a is irreducible we know that either $a' \in A^\times$ or $x \in A^\times$. But we can't have that $a' \in A^\times$ since that would imply $\langle a' \rangle = A$. So, we have that $x \in A^\times$ and thus $a \sim a'$.

($\Longleftarrow$)
Since $\langle a\rangle \neq \{0\}$ and $\langle a\rangle \neq A$, we know that $a \notin A^\times \cup \{0\}$. Now suppose $a = bc$. Then we know that $\langle a\rangle \subseteq \langle b\rangle$. So by the maximality of a we have that either $\langle b\rangle = \langle a\rangle$ or $\langle b\rangle = A$. In the latter case, we know b is a unit. As for the former case, we thus know that $bc = a = bu$ for some $u \in A^\times$. By cancellation we thus get that $c = u$. So, c is a unit. ■

Here are a few more definitions:

- A is a Principal Ideal Domain (P.I.D.) if every ideal in A is principal.
- A is a Euclidean Domain (E.D.) if there exists a function $N : A \to \mathbb{Z}^{\geq 0}$ such that $N(0) = 0$ and for all $a, b \in A$ with $b \neq 0$ there exists $q, r \in A$ such that $a = bq + r$ and $N(r) < N(b)$.

  Here are some examples of Euclidean domains:
  - $\mathbb{Z}$ equipped with $N(a) := |a|$
  - $F[x]$ (where F is a field) equipped with $N(f(x)) := 2^{\deg(f)}$ (we use the convention $2^{-\infty} = 0$).

Theorem: If A is an E.D. then we know A is a P.I.D.

Proof:
Suppose $\mathfrak{a} \lhd D$. Without loss of generality we can assume $\mathfrak{a} \neq \{0\}$ (since $\{0\}$ is a principal ideal). Therefore, we can choose $a_0 \in \mathfrak{a} - \{0\}$ with the property that:

$$N(a_0) = \min\{N(a) : a \in \mathfrak{a} - \{0\}\}.$$

Now since $a_0 \in \mathfrak{a}$ we trivially have that $\langle a_0\rangle \subseteq \mathfrak{a}$. We also claim that $\mathfrak{a} \subseteq \langle a_0\rangle$. After all, suppose $a \in \mathfrak{a}$. Then there exists $q, r \in A$ such that $a = a_0q + r$ and $N(r) < N(a_0)$. But also note that $r = a - a_0q \in \mathfrak{a}$. Hence, we must have that $r = 0$ or else we'd be forced to have that $N(r) \geq N(a_0)$ because of how we chose a_0. With that, we've shown for any $a \in \mathfrak{a}$ that there exists $q \in A$ such that $a = a_0q$. This proves that $\mathfrak{a} \subseteq \langle a_0\rangle$. ■

With that I hope I'm prepared to cram the homework tomorrow.

12/3/2025

Math 200a Homework:

Set 9 Problem 1: Suppose A is a nonzero unital commutative ring.

(a) Suppose $\mathfrak{a}$ is an ideal of A and let $\pi_\mathfrak{a} : A \to A/\mathfrak{a}$ be such that $\pi_\mathfrak{a}(x) := x + \mathfrak{a}$. Then $\pi_\mathfrak{a} : A[x] \to (A/\mathfrak{a})[x]$ given by:

$$\pi_\mathfrak{a}(\textstyle\sum_{i=0}^n a_ix^i) = \sum_{i=0}^n \pi_\mathfrak{a}(a_i)x^i$$

is a surjective ring homomorphism and its kernel is $\mathfrak{a}[x] := \{\sum_{i=0}^n a_ix^i : n \in \mathbb{Z}^+, a_i \in \mathfrak{a}\}$.

To prove that we did make a well-defined ring homomorphism, I'm going to show something slightly more general. Suppose A, A' are arbitrary rings and $\theta : A \to A'$ is a ring homomorphism. Then we can extend θ to a ring homomorphism from $A[x]$ to $A'[x]$ by setting $\theta(\sum_{i=0}^n a_ix^i) = \sum_{i=0}^n \theta(a_i)x^i$.

Suppose $a_m x^m + \ldots + a_0$ and $b_n x^n + \ldots + b_0$ are polynomials in $A[x]$ and without loss of generality suppose $m = n$ (we can do this by just padding one of the polynomial with zero terms).

- $$\begin{aligned}\theta(\textstyle\sum_{i=0}^n a_i x^i + \sum_{i=0}^n b_i x^i) &= \textstyle\sum_{i=0}^n \theta(a_i + b_i)x^i \\ &= \textstyle\sum_{i=0}^n \theta(a_i)x^i + \sum_{i=1}^n \theta(b_i)x^i \\ &= \theta(\textstyle\sum_{i=0}^n a_i x^i) + \theta(\sum_{i=0}^n b_i x^i).\end{aligned}$$

- $$\begin{aligned}\theta((\textstyle\sum_{i=0}^n a_i x^i)(\sum_{i=0}^n b_i x^i)) &= \theta(\textstyle\sum_{i=0}^{2n}(\sum_{j+k=i} a_j b_k)x^i) \\ &= \textstyle\sum_{i=0}^{2n} \theta(\sum_{j+k=i} a_j b_k)x^i \\ &= \textstyle\sum_{i=0}^{2n} \sum_{j+k=i} \theta(a_j)\theta(b_k)x^i \\ &= (\textstyle\sum_{i=0}^n \theta(a_i)x^i)(\sum_{i=0}^n \theta(b_i)x^i) \\ &= \theta(\textstyle\sum_{i=0}^n a_i x^i)\theta(\sum_{i=0}^n b_i x^i).\end{aligned}$$

Also if $\theta : A \to A'$ is unital, then so is θ's extension to $A[x]$.

$\pi_{\mathfrak{a}}$ is surjective because it's restriction to just the constant polynomials is surjective. Also $\sum_{i=0}^n a_i x^i \in \ker(\pi_{\mathfrak{a}})$ iff $\pi_{\mathfrak{a}}(a_i) = 0 + \mathfrak{a}$ for all i. Yet this happens precisely when $a_i \in \mathfrak{a}$ for each i.

(b) Prove that if $\mathfrak{p} \in \mathrm{Spec}(A)$ then $\mathfrak{p}[x] \in \mathrm{Spec}(A[x])$.

We know $\mathfrak{p}[x]$ is an ideal of $A[x]$ since it is the kernel of the ring homomorphism $\pi_{\mathfrak{p}} : A[x] \to (A/\mathfrak{p})[x]$ defined in part (a). Meanwhile, to show that $\mathfrak{p}[x]$ is prime it suffices to instead demonstrate that $A[x]/\mathfrak{p}[x]$ is an integral domain. Luckily, by the first isomorphism theorem for rings (see my math 100b notes), we know that $A[x]/\mathfrak{p}[x] = A[x]/\ker(\pi_{\mathfrak{p}}) \cong (A/\mathfrak{p})[x]$ by a unital ring isomorphism. Next note that because $\mathfrak{p} \in \mathrm{Spec}(A)$, we know that $A/\mathfrak{p}$ is an integral domain. In turn, we also have that $(A/\mathfrak{p})[x]$ is an integral domain. And since $A[x]/\mathfrak{p}[x]$ is isomorphic to an integral domain, it must itself be an integral domain.

(c) Prove that $\mathrm{Nil}(A[x]) = \mathrm{Nil}(A)[x]$.

To start off, by part (b) we know that:

$$\mathrm{Nil}(A[x]) = \bigcap_{\mathfrak{p}' \in \mathrm{Spec}(A[x])} \mathfrak{p}' \supseteq \bigcap_{\mathfrak{p} \in \mathrm{Spec}(A)} \mathfrak{p}[x]$$

Also note that a polynomial f is in $\bigcap_{\mathfrak{p} \in \mathrm{Spec}(A)} \mathfrak{p}[x]$ if and only if all the coefficients of f are in each $\mathfrak{p} \in \mathrm{Spec}(A)$. Hence, we have that:

$$\bigcap_{\mathfrak{p} \in \mathrm{Spec}(A)} \mathfrak{p}[x] = \left(\bigcap_{\mathfrak{p} \in \mathrm{Spec}(A)} \mathfrak{p}\right)[x] = \mathrm{Nil}(A)[x]$$

Thus, we've shown that $\mathrm{Nil}(A)[x] \subseteq \mathrm{Nil}(A[x])$. As for showing that this inclusion isn't proper, suppose $f \in \mathrm{Nil}(A[x])$ with $f \notin \mathrm{Nil}(A)[x]$. Then we must have $\mathrm{Nil}(\frac{A[x]}{\mathrm{Nil}(A)[x]}) \neq \{0 + \mathrm{Nil}(A)[x]\}$ since $f + \mathrm{Nil}(A)[x] \neq 0 + \mathrm{Nil}(A)[x]$ and yet $f + \mathrm{Nil}(A)[x] \in \mathrm{Nil}(\frac{A[x]}{\mathrm{Nil}(A)[x]})$.

But now by part (a) we know that $\frac{A[x]}{\mathrm{Nil}(A)[x]}$ is isomorphic to $(\frac{A}{\mathrm{Nil}(A)})[x]$. Hence, we must have that $\mathrm{Nil}((\frac{A}{\mathrm{Nil}(A)})[x])$ isn't trivial either. This is a contradiction. After all suppose $g \in (\frac{A}{\mathrm{Nil}(A)})[x]$ with $g(x) \neq (0 + \mathrm{Nil}(A))$. Then as $\mathrm{Nil}(A/\mathrm{Nil}(A))$ is trivial and the leading term of g isn't zero, we must have that $\mathrm{lt}(g(x)^k) = \mathrm{lt}(g(x))^k \neq 0$ for all positive integers k. In particular, this means that:

$$g(x)^k \neq (0 + \mathrm{Nil}(A)) \text{ for all } g(x) \in (\tfrac{A}{\mathrm{Nil}(A)})[x].$$

(d) Prove that:

$$A[x]^\times = \{a_0 + a_1 x + \ldots + a_n x^n : n \in \mathbb{Z}^+, a_0 \in A^\times, \text{ and } a_1, \ldots, a_n \in \mathrm{Nil}(A)\}.$$

Let S be the set we're claiming above that $A[x]^\times$ equals. Then by part (c) plust the fact that it's obvious that $A^\times \subseteq A[x]^\times$, we know that $S \subseteq A[x]^\times + \mathrm{Nil}(A[x]) \subseteq A[x]^\times$.

To show the other inclusion, suppose $f \in A[x]^\times$ and let $g = f^{-1}$. Then for any $\mathfrak{p} \in \mathrm{Spec}(A)$ consider the ring homomorphism $\pi_\mathfrak{p} : A[x] \to (A/\mathfrak{p})[x]$ from part (a). Since $f(x)g(x) = 1$ we must have that $\pi_\mathfrak{p}(f(x))\pi_\mathfrak{p}(g(x)) = (1 + \mathfrak{p})$. So, $\pi_\mathfrak{p}(f(x))$ is a unit in $(A/\mathfrak{p})[x]$. Yet as $A/\mathfrak{p}$ is an integral domain since $\mathfrak{p}$ is a prime ideal, we know that $(A/\mathfrak{p})[x]^\times = (A/\mathfrak{p})^\times$. Therefore, if $f(x) = a_0 + a_1 x + \ldots + a_n x^n$ then for any $\mathfrak{p} \in \mathrm{Spec}(A)$ we must have that:

$$a_0 + \mathfrak{p} \in (A/\mathfrak{p})^\times \text{ and } a_k \in \mathfrak{p} \text{ for all } k \geq 1.$$

It follows that $a_k \in \mathrm{Nil}(A) = \bigcap_{\mathfrak{p} \in \mathrm{Spec}(A)} \mathfrak{p}$ for all $k \geq 1$.

Finally, to show that $a_0 \in A^\times$ just note that if $g(x) = b_0 + \ldots b_m x^m$ then the constant term of fg is $a_0 b_0 = 1$. This proves that $f \in S$. So $A[x]^\times \subseteq S$. ■

Set 9 Problem 5: Suppose A is a nonzero unital commutative ring and $\mathfrak{a}$ is an ideal of A. Let:

$$\sqrt{\mathfrak{a}} := \{x \in A : \exists n \in \mathbb{Z}^+ \; s.t. \; x^n \in \mathfrak{a}\}.$$

(a) Prove that $\sqrt{\mathfrak{a}}$ is an ideal of A and that $\mathrm{Nil}(A/\mathfrak{a}) = \sqrt{\mathfrak{a}}/\mathfrak{a}$.

First we show that $\sqrt{\mathfrak{a}}$ is an ideal.

Suppose $x, y \in \sqrt{\mathfrak{a}}$ and let $a, b \in A$. We want to show that $ax + by \in \sqrt{\mathfrak{a}}$. So, pick $n, m \in \mathbb{Z}^+$ such that $x^n \in \mathfrak{a}$ and $y^m \in \mathfrak{a}$. Then for any $k > n + m$, we have that:

$$(ax + by)^k = \sum_{i=0}^{k} \binom{k}{i} a^i b^{k-i} x^i y^{k-i}$$

Also, because we picked K large enough, we know that either x^n or y^m divides each term of the right hand sum. Hence, each summand is in $\mathfrak{a}$ and in turn we know that $(ax + by)^k \in \mathfrak{a}$. This proves that $ax + by \in \sqrt{\mathfrak{a}}$.

With that the other claim is now easy. If $(x + \mathfrak{a})^n = 0 + \mathfrak{a}$ then we know that $x^n \in \mathfrak{a}$. Hence, $\mathrm{Nil}(A/\mathfrak{a}) \subseteq \sqrt{\mathfrak{a}}/\mathfrak{a}$. On the other hand, if $x^n \in \mathfrak{a}$ then we know that $(x + \mathfrak{a})^n = 0 + \mathfrak{a}$. Therefore, $x + \mathfrak{a} \in \mathrm{Nil}(A/\mathfrak{a})$ and we've shown that $\mathrm{Nil}(A/\mathfrak{a}) \supseteq \sqrt{\mathfrak{a}}/\mathfrak{a}$.

(b) Prove that $\sqrt{\mathfrak{a}} = \bigcap_{\mathfrak{p} \in V(\mathfrak{a})} \mathfrak{p}$.

To prove this statement it is helpful to refine our statement of the correspondance theorem for ring ideals.

Recall that if $\mathfrak{a}$ is an ideal of a ring A, then there is a bijective correspondance $\pi : \{\mathfrak{b} \in \mathrm{Ideal}(A) : \mathfrak{a} \subseteq \mathfrak{b}\} \to \mathrm{Ideal}(A/\mathfrak{a})$ given by $\mathfrak{b} \mapsto \mathfrak{b}/\mathfrak{a}$. I now claim that $\mathfrak{b}$ is prime in A if and only if $\pi(\mathfrak{b})$ is prime in $A/\mathfrak{a}$.

Proof:

Lemma: If $\mathfrak{b}, \mathfrak{c}$ are ideals of A containing $\mathfrak{a}$ then $(\mathfrak{b}/\mathfrak{a})(\mathfrak{c}/\mathfrak{a}) = (\mathfrak{b}\mathfrak{c})/\mathfrak{a}$.

To see this, note that $(\mathfrak{b}/\mathfrak{a})(\mathfrak{c}/\mathfrak{a})$ is the ideal generated by the elements $bc + \mathfrak{a}$ as (b, c) ranges over $\mathfrak{b} \times \mathfrak{c}$. And as all those elements are in $(\mathfrak{b}\mathfrak{c})/\mathfrak{a}$, we know that $(\mathfrak{b}/\mathfrak{a})(\mathfrak{c}/\mathfrak{a}) \subseteq (\mathfrak{b}\mathfrak{c})/\mathfrak{a}$. On the other hand, note that any element of $(\mathfrak{b}\mathfrak{c})/\mathfrak{a}$ can be written $b_1c_1 + \ldots + b_nc_n + \mathfrak{a} = \sum_{i=1}^{n}(b_i + \mathfrak{a})(c_i + \mathfrak{a}) \in (\mathfrak{b}/\mathfrak{a})(\mathfrak{c}/\mathfrak{a})$.

Now suppose $\mathfrak{p}$ is a prime ideal of A containing $\mathfrak{a}$. Then suppose $\mathfrak{p}/\mathfrak{a} \mid (\mathfrak{b}/\mathfrak{a})(\mathfrak{c}/\mathfrak{a})$ where $\mathfrak{b}, \mathfrak{c} \lhd A$ and contain $\mathfrak{a}$. By our lemma we thus have that $\mathfrak{p}/\mathfrak{a} \supseteq (\mathfrak{b}\mathfrak{c})/\mathfrak{a}$ and from there it is clear that $\mathfrak{p} \supseteq \mathfrak{b}\mathfrak{c}$. Since $\mathfrak{p}$ is a prime ideal, we in turn must have that $\mathfrak{p} \supseteq \mathfrak{b}$ or $\mathfrak{p} \supseteq \mathfrak{c}$. And that let's us conclude that either $\mathfrak{p}/\mathfrak{a} \mid \mathfrak{b}/\mathfrak{a}$ or $\mathfrak{p}/\mathfrak{a} \mid \mathfrak{c}/\mathfrak{a}$. This proves that $\mathfrak{p}/\mathfrak{a}$ is a prime ideal.

Conversely, suppose $\mathfrak{p}, \mathfrak{b}, \mathfrak{c} \lhd A$ and all contain $\mathfrak{a}$. Also suppose that $\mathfrak{p} \mid \mathfrak{b}\mathfrak{c}$ and that $\mathfrak{p}/\mathfrak{a}$ is a prime ideal in $A/\mathfrak{a}$. Then since $\mathfrak{b}\mathfrak{c} \subseteq \mathfrak{p}$ we know that $(\mathfrak{b}/\mathfrak{a})(\mathfrak{c}/\mathfrak{a}) = (\mathfrak{b}\mathfrak{c})/\mathfrak{a} \subseteq \mathfrak{p}/\mathfrak{a}$. By the prime-ness of $\mathfrak{p}/\mathfrak{a}$ we in turn know that either $\mathfrak{p}/\mathfrak{a} \supseteq \mathfrak{b}/\mathfrak{a}$ or $\mathfrak{p}/\mathfrak{a} \supseteq \mathfrak{c}/\mathfrak{a}$. That implies that either $\mathfrak{p} \supseteq \mathfrak{b}$ or $\mathfrak{p} \supseteq \mathfrak{c}$. So, we've shown that $\mathfrak{p}$ is a prime ideal. ■

The important take-away here is that the correspondance theorem gives us a bijection from $V(\mathfrak{a})$ to $\mathrm{Spec}(A/\mathfrak{a})$.

Going back to the homework problem, note by part (a) plus the note right above that $\sqrt{\mathfrak{a}}/\mathfrak{a} = \mathrm{Nil}(A/\mathfrak{a}) = \bigcap\limits_{\mathfrak{p}' \in \mathrm{Spec}(A/\mathfrak{a})} \mathfrak{p}' = \bigcap\limits_{\mathfrak{p} \in V(\mathfrak{a})} \mathfrak{p}/\mathfrak{a} = (\bigcap\limits_{\mathfrak{p} \in V(\mathfrak{a})} \mathfrak{p})/\mathfrak{a}$.

Therefore, by another application of the ideal correspondance theorem we have that $\sqrt{\mathfrak{a}} = \bigcap\limits_{\mathfrak{p} \in V(\mathfrak{a})} \mathfrak{p}$. ■

(I wrote the stuff below out before having it obsoleted by my other reasoning. Still, I feel bad just erasing it all.

Suppose $x \in \sqrt{\mathfrak{a}}$. Then it is clear that $x \in \sqrt{\mathfrak{p}}$ for all $\mathfrak{p} \in V(\mathfrak{a})$ since $x^n \in \mathfrak{a}$ and $\mathfrak{a} \subseteq \mathfrak{p}$ implies that $x^n \in \mathfrak{p}$. So, by part (a) we know that $x + \mathfrak{p} \in \mathrm{Nil}(A/\mathfrak{p})$ for all $\mathfrak{p} \in V(\mathfrak{a})$. Yet note that each $A/\mathfrak{p}$ is an integral domain and thus has a trivial nilradical. It follows that we must have that $x \in \mathfrak{p}$ for all $\mathfrak{p} \in V(\mathfrak{a})$. This proves that $\sqrt{\mathfrak{a}} \subseteq \bigcap_{\mathfrak{p} \in V(\mathfrak{a})} \mathfrak{p}$.)

Set 9 Problem 2: For every commutative unital ring A prove that:

$$A[x]/\langle x^2 - 2\rangle \cong \left\{ \begin{pmatrix} a & b \\ 2b & a \end{pmatrix} : a, b \in A \right\}.$$

Proof:
Note that we can identify A as a subring of $M_2(A)$ by identifying a and $\left(\begin{smallmatrix} a & 0 \\ 0 & a \end{smallmatrix}\right)$. Having done that, it's now clear that the evaluation map e sending $f(x) \in A[x]$ to $f(\left(\begin{smallmatrix} 0 & 1 \\ 2 & 0 \end{smallmatrix}\right))$ is a well-defined ring homomorphism.

Next, note that $x^2 - 2$ is in the kernel of e. After all,
$$\left(\begin{smallmatrix} 0 & 1 \\ 2 & 0 \end{smallmatrix}\right)^2 - \left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right) - \left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right).$$

Furthemore, if f is any other polynomial in $A[x]$, then because the leading term of $x^2 - 2$ is a unit (namely 1) we can apply the long division algorithm to get unique polynomials $q, r \in A[x]$ such that $f(x) = q(x)(x^2 - 2) + r(x)$ and $\deg(r) < 2 = \deg(x^2 - 2)$. In particular, as $\deg(r) = 2$, we know that there exists $a, b \in A$ such that $r(x) = bx + a$. And then by applying e to $f(x)$ we get that:
$$e(f(x)) = e(q(x))e(x^2 - 2) + e(bx + a) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right) + \left(\begin{smallmatrix} a & b \\ 2b & a \end{smallmatrix}\right)$$

This shows that:

- $e(f(x)) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ iff $r(x) = bx + a = 0$. Hence, we've proven that $\ker(e) = \langle x^2 - 2 \rangle$.

- $\operatorname{im}(e) = \{\left(\begin{smallmatrix} a & b \\ 2b & a \end{smallmatrix}\right) : a, b \in A\}$ since the latter is the set of all values that the polynomial $r(x)$ can take on.

By the first isomorphism theorem for rings we are now done. ■

Set 9 Problem 4: Let $\omega := \frac{-1+i\sqrt{3}}{2} = e^{i2\pi/3}$ and $\mathbb{Z}[\omega] := \{a + b\omega : a, b \in \mathbb{Z}\}$

Note that $\mathbb{Z}[\omega]$ is a subring of $\mathbb{C}$. After all, it is clear that $\mathbb{Z}[\Omega]$ is closed under subtraction. Meanwhile, note that:
$$\omega^2 = (\tfrac{-1+i\sqrt{3}}{2})^2 = \tfrac{1-2i\sqrt{3}-3}{4} = \tfrac{-1-i\sqrt{3}}{2} = -1 - \tfrac{-1+i\sqrt{3}}{2} = -1 - \omega$$

Another way you can see this is by noting that ω is a root of the polynomial $x^3 - 1 = (x^2 + x + 1)(x - 1)$. And since $\omega - 1 \neq 0$ we know that $\omega^2 + \omega + 1 = 0$.

Therefore, $\mathbb{Z}[\omega]$ is closed under multiplication with:
$$(a + b\omega)(c + d\omega) = (ab - bd) + (ad + bc - bd)\omega$$

I'll also mention that $\mathbb{Z}[\omega]$ and $\mathbb{C}$ share a multiplicative identity.

(a) Let $N(z) := |z|^2$ and check that $N(a + b\omega) = a^2 - ab + b^2$ for every $a, b \in \mathbb{R}$.

Note that $\omega \cdot (-1 - \omega) = \omega \cdot \omega^2 = 1 = \omega \cdot \overline{\omega}$. Therefore, $-1 - \omega = \overline{\omega}$ and so we can say that:
$$\begin{aligned} |a + b\omega|^2 = (a + b\omega)(\overline{a + b\omega}) &= (a + b\omega)(a + b\overline{\omega}) \\ &= (a + b\omega)(a - b - b\omega) \\ &= a^2 - ab - ab\omega + ab\omega - b^2\omega - b^2\omega^2 \\ &= a^2 - ab - b^2\omega + b^2 + b^2\omega \\ &= a^2 - ab + b^2. \end{aligned}$$

(b) Prove that for every $z_1 \in \mathbb{Z}[\omega]$ and $z_2 \in \mathbb{Z}[\omega] - \{0\}$ there are $q, r \in \mathbb{Z}[\omega]$ such that $z_1 = z_2 q + r$ and $N(r) < N(z_2)$.

Consider the fraction $\frac{z_1}{z_2}$ (which we can do since $z_2 \neq 0$). Importantly, $(1, 0)$ and $(-1/2, \sqrt{3}/2)$ is a basis for $\mathbb{R}^2 \cong \mathbb{C}$ when the latter is viewed as a real vector space. Hence, we can find unique numbers $x, y \in \mathbb{R}$ such that $\frac{z_1}{z_2} = x + y\omega$.

Pick $m, n \in \mathbb{Z}$ such that $|x - m| \leq 1/2$ and $|y - n| \leq 1/2$ and set $q := m + n\omega$ in $\mathbb{Z}[\omega]$. If either $|x - m| < 1/2$ or $|y - n| < 1/2$ then we trivially have by the triangle inequality that:

$$|\tfrac{z_1}{z_2} - q| = |x + y\omega - m - n\omega| \leq |x - m||1| + |y - n||\omega| < 1/2 + 1/2 = 1$$

Meanwhile, if both $|x - m| = 1/2$ and $|y - n| = 1/2$ then we know that either $|\frac{z_1}{z_2} - q| = \frac{1}{2}|1 + \omega|$ or $|\frac{z_1}{z_2} - q| = \frac{1}{2}|1 - \omega|$. Yet as ω and 1 aren't collinear and both have a magnitude of 1, it's clear that $\frac{1}{2}|1 \pm \omega| < \frac{1}{2}(|1| + |\omega|) = 1$. Therefore, in general we have shown that $|\frac{z_1}{z_2} - q| < 1$.

Finally, set $r := z_1 - qz_2$. Then $r \in \mathbb{Z}[\omega]$ and $|r| = |z_1 - qz_2| = |z_2| \cdot |\frac{z_1}{z_2} - q| < |z_2|$. It follows that $z_1 = z_2 q + r$ with $N(r) < N(z_2)$.

(c) Prove that $\mathbb{Z}[\omega]$ is a Euclidean domain and so it is a PID.

It's clear that $N(z) \geq 0$ with $N(z) = 0$ iff $z = 0$. Also, we know that N is integer-valued on $\mathbb{Z}[\omega] - \{0\}$ by part (a). Then the rest of the proof that $\mathbb{Z}[\omega]$ is a Euclidean domain is just part (b).

(d) Prove that $\mathbb{Z}[\omega]^\times = \{\pm 1, \pm\omega, \pm\omega^2\}$.

To start off, it's clear that $\{\pm 1, \pm\omega, \pm\omega^2\} \subseteq \mathbb{Z}[\omega]^\times$. After all, 1 and -1 are their own multiplicative inverses, $\omega^{-1} = \omega^2$, and $(-\omega)^{-1} = -\omega^2$.

To prove that those six elements of $\mathbb{Z}[\omega]$ are the only units, note that if $z_1, z_2 \in \mathbb{Z}[\omega]$ then $N(z_1 z_2) = |z_1 z_2|^2 = |z_1|^2 |z_2|^2 = N(z_1)N(z_2)$. Therefore, if $z_1 z_2 = 1$, we must have that $N(z_1) = \frac{1}{N(z_2)}$. Yet as both $N(z_1)$ and $N(z_2)$ are positive integers, the only way this is possible is if $N(z_1) = N(z_2) = 1$.

So, we can conclude that if $z_1 = a + b\omega$ is a unit then we must have that $N(z_1) = a^2 - ab + b^2 = 1$. After rearranging with the quadratic formula, this is the same as saying that:

$$a = \frac{b \pm \sqrt{-3b^2 + 4}}{2} \text{ (where as a reminder } a, b \in \mathbb{Z}\text{)}$$

If b is an integer with $b \notin \{-1, 0, +1\}$ then it is clear there are no real-valued solutions for a. Meanwhile, for each of those three possible values of b there are 2 real solutions for a. This proves that there can be no more than 6 units in $\mathbb{Z}[\omega]$. ■

Set 9 Problem 3: Prove that $\langle x, 2\rangle$ in $\mathbb{Z}[x]$ is not a principal ideal.

Suppose $\mathfrak{a} := \langle x, 2\rangle$ is principal. Then there exists a nonzero $f \in \mathbb{Z}[x]$ such that $\mathfrak{a} = f(x)\cdot\mathbb{Z}[x]$. But then since $2 \in \mathfrak{a}$ and $x \in \mathfrak{a}$, we know that f must divide both 2 and x.

Now since $\mathbb{Z}$ is an integral domain, we know that $\deg(fg) = \deg(f) + \deg(g) \geq \deg(f)$ for all $g \in \mathbb{Z}[x]$ with $g \neq 0$. Consequently, if $f(x)g(x) = 2$ we must have that $0 = \deg(fg) \geq \deg(f)$. And this proves that f is a constant polynomial. Going a step further, the only divisors of 2 in $\mathbb{Z}$ are ± 1 and ± 2. So, $f(x)$ must equal one of those constants. Yet note that if $f(x) = \pm 2$ then every polynomial in $\langle f(x)\rangle$ with have only even coefficients. This contradicts that $x \in \langle f(x)\rangle$.

So, we've narrowed down that $f(x) = 1$ or -1. Yet even here we have a contradiction. After all, $f(x) = \pm 1$ would imply that $\langle 2, x\rangle = \mathbb{Z}[x]$. Yet, $\langle 2, x\rangle = 2\mathbb{Z}[x] + x\mathbb{Z}[x]$ consists precisely of the polynomials with integer coefficients whose constant term is even. Hence, $\langle 2, x\rangle \neq \mathbb{Z}[x]$.

We thus conclude that no such f exists and $\mathfrak{a}$ is not principal. ■

12/4/2025

Math 220a Homework:

Exercise V.1.4: Let $f(z) = \frac{1}{z(z-1)(z-2)}$. Then give the Laurent Expansion of $f(z)$ in each of the following annuli:

(a) $A(0; 0, 1)$

By doing partial fractions we can solve that $f(z) = \frac{1}{2z} - \frac{1}{z-1} + \frac{1}{2(z-2)}$. Next note that if $|z| < 1$ then $-\frac{1}{z-1} = \frac{1}{1-z} = \sum_{n=0}^{\infty} z^n$. Meanwhile, if $|z| < 2$ then:

$$\frac{1}{2(z-2)} = -\frac{1}{4}\cdot\frac{1}{1-\frac{z}{2}} = -\frac{1}{4}\sum_{n=0}^{\infty}\frac{z^n}{2^n}.$$

Therefore, for $z \in A(0;0,1)$ we have that $f(z) = \frac{1}{2}z^{-1} + \sum\limits_{n=0}^{\infty}(1 - \frac{1}{2^{n+2}})z^n$.

(b) $A(0; 1, 2)$

For this annulus, our power series for $\frac{-1}{z-1}$ from part (a) doesn't work any more. Hence, we need to find a new Laurent series for $\frac{-1}{z-1}$ when $|z| > 1$.

Fortunately, note that $|z^{-1}| < 1$ when $|z| > 1$. Therefore:

$$\frac{-1}{z-1} = \frac{1}{\frac{1}{z^{-1}}-1} = \frac{-z^{-1}}{1-z^{-1}} = -z^{-1}\sum_{n=0}^{\infty} z^{-n} = \sum_{n=1}^{\infty} -z^{-n}$$

And this let's us conclude for $z \in A(0;1,2)$ that $f(z) = -\sum\limits_{n=2}^{\infty} z^{-n} - \frac{1}{2}z^{-1} + \sum\limits_{n=0}^{\infty}\frac{-1}{2^{n+2}}z^n$.

(c) $A(0;2,\infty)$

For this annulus, we don't need to change our treatment of $\frac{1}{2z} - \frac{1}{z-1}$ from part (b) at all. As for the $\frac{1}{2(z-2)}$ term, note that when $|z| > 2$ then $|z^{-1}| < 1/2$. Hence:

$$\frac{1}{2(z-2)} = \frac{1}{2(\frac{1}{z^{-1}}-2)} = \frac{z^{-1}}{2(1-2z^{-1})} = \frac{z^{-1}}{2}\sum_{n=0}^{\infty}(2z^{-1})^n = \sum_{n=0}^{\infty} 2^{n-1}z^{-(n+1)} = \sum_{n=1}^{\infty} 2^{n-2}z^{-n}.$$

So for $z \in A(0;2,\infty)$ we have that:

$$f(z) = \tfrac{1}{2}z^{-1} + \sum_{n=1}^{\infty} -z^{-n} + \sum_{n=1}^{\infty} 2^{n-2}z^{-n} = \sum_{n=2}^{\infty}(2^{n-2}-1)z^{-n} + 0z^{-1}. \blacksquare$$

Exercise V.1.5: Show that $f(z) = \tan(z)$ is analytic in $\mathbb{C}$ except for simple poles (i.e. order 1 poles) at $z = \frac{\pi}{2} + n\pi$ for each integer n. Determine the singular part of f at each of these poles.

Since $\tan(z) := \sin(z)/\cos(z)$ and $\sin(z)$ and $\cos(z)$ are analytic, we know $\tan(z)$ is anaytic anywhere $\cos(z) \neq 0$. Also, note that $\cos(z) = \frac{e^{iz}+e^{-iz}}{2}$. So, $\cos(0) = 0$ implies that $e^{iz} = -e^{-iz} = e^{i\pi - iz}$ for some $n \in \mathbb{Z}$. Thus, we we can say that $iz = i\pi - iz + i2n\pi$ for some $n \in \mathbb{Z}$. Or in other words, $z = \frac{(2n+1)\pi}{2} = n\pi + \frac{\pi}{2}$ for some $n \in \mathbb{Z}$.

Thus $\cos(z)$ has the same zeros in the complex plane as in the real number line and in turn $\tan(z)$ is analytic except at $z = \frac{\pi}{2} + n\pi$ for each integer n.

Next, we show that each of the singularities of $\tan(z)$ are simple poles. Note that if $\cos(t) = 0$ then $\cos'(t) = \sin(t) \neq 0$ for any $t \in \mathbb{R}$. Hence, all of the zeros of $\cos$ are simple and for any singularity p of $\cos$ we can write $\cos(z) = (z-p)h(z)$ for some analytic function $h : \mathbb{C} \to \mathbb{C}$ with $h(p) \neq 0$. But now $\tan(z) = \frac{1}{z-p} \cdot \frac{\sin(z)}{h(z)}$ where $\frac{\sin}{h}$ is analytic (and also crucially nonzero at p). This proves that each singularity p of $\tan(z)$ is a simple pole.

Finally, we now know that the Laurent series of $\tan(z)$ on the punctured disk about any pole p has the form $\sum_{n=-1}^{\infty} a_n(z-p)^n$. It follows that $(z-p)\tan(z)$ is represented by a power series about p whose constant term is a_{-1}. So, if we can calculate

$$a_{-1} = \lim_{z\to p}(z-p)\tan(z),$$

then we will have that the singular part of f at the pole p is $\frac{a_{-1}}{z-p}$.

Note that $\tan(z+\pi) = \tan(z)$. After all:

$$\tan(z+\pi) = -i\frac{e^{i(z+\pi)}-e^{-i(z+\pi)}}{e^{i(z+\pi)}+e^{-i(z+\pi)}} = -i\frac{-e^{iz}+e^{-iz}}{-e^{iz}-e^{-iz}} = -i\frac{e^{iz}-e^{-iz}}{e^{iz}+e^{-iz}} = \tan(z)$$

Therefore, for any $n \in \mathbb{Z}$ we have that:

$$\lim_{z\to\frac{\pi}{2}+n\pi}(z - n\pi - \tfrac{\pi}{2})\tan(z) = \lim_{z\to\frac{\pi}{2}}(z-\tfrac{\pi}{2})\tan(z+n\pi) = \lim_{z\to\frac{\pi}{2}}(z-\tfrac{\pi}{2})\tan(z)$$

Since we already know the limit exists and are just trying to numerically evaluate it, we can approach the limit via any path. Hence, it suffices to assume $z \in \mathbb{R}$ and that is nice because we can now use L'Hôpital's rule to quickly evaluate that:

$$a_{-1} = \lim_{x\to\frac{\pi}{2}}\frac{(x-\frac{\pi}{2})\sin(x)}{\cos(x)} = \lim_{x\to\frac{\pi}{2}}\frac{\sin(x)+(x-\frac{\pi}{2})\cos(x)}{-\sin(x)} = -1.$$

So, the singular part at $n\pi + \frac{\pi}{2}$ is $\frac{-1}{z-n\pi-\frac{\pi}{2}}$ for each $n \in \mathbb{Z}$. $\blacksquare$

Exercise V.1.6: If $f : G \to \mathbb{C}$ is analytic except for poles, show that the poles of f cannot have a limit point in G.

Given an analytic function h let $Z(h)$ and $P(h)$ denote the zero set and pole set respectively of h. Now, if f is the function from the problem statement, define:

$g(z) = 1/f(z)$ when $z \in G - (Z(f) \cup P(f))$ and $g(z) = 0$ when $z \in P(f)$

It is trivial to see that $Z(g) = P(f)$. Additionally, we claim that g is analytic on $G - Z(f)$ and that $P(g) = Z(f)$.

To see this first note that if $f(z_0) \neq 0$ then there is some neighborhood N about z_0 on which $g(z) = 1/f(z)$ for all $z \in N$. In turn, it is easy to see that g is analytic at z_0. Meanwhile, if f has a pole at z_0, then we know that $\lim_{z \to z_0} 1/f(z) = 0$. Hence, g is continuous at z_0. Furthermore, if we let $r > 0$ be such that f is analytic on $A(z_0; 0, r)$, then we know that g is continuous on $B_r(z_0)$ and differentiable on $B_r(z_0) - \{z_0\}$. This implies that g is actually differentiable and in turn analytic on all of $B_r(z_0)$ including z_0.

Finally, note that since $f(z) \to 0$ as $z \to z_0$ for any $z_0 \in Z(f)$, we know that $g(z) \to \infty$ as $z \to z_0$ for any $z_0 \in Z(f)$.

But now we know that $Z(g)$ can have no limit points in the open set $G - Z(f)$. (Note: we know $Z(f)$ is closed in G because we can view f as a continuous well-defined function from G to $\mathbb{C}_\infty$) Equivalently, this means that $P(f)$ can have no limits point in $G - Z(f)$. Finally, since f is continuous at all points in $Z(f)$, we can find a pole-free neighborhood of z_0 for each $z_0 \in Z(f)$ (namely by just taking a neighborhood of z_0 where f is bounded). So $P(f)$ has no limit points in $Z(f)$ either. ■

Exercise V.1.7: Let f be an analytic function with an isolated singularity at $z = a$ and suppose f is not the constant zero function around a. Show that if either of the two following equations hold:

$(*)$ $\lim_{z \to a} |z - a|^s \cdot |f(z)| = 0$,

$(**)$ $\lim_{z \to a} |z - a|^s \cdot |f(z)| = \infty$,

then there is an integer m such that $(*)$ holds if $s > m$ and $(**)$ holds if $s < m$.

Note that if $\lim_{z \to a} |z-a|^s \cdot |f(z)| = 0$ then we also know that $\lim_{z \to a} |z-a|^r \cdot |f(z)| = 0$ for any $r > s$. After all, $|z-a|^r \cdot |f(z)| = |z-a|^{r-s} \cdot (|z-a|^s \cdot |f(z)|)$ and $|z-a|^b \to 0$ as $z \to a$ so long as $b > 0$. Consequently, if $(*)$ holds for some s then we can find a positive integer n such that $|z-a|^n \cdot |f(z)| = |(z-a)^n f(z)| \to 0$ as $z \to a$. Next, note that $g(z) = (z-a)^n f(z)$ when $z \neq a$ and $g(z) = 0$ is analytic on some ball around a. Hence, we can find coefficients $\{b_k\}_{k \in \mathbb{Z}}$ such that:

$$f(z) = \sum_{k=-n}^{\infty} b_k (z-a)^k$$

Meanwhile, suppose $\lim_{z \to a} |z-a|^s \cdot |f(z)| = \infty$. Then similarly to before we know that $\lim_{z \to a} |z-a|^r \cdot |f(z)| = \infty$ for any $r < s$ since $|z-a|^b \to \infty$ as $z \to a$ so long as $b < 0$ and $|z-a|^r \cdot |f(z)| = |z-a|^{r-s} \cdot (|z-a|^s \cdot |f(z)|)$. Consequently, if $(**)$ holds for some s then we can find a negative integer n' such that $|z-a|^{n'} \cdot |f(z)| = |(z-a)^{n'} f(z)| \to \infty$ as $z \to a$. This proves that $(z-a)^{n'} f(z)$ has a pole at a and so there exists some negative integer n'' such that:

$$(z-a)^{n'} f(z) = \sum_{k=n''}^{\infty} b_k (z-a)^k$$

Finally, setting $n = n'' - n'$ we have (after relabeling the coefficients b_k) that:

$$f(z) = \sum_{k=n}^{\infty} b_k (z-a)^k$$

Thus if either $(*)$ or $(**)$ holds, we've proven that there is an integer $n \in \mathbb{Z}$ such that $f(z) = \sum_{k=n}^{\infty} b_k(z-a)^k$. Now as f is not the zero function, we must have that at least one b_k in the expansion is not 0. By the well-ordering of the integers greater than n, we can thus pick a least integer M such that $b_M \neq 0$. In other words,

$$f(z) = \sum_{k=M}^{\infty} b_k(z-a)^k \text{ where } b_M \neq 0$$

Finally, set $m := -M$. Then it is not difficult to see that $(*)$ holds if $s > m$ and $(**)$ holds if $s < m$.

Write $f(z) = (z-a)^{-m} \cdot \sum_{n=0}^{\infty} b_{n-m}(z-a)^n$. That way, we know for any $s \in \mathbb{R}$ that:

$$|z-a|^s \cdot |f(z)| = |z-a|^{s-m} \cdot \left|\sum_{n=0}^{\infty} b_{n-m}(z-a)^n\right|.$$

Next note that $\left|\sum_{n=0}^{\infty} b_{n-m}(z-a)^n\right| \to b_{-m} \neq 0$ as $z \to a$. Therefore, we know that:

$$\lim_{z\to a} |z-a|^s \cdot |f(z)| = b_{-m} \cdot \lim_{z\to a} |z-a|^{s-m}.$$

Now if $s > m$ then $|z-a|^{s-m} \to 0$ as $z \to a$. Meanwhile, if $s < m$ then $|z-a|^{s-m} \to \infty$ as $z \to a$. ■

Exercise V.1.8: Let f, a, and m be as in the last exercise. Then show:

(a) $m = 0$ iff $z = a$ is a removable singularity and $f(a) \neq 0$.

(b) $m < 0$ iff $z = a$ is a removable singularity and f has a zero at $z = a$ of order $-m$

(c) $m > 0$ iff $z = a$ is a pole of f of order m.

If such an m exists then all the $(\Longrightarrow)$ directions of the above statements are obvious. After all, in the previous exercise I showed that $-m$ is the least integer index of the nonzero coefficients of the Laurent series of f in the punctured ball about a. Furthermore, by the prior exercise I know m exists if f satisfies $(*)$ or $(**)$ for some $s \in \mathbb{R}$. So, it suffices to show that if f has a removable singularity or pole at a then $(*)$ or $(**)$ holds for some $s \in \mathbb{R}$.

Luckily, f has a removable singularity at a iff $|z-a|^1 \cdot |f(z)| \to 0$ as $z \to a$, thus showing that $(*)$ holds for some $s \in \mathbb{R}$. Meanwhile, if f has a pole at a then we know that $|z-a|^0 \cdot |f(z)| \to \infty$ as $z \to a$, thus showing that $(**)$ holds for some $s \in \mathbb{R}$. ■

Exercise V.1.9: A function f has an essential singularity at $z = a$ iff neither $(*)$ nor $(**)$ holds for any $s \in \mathbb{R}$.

This is literally just asking for the contrapositive in spirit of V.1.8.

After all, if $(*)$ or $(**)$ holds for any s then we can find m as in exercise 7. Then by the forward directions of exercise 8 we know that f has a removable singularity or pole at a. This proves that a is not an essential singularity. Meanwhile, if f doesn't have an essential singularity at a, then it must have a pole or removable singularity. So, by the backwards direction of exercise 8 we can find $m \in \mathbb{Z}$ with $(*)$ holding for all $s > m$. ■

12/5/2025

Math 241a Notes:

I'm going back to covering Zimmer's book: *Essential Results Of Functional Analysis* instead of Rudin's book. As a bit of notation, if $\mathcal{X}$ is a normed vector space we let $\mathcal{X}_r$ denote the set of all points in $\mathcal{X}$ with $\|x\| \leq r$.

Theorem 2.2.3: Let $\mathcal{X}$ be a Banach space and G be a compact group. Let $\pi : G \to \mathrm{Iso}(\mathcal{X})$ be a continuous isometric representation of G (where $\mathrm{Iso}(\mathcal{X})$ has the strong operator topology). Let $A \subseteq \mathcal{X}_1^*$ be a nonempty compact convex G-invariant subset (for the adjoint π^* of π acting on $\mathcal{X}^*$) where $\mathcal{X}_1^*$ has the weak* topology. Then there is a G-fixed point in A (meaning $\pi^*(g)\lambda = \lambda$ for all $g \in G$).

Lemma 2.2.4: Fix any $x \in \mathcal{X}$ and define $\|\lambda\|_0$ for all $\lambda \in \mathcal{X}^*$ by

$$\|\lambda\|_0 = \sup\{|\lambda(\pi(g)x)| : g \in G\}.$$

Then:

(i) $\|\cdot\|_0$ is a well-defined semi-norm on $\mathcal{X}^*$.

(ii) $\|\cdot\|_0$ is G-invariant (i.e. for all $\lambda \in \mathcal{X}^*$ and $h \in G$ we have that $\|\pi^*(h)\lambda\|_0 = \|\lambda\|_0$).

(iii) For any $0 \leq r < \infty$, $\|\cdot\|_0 : \mathcal{X}^* \to \mathbb{R}$ is continuous on $\mathcal{X}_r^*$ with the weak* topology.

(iv) If $\lambda(x) \neq 0$ then $\|\lambda\|_0 > 0$.

Proof:

To start off, since π is a continuous representation (where $\mathrm{Iso}(\mathcal{X})$ has the strong operator topology), we know that the map $g \mapsto \pi(g)x$ is continuous. After all, $\|\pi(g_i)y - \pi(g)y\| \to 0$ as $g_i \to g$ for any $y \in \mathcal{X}$ by the definition of the strong operator topology. It follows that the map $g \mapsto \pi(g)x$ has a compact image in $\mathcal{X}$. And as $|\lambda|$ is continuous on $\mathcal{X}$ we know that $\sup\{|\lambda(\pi(g)x)| : g \in G\}$ exists.

Now the rest of (i) and (iv) are trivial to see. Meanwhile, to show (ii) note that:

$$\begin{aligned}\|\pi^*(h)\lambda\|_0 = \|\lambda \circ \pi(h^{-1})\|_0 &= \sup\{|\lambda(\pi(h^{-1})\pi(g)x)| : g \in G\} \\ &= \sup\{|\lambda(\pi(h^{-1}g)x)| : g \in G\} \\ &= \sup\{|\lambda(\pi(g)x)| : g \in G\} = \|\lambda\|_0\end{aligned}$$

Finally, we prove (iii). Suppose $\lambda_\alpha \to \lambda$ in the weak* topology where $\|\lambda_\alpha\|, \|\lambda\| \leq r$. Then $|\|\lambda_\alpha\|_0 - \|\lambda\|_0| \leq \|\lambda_\alpha - \lambda\|_0$ and we claim the latter goes to zero. After all, let $\beta_\alpha = \lambda_\alpha - \lambda$ and then fix any $\varepsilon > 0$.

Since $\{\pi(g)x : g \in G\}$ is compact (and thus totally bounded) in $\mathcal{X}$ (viewed as a metric space with the norm topology), we can choose a finite set $g_1, \ldots, g_n \in G$ such that for all $g \in G$ there exists i with $\|\pi(g)x - \pi(g_i)x\| < \varepsilon/4r$. In turn, because $\beta_\alpha \to 0$ we know that eventually $|\beta_\alpha(\pi(g_i)x)| < \varepsilon/2$ for each i. Also, we know that $\|\beta_\alpha\|_{\mathrm{op}} \leq 2r$. Therefore, we can conclude that for any $g \in G$ eventually we have that:

$$\begin{aligned}|\beta_\alpha(\pi(g)x)| &\leq |\beta_\alpha(\pi(g)x - \pi(g_i)x)| + |\beta_\alpha(\pi(g_i)x)| \\ &< 2r\|\pi(g)x - \pi(g_i)x\| + \varepsilon/2 \\ &< 2r \cdot \tfrac{\varepsilon}{4r} + \varepsilon/2 = \varepsilon\end{aligned}$$

This proves that $\|\lambda_\alpha - \lambda\|_0 = \|\beta_\alpha\|_0 \to 0$.

Now we go to actually prove theorem 2.2.3. Consider the family:

$$\mathcal{F} := \{B \subseteq A : B \text{ is nonempty, compact, convex, and } G\text{-invariant}\}.$$

Note that $\mathcal{F}$ is nonempty as $A \in \mathcal{F}$. Furthermore, suppose $\mathcal{F}_0$ is a subcollection of $\mathcal{F}$ that is simply ordered by inclusion. Then upon setting $C = \bigcap_{B \in \mathcal{F}_0} B$ we have that:

- C is nonempty by the finite intersection property of A.
- As $\mathcal{X}^*$ is Hausdorff, we know all the $B \in \mathcal{F}$ are closed. Thererefore C being an intersection of closed sets is itself a closed subset of the compact set A and is therefore compact itself.
- C is convex because intersections of convex sets are convex.
- Finally, note that if $\lambda \in B$ for all $B \in \mathcal{F}_0$ then so is $\pi^*(g)\lambda$ since each B is G-invariant. Hence, $\lambda \in C \implies \pi^*(g)\lambda \in C$ and we've prove that C is G-invariant.

It follows by Zorn's lemma that $\mathcal{F}$ has a minimal set. For ease of notation we'll just set A equal to that minimal nonempty, compact, convex, and G-invariant set.

Now we claim that A is a singleton. After all, suppose not. Then we can find $x \in E$ such that $\lambda_1(x) \neq \lambda_2(x)$ for some $\lambda_1, \lambda_2 \in A$. Next, we define the seminorm $\|\cdot\|_0$ as in lemma 2.2.4 using x.

- For each $\lambda \in A$ and $r > 0$ we let $B_r(\lambda) := \{\beta \in A : \|\lambda - \beta\|_0 < r\}$. By part (iii) of lemma 2.2.4 we know that each $B_r(\lambda)$ is open in A. Also, by the triangle inequality we know that each $B_r(\lambda)$ is convex.
- Also, for each $\lambda \in A$ and $r \geq 0$, we define $\overline{B}_r(\lambda) := \{\beta \in A : \|\lambda - \beta\|_0 \leq r\}$. By part (iii) of lemma 2.2.4 we know that each $\overline{B}_r(\lambda)$ is closed in A. Also, by the triangle inequality we know that each $\overline{B}_r(\lambda)$ is convex.
- Let $d = \sup\{\|\lambda - \beta\|_0 : \lambda, \beta \in A\}$. Since A is compact and $\|\cdot\|_0$ is continuous on A we know $d < \infty$. Also, since $\|\lambda_1(x) - \lambda_2(x)\|_0 > 0$ (where $\lambda_1, \lambda_2 \in A$ are from before) we know $d > 0$.

Since A is compact, we can find a finite set $\omega_1, \ldots, \omega_n \in A$ such that $A = \bigcup_{i=1}^n B_{d/2}(\omega_i)$. So, set $\omega := \frac{1}{n}\sum_{i=1}^n \omega_i$. Since A is convex, we know that $\omega \in A$. Also, for any $\lambda \in A$ we have that:

$$\|\omega - \lambda\|_0 \leq \tfrac{1}{n}\textstyle\sum_{i=1}^n \|\omega_i - \lambda\|_0 \leq \tfrac{1}{n}(\tfrac{d}{2} + (n-1)d) = \tfrac{2n-1}{2n}d$$

Therefore, upon setting $r = \frac{2n-1}{2n}d$ we have that $\overline{B}_r(\omega) = A$. But now in turn we know that $\omega \in \bigcap_{\lambda \in A} \overline{B}_r(\lambda)$. Hence, $B := \bigcap_{\lambda \in A} \overline{B}_r(\lambda)$ is a nonempty compact convex subset of A.

Finally, to see that B is G-invariant, note that because $\|\cdot\|_0$ is G-invariant we know that $\overline{B}_r(\pi^*(g)\lambda) = \pi^*(g)\overline{B}_r(\lambda)$ for all $\lambda \in A$. Therefore for each $g \in G$ we have that:

$$\pi^*(g)B = \pi^*(g)\bigcap_{\lambda \in A}\overline{B}_r(\lambda) = \bigcap_{\lambda \in A}\pi^*(g)\overline{B}_r(\lambda) = \bigcap_{\lambda \in A}\overline{B}_r(\pi^*(g)\lambda) = \bigcap_{\lambda \in A}\overline{B}_r(\lambda) = B$$

And now we have a contradiction since A was minimal.

Now that we know $A = \{\lambda\}$ for some $\lambda \in \mathcal{X}^*$, the fact that A is G-invariant must mean that $\pi^*(g)\lambda = \lambda$ for all $g \in G$. Hence we are done. ■

Corollary 2.2.5: If G is any compact group and X is any compact Hausdorff space on which G acts continuously, then there is a G-invariant Radon probability measure on X.

Proof:

Do the construction on *pages 439-440* and then just use the prior theorem instead of the Kakutani-Markov fixed point theorem. ■

I'll also note that if X isn't a compact metric space, then we don't necessarily have that all probability measures are Radon. That said, the proof of corollary 1.1.29 on page 282 still shows that the space of Radon probability measures is compact in the weak* topology of $C_c(X)^* = C(X)^*$.

Note: If X is not compact but just locally compact, then we no longer have that the constant 1 function is in $C_c(X)$ and hence the reasoning in said proof on page 282 breaks down.

Also, the space of Radon probability measures is convex since convex linear combinations of probability measures gives another probability measure and the space of finite Radon measures is closed under scalar multiplication and addition.

Finally, we know the space of Radon probability measures is G-invariant because of the following theorem plus the fact that measure pushforwards preserve the positivity of a measure and the total measure.

If X and Y are topological spaces, a function $f : X \to Y$ is called proper if $f^{-1}(K)$ is compact for all compact $K \subseteq Y$. Here are two easy sufficient conditions for when $f : X \to Y$ is proper.

- If $f : X \to Y$ is continuous, X is compact, and Y is Hausdorff, then f is proper.

  Why? Suppose $K \subseteq Y$ is compact. Then we know that K is closed. In turn $f^{-1}(K)$ is a closed subset of the compact set X. So, $f^{-1}(K)$ is compact.

- If $f : X \to Y$ is a bijective function with a continuous inverse, then f is proper.

  Why? If f^{-1} is continuous and K is compact then $f^{-1}(K)$ is also compact.

Lemma: Suppose $f : X \to Y$ is proper and continuous, and Y is an LCH space. Then f is a closed map (meaning $f(C)$ is closed for all closed sets C).

Suppose $y \in Y - f(C)$. Then we know there is a precompact open set $V \subseteq Y$ containing y. In turn, since f is proper we know that $f^{-1}(\overline{V})$ is compact in X.

Let $E = C \cap f^{-1}(\overline{V})$. Then we know E is a closed subset of the compact set $f^{-1}(\overline{V})$. Hence, E is compact. In turn we also know $f(E)$ is compact. And since Y is Hausdorff, we know that $f(E)$ is closed.

Finally, set $U = V - f(E)$. Then U is open in Y. Also, since $y \notin f(C)$ we know that $y \notin f(E) \subseteq f(C)$. Therefore, we know that $y \in U$. Thirdly, note that if $y' \in f(C) \cap U$ then there must exist $x \in C - E = C - f^{-1}(\overline{V})$ such that $f(x) = y$. But then we have a contradiction since that implies $y \notin \overline{V}$ and yet we know that $U \subseteq \overline{V}$. Hence, we conclude that U is disjoint from $f(C)$.

All in all, this proves that $y \notin \overline{f(C)}$. And since y was arbitrary we know that $\overline{f(C)} = f(C)$. Hence, f is closed. ■

Theorem: Suppose X and Y are LCH spaces, μ is a Radon measure on X, and $\varphi : X \to Y$ is a proper, continuous map. Then $\varphi_*\mu$ is Radon.

Proof:

- If K is compact then $\varphi_*\mu(K) = \mu(\varphi^{-1}(K))$ is finite since μ is Radon and $\varphi^{-1}(K)$ is compact on account of φ being proper.

- Let U be any open set in Y. Then by the continuity of φ we know that $\varphi^{-1}(U)$ is open in X. Furthermore, for any compact $K \subseteq \varphi^{-1}(U)$ we know that $\varphi(K)$ is a compact subset of U (again since φ is continuous). Also, $\varphi^{-1}(\varphi(K)) \subseteq \varphi^{-1}(U)$ for any compact K and clearly $K \subseteq \varphi^{-1}(\varphi(K))$. Hence:
$$\begin{aligned}\varphi_*\mu(U) = \mu(\varphi^{-1}(U)) &= \sup_{K \subseteq \varphi^{-1}(U)} \mu(K) \\ &\leq \sup_{K \subseteq \varphi^{-1}(U)} \mu(\varphi^{-1}(\varphi(K))) \leq \sup_{K \subseteq U} \mu(\varphi^{-1}(K)) = \sup_{K \subseteq U} \varphi_*\mu(K)\end{aligned}$$

  This proves that $\varphi_*\mu$ is inner regular on open sets.

  As a side note, when proving inner regularity we didn't need φ to be proper.

- Suppose E is any Borel set in Y. Then let U be any open set in X with $\varphi^{-1}(E) \subseteq U$. We claim there is an open set $V \subseteq Y$ such that $E \subseteq V$ and $\varphi^{-1}(V) \subseteq U$.

  To see why, note by our prior lemma that φ is a closed map. Therefore, we know $\varphi(X - U)$ is closed in Y. Furthermore, $E \cap \varphi(X - U) = \emptyset$. After all, if $y \in E \cap \varphi(X - U)$ then we know there is $x \in X - U$ such that $\varphi(x) = y \in E$. But this is impossible since $\varphi^{-1}(E) \subseteq U$. Therefore, if we set $V = Y - \varphi(X - U)$ then we have that V is an open set containing E. To show that $\varphi^{-1}(V) \subseteq U$, suppose $x \in \varphi^{-1}(V) \cap (X - U)$. Then we have a contradiction as $y = \varphi(x) \in V \cap \varphi(X - U)$.

  Now it is easy to see that the outer regularity of μ implies the outer regularity of $\varphi_*\mu$. ■

One quick observation:

Corollary 2.2.5 gives us another proof of the extistance of Haar measures on compact groups. After all, if G is a compact group then we can apply that corollary to the action $G \curvearrowright G$ via left translation to get a probability measure that is invariant under left translations.

12/7/2025

Math 220a Notes (for my final tomorrow):

If f has an isolated singularity at a, then we can consider the Laurent series $\sum_{n=-\infty}^{\infty} b_n(z-a)^n$ of f on some annulus $A(a;0,r)$. Then, we say $\mathrm{Res}(f;a) \coloneqq b_{-1}$ is the residue of f at $z=a$.

The Residue Theorem: Let f be analytic in the region G except for the isolated singularities $a_1, a_2, \ldots, a_n$. If γ is a closed piecewise C^1 curve in G which does not pass through any of the points a_k and if $\gamma \approx_G 0$ then:

$$\frac{1}{2\pi i}\int_\gamma f = \sum_{k=1}^{n} n(\gamma;a_k)\mathrm{Res}(f;a_k)$$

Proof:
Let $m_k = n(\gamma;a_k)$ for $1 \leq k \leq m$. Then for each a_k pick some radius r_k such that $\overline{B_{r_k}(a_k)} \subseteq G - \{\gamma\}$ and each of the $\overline{B_{r_k}(a_k)}$ are disjoint. This can be done by induction because for any $J < m$ we have that $G - \{\gamma\} - \bigcup_{k=1}^{J} \overline{B_{r_k}(a_k)} - \bigcup_{k=J+2}^{m}\{a_k\}$ is an open set containing a_{J+1}.

Now set $\gamma_k = a_k + r_k e^{(-2\pi i m_k t)}$ for each $k \in \{1, \ldots, m\}$. That way:

$n(\gamma;a_j) + \sum_{k=1}^{m} n(\gamma_k;a_j) = 0$ for all $j \in \{1,\ldots,m\}$.

Also since we have that $n(\gamma_k;a) = 0$ when $a \notin G$ since $\gamma_k \sim_G 0$, we can conclude via Cauchy's theorem that $\int_\gamma f + \sum_{k=1}^{m}\int_{\gamma_k} f = 0$.

But now note that if $f = \sum_{n=-\infty}^{\infty} b_n(z-a_k)^n$ is the Laurent series of f on the punctured disk about a_k, we know that that sum converges uniformly on $\{\gamma_k\}$. Hence by *(Conway) lemma IV.2.7 on page 368* we can conclude that:

$$\int_{\gamma_k} f = \int_{\gamma_k} \sum_{n=-\infty}^{\infty} b_n(z-a_k)^n = \sum_{n=-\infty}^{\infty} b_n \int_{\gamma_k}(z-a_k)^n.$$

But note if $n \neq -1$ then $(z-a)^n$ has a primitive and so $\int_{\gamma_k}(z-a)^n = 0$.

As a side note: Gee I really overcomplicated my solution to exercise IV.5.5 (see *pages 405-406*). You can tell that I was handed a "Cauchy theorem"-shaped hammer prior to doing that exercise and then viewed that exercise as a "Cauchy theorem"-shaped nail.

Therefore:

$$\begin{aligned}\sum_{n=-\infty}^{\infty} b_n \int_{\gamma_k}(z-a_k)^n = b_{-1}\int_{\gamma_k}(z-a_k)^{-1} &= \mathrm{Res}(f;a_k)\cdot 2\pi i n(\gamma_k;a_k) \\ &= -2\pi i \mathrm{Res}(f;a_k) n(\gamma;a_k)\end{aligned}$$

And this proves that $\int_\gamma f = +2\pi i \sum_{k=1}^{m} \mathrm{Res}(f;a_k)n(\gamma;a_k)$. ■

While we can use integrals to calculate residues, it's more common to use residues to calculate integrals. Hence, we need a separate method for calculating residues.

<u>(Conway) Proposition V.2.4:</u> Suppose f has a pole of order m at $z = a$ and put:

$g(z) \coloneqq (z-a)^m f(z)$ when $z \neq a$ and $g(a) \coloneqq \lim_{z\to a}(z-a)^m f(z)$.

Then $\mathrm{Res}(f;a) = \frac{1}{(m-1)!}g^{(m-1)}(a)$.

Why?

We know g is analytic on a neighborhood of a with $g(z) = \sum_{n=0}^{\infty} b_n(z-a)^n$ (where $b_n = g^{(n)}/n!$). In turn, we have that the Laurent series of f in the punctured disk about a is:

$$f(z) = \frac{b_0}{(z-a)^m} + \cdots + \frac{b_{m-1}}{(z-a)} + \sum_{k=0}^{\infty} b_{m+k}(z-a)^k$$

Therefore, $\mathrm{Res}(f;a) = b_{m-1} = \frac{1}{(m-1)!}g^{(m-1)}(a)$. ■

I'm now going to solve a bunch of integrals using residues:

- Let $I_1 = \int_0^\infty \frac{\cos(x)}{1+x^2}\mathrm{d}x$.

  As $\frac{\cos(x)}{1+x^2}$ is an even function we know that $I_1 = \frac{1}{2}\int_{-\infty}^{\infty}\frac{\cos(x)}{1+x^2}\mathrm{d}x$. Also, because $\cos(x) = \mathrm{Re}(e^{ix})$ we can say that $I_1 = \frac{1}{2}\mathrm{Re}(\int_{-\infty}^{\infty}\frac{e^{ix}}{1+x^2}\mathrm{d}x)$. So, we now seek to evaluate $\int_{-\infty}^{\infty} f(x)\mathrm{d}x$ where $f : \mathbb{C} - \{\pm i\} \to \mathbb{C}$ is an analytic function defined by $f(z) = \frac{e^{iz}}{1+z^2}$.

  To start off, because $|f(x)| = \frac{1}{1+x^2}$ has a finite integral on $\mathbb{R}$, we can say without any ambiguity that $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \lim_{r\to\infty}\int_{-r}^{r} f(x)\mathrm{d}x$. Next, for any $r > 0$ we let γ_r be the path $[-r, r]$ concatenated with the path μ_r given by re^{it} where $t \in [0, \pi]$.

  A fact that I'll be using a bunch in this section is that if $a, b \in \mathbb{R}$ and f is an integrable function then the countour integral $\int_{[a,b]} f(z)\mathrm{d}z$ is equal to the integral $\int_a^b f(x)\mathrm{d}x$. To see this, just note by a simple substitution that:

  $$\int_{[a,b]} f(z)\mathrm{d}z = \int_0^1 f(a + t(b-a))\cdot(b-a)\mathrm{d}t = \int_a^b f(x)\mathrm{d}x$$

  By the Residue theorem, we have that $\int_{-r}^{r} f(x)\mathrm{d}x + \int_{\mu_r} f(z)\mathrm{d}z = 2\pi i \cdot \mathrm{Res}(f;i)$. Equivalently, this means that:

  $$\int_{-r}^{r} f(x)\mathrm{d}x = \mathrm{Res}(f;i) - \int_{\mu_r}\frac{e^{iw}}{1+w^2}\mathrm{d}w = \mathrm{Res}(f;i) - \int_{\mu_r}\frac{e^{iw}}{1+w^2}\mathrm{d}w$$

  Next note that $\left|\int_{\mu_r}\frac{e^{iw}}{1+w^2}\mathrm{d}w\right| = \left|\int_0^\pi \frac{e^{i(r\exp(it))}}{1+r^2e^{2it}}\cdot rie^{it}\mathrm{d}t\right| \leq r\int_0^\pi\left|\frac{e^{i(r\exp(it))}}{1+r^2e^{2it}}\right|\mathrm{d}t$.

  Yet $|e^{i(r\exp(t))}| = e^{r\cdot\mathrm{Re}(i\cos(t)-\sin(t))} = e^{-r\sin(t)} \leq 1$ when $t \in [0,\pi]$. Meanwhile, if $r > 1$ then we know that $|1 + r^2e^{2it}| \geq r^2 - 1$. Hence:

  $$r\int_0^\pi\left|\frac{e^{i(r\exp(it))}}{1+r^2e^{2it}}\right|\mathrm{d}t \leq r\int_0^\pi \frac{1}{r^2-1}\mathrm{d}t = \frac{\pi r}{r^2-1}$$

  But now the latter expression goes to 0 as $r\to\infty$. Therefore:

  $$\begin{aligned}\int_{-\infty}^{\infty} f(x)\mathrm{d}x &= \lim_{r\to\infty}\int_{-r}^{r} f(x)\mathrm{d}x \\ &= 2\pi i\mathrm{Res}(f;i) - \lim_{r\to\infty}\int_{\mu_r} f(w)\mathrm{d}w = 2\pi i\mathrm{Res}(f;i) - 0\end{aligned}$$

Finally, we calculate $\mathrm{Res}(f;i)$. Note that f has a pole at $+i$ and that $(z-i)f(z) = \frac{e^{iz}}{(z+i)}$ is analytic at $+i$. Therefore, by proposition V.2.4 we have that:

$$\mathrm{Res}(f;a) = \frac{1}{(1-1)!} \cdot [\frac{e^{iz}}{(z+i)}]^{(1-1)}(+i) = \frac{e^{i(+i)}}{i+i} = \frac{1}{2ei}$$

Therefore $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 2\pi i \cdot \frac{1}{2ei} = \frac{\pi}{e}$. And finally, we have that:

$$\boxed{\int_0^\infty \frac{\cos(x)}{1+x^2}\mathrm{d}x = I_1 = \frac{1}{2}\mathrm{Re}(\frac{\pi}{e}) = \frac{\pi}{2e}.} \blacksquare$$

- Let $I_2 = \int_0^\pi \frac{1}{a+\cos(\theta)}\mathrm{d}\theta$ where $a > 1$.

  To start off, note that if $z = e^{i\theta}$ then we have that $z^{-1} = \overline{z}$ for all $\theta \in \mathbb{R}$. In turn:

  $$a + \cos(\theta) = a + \mathrm{Re}(z) = a + \frac{z}{2} + \frac{1}{2z} = \frac{z^2+2az+1}{2z}$$

  Also note that $\int_0^\pi \frac{1}{a+\cos(\theta)}\mathrm{d}\theta = \int_\pi^{2\pi} \frac{1}{a+\cos(\theta)}\mathrm{d}\theta$. Therefore, after letting γ be the path e^{it} where $t \in [0, 2\pi]$, we have that:

  $$I_2 = \frac{1}{2}\int_0^{2\pi} \frac{1}{a+\cos(\theta)}\mathrm{d}\theta = \frac{1}{2}\int_0^{2\pi} \frac{2(e^{i\theta})}{(e^{i\theta})^2+2a(e^{i\theta})+1}\mathrm{d}\theta = -i\int_\gamma \frac{1}{z^2+2az+1}\mathrm{d}z.$$

  Next, we can apply the quadratic formula to get that $z^2+2az+1 = (z-\alpha)(z-\beta)$ where $\alpha = -a+\sqrt{a^2-1}$ and $\beta = -a-\sqrt{a^2-1}$.

  Since $|a| > 1$ and $\sqrt{a^2-1}$ has the same sign as a, we know $|\beta| > 1$. Meanwhile, note that $h(x) = -x+\sqrt{x^2-1} < 0$ and $h'(x) = -1 + \frac{x}{\sqrt{x^2-1}} > 0$ when $x > 1$. Therefore, as h is strictly increasing on $(1,\infty)$ and $h(1) = -1$, we can conclude that $|\alpha| = |h(a)| < 1$

  Therefore, by the Residue theorem we have that:

  $$\int_\gamma \frac{1}{z^2+2az+1}\mathrm{d}z = 2\pi i(1\cdot\mathrm{Res}(f;\alpha)) + 2\pi i(0\cdot\mathrm{Res}(f;\beta)) = 2\pi i\mathrm{Res}(f;\alpha).$$

  In other words, $I_2 = 2\pi\mathrm{Res}(f;\alpha)$. Finally, by proposition V.2.4 we can calculate that:

  $$\mathrm{Res}(f;\alpha) = \frac{1}{\alpha-\beta} = \frac{1}{(-a+\sqrt{a^2-1})-(-a-\sqrt{a^2+1})} = \frac{1}{2\sqrt{a^2-1}}.$$

  Therefore, $\boxed{\int_0^\pi \frac{1}{a+\cos(x)}\mathrm{d}x = I_2 = \frac{2\pi}{\sqrt{a^2-1}} = \frac{\pi}{\sqrt{a^2-1}}.}$ $\blacksquare$

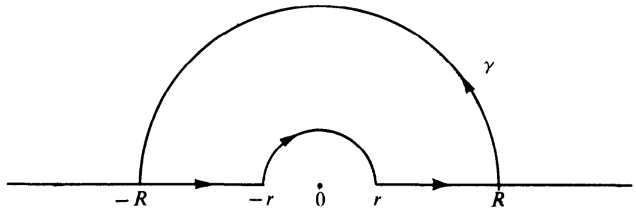
- Let $I_3 = \int_0^\infty \frac{\log(x)}{1+x^2}\mathrm{d}x$. Then note that this integral is well-defined. After all, $\int_0^\infty |\frac{\log(x)}{1+x^2}|\mathrm{d}x \leq \int_0^1 -\log(x)\mathrm{d}x + \int_1^\infty \frac{1}{1+x^2}\mathrm{d}x$ and the latter two integrals are both finite.

  To solve for I_3 let G be the region $\{z \in \mathbb{C} : z \neq 0 \text{ and } -\pi/2 < \arg(z) < 3\pi/2\}$. Then if $z = |z|e^{i\theta} \in G$ (where $-\pi/2 < \arg(z) < 3\pi/2$) we define:

  $$\ell(z) = \log|z| + i\theta.$$

  Then $\ell(x) = \log(x)$ for all $x > 0$. So, we can say that $I_3 = \int_0^\infty \frac{\ell(x)}{1+x^2}\mathrm{d}x$.

Now, let $0 < r < R < \infty$ and define the closed path γ by concatenating the paths $[r, R]$, Re^{it} (for $t \in [0, \pi]$), $[-R, -r]$, and $re^{i(\pi - t)}$ (for $t \in [0, \pi]$). Here is a drawing that I'm shamelessly plagiarizing from Conway.



Therefore:

$$\int_\gamma \frac{\ell(z)}{1+z^2}\mathrm{d}z = \int_r^R \frac{\log(x)}{1+x^2}\mathrm{d}x + iR\int_0^\pi \frac{\log(R)+i\theta}{1+R^2e^{2i\theta}}e^{i\theta}\mathrm{d}\theta + \int_{-R}^{-r} \frac{\log|x|+\pi i}{1+x^2}\mathrm{d}x + ir\int_\pi^0 \frac{\log(r)+i\theta}{1+r^2e^{2i\theta}}e^{i\theta}\mathrm{d}\theta.$$

If $r < 1$ and $R > 1$, then the only singularity of $\frac{\ell(z)}{1+z^2}$ inside the curve is $z = i$. And since that singularity is a simple pole we can calculate via proposition V.2.4 that $\mathrm{Res}(\frac{\ell(z)}{1+z^2}; i) = [\frac{\ell(z)}{z+i}](i) = \frac{\ell(i)}{i+i} = \frac{\log|i| + \pi i/2}{2i} = \frac{\pi}{4}$. In turn, by the residue theorem we have that $\frac{\pi^2 i}{2} = \int_\gamma \frac{\ell(z)}{1+z^2}\mathrm{d}z$.

Next note that $\int_r^R \frac{\log(x)}{1+x^2}\mathrm{d}x + \int_{-R}^{-r} \frac{\log|x|+\pi i}{1+x^2}\mathrm{d}x = 2\int_r^R \frac{\log(x)}{1+x^2}\mathrm{d}x + \pi i \int_r^R \frac{1}{1+x^2}\mathrm{d}x$. It follows (without mattering how we take $r \to 0$ and $R \to \infty$) that:

$$\lim_{\substack{r\to 0^+ \\ R\to\infty}} \left(\int_r^R \frac{\log(x)}{1+x^2}\mathrm{d}x + \int_{-R}^{-r} \frac{\log|x|+\pi i}{1+x^2}\mathrm{d}x\right) = 2I_3 + \pi i \lim_{\substack{r\to 0^+ \\ R\to\infty}} \int_r^R \frac{1}{1+x^2}\mathrm{d}x = 2I_3 + \frac{\pi^2 i}{2}.$$

Yet we must also have that:

$$\lim_{\substack{r\to 0^+ \\ R\to\infty}} \left(\int_r^R \frac{\log(x)}{1+x^2}\mathrm{d}x + \int_{-R}^{-r} \frac{\log|x|+\pi i}{1+x^2}\mathrm{d}x\right)$$

$$= \lim_{\substack{r\to 0^+ \\ R\to\infty}} \left(\frac{i\pi^2}{2} - iR\int_0^\pi \frac{\log(R)+i\theta}{1+R^2e^{2i\theta}}e^{i\theta}\mathrm{d}\theta - ir\int_\pi^0 \frac{\log(r)+i\theta}{1+r^2e^{2i\theta}}e^{i\theta}\mathrm{d}\theta\right).$$

Therefore, we can conclude that:

$$I_3 = \frac{i}{2}\left(\lim_{r\to 0^+} r\int_0^\pi \frac{\log(r)+i\theta}{1+r^2e^{2i\theta}}e^{i\theta}\mathrm{d}\theta - \lim_{R\to\infty} R\int_0^\pi \frac{\log(R)+i\theta}{1+R^2e^{2i\theta}}e^{i\theta}\mathrm{d}\theta\right)$$

At last, we show that both of those limits are 0. Suppose $\rho > 0$ with $\rho \neq 1$. Then:

$$\left|\rho\int_0^\pi \frac{\log(\rho)+i\theta}{1+\rho^2e^{2i\theta}}e^{i\theta}\mathrm{d}\theta\right| \leq \rho\int_0^\pi \left|\frac{\log(\rho)+i\theta}{1+\rho^2e^{2i\theta}}\right|\mathrm{d}\theta \leq \rho\int_0^\pi \frac{|\log(\rho)|}{|1-\rho^2|}\mathrm{d}\theta + \rho\int_0^\pi \frac{\theta}{|1-\rho^2|}\mathrm{d}\theta$$

$$= \frac{\rho\pi|\log(\rho)|}{|1-\rho^2|} + \frac{\rho\pi^2}{2|1-\rho^2|}$$

Now it's easy to see that $\frac{\rho\pi^2}{2|1-\rho^2|} \to 0$ if either $\rho \to 0^+$ or $\rho \to \infty$. Meanwhile, note that:

- $\lim_{\rho\to 0^+} \frac{\rho\pi|\log(\rho)|}{|1-\rho^2|} = \lim_{\rho\to 0^+} \frac{-\rho\pi\log(\rho)}{1-\rho^2} = -\pi \lim_{\rho\to 0^+} \frac{\log(\rho)}{\frac{1-\rho^2}{\rho}} = -\pi \lim_{\rho\to 0^+} \frac{\rho^{-1}}{\frac{-1-\rho^2}{\rho^2}} = -\pi \lim_{\rho\to 0^+} \frac{\rho}{-1-\rho^2} = 0.$

- $\lim_{\rho\to\infty} \frac{\rho\pi|\log(\rho)|}{|1-\rho^2|} = \lim_{\rho\to\infty} \frac{\rho\pi\log(\rho)}{\rho^2-1} = \pi \lim_{\rho\to\infty} \frac{\log(\rho)+1}{2\rho} = \pi \lim_{\rho\to\infty} \frac{\rho^{-1}}{2} = 0.$

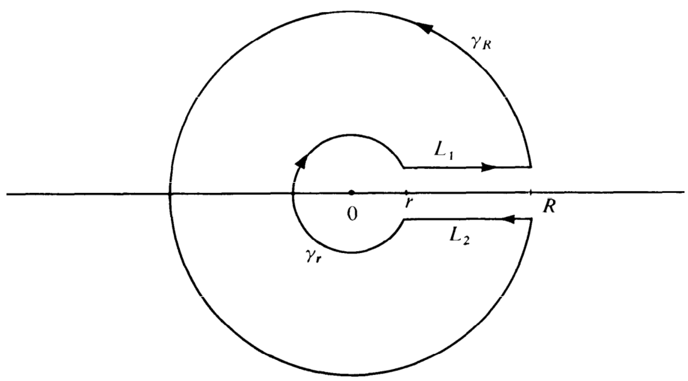
So, we've proven that $\boxed{I_3 = 0.}$ ■

- Suppose $0 < c < 1$ and let $I_4 = \int_0^\infty \frac{x^{-c}}{1+x}\mathrm{d}x$. Note that:
$$\int_0^\infty |\tfrac{x^{-c}}{1+x}|\mathrm{d}x \leq \int_0^1 x^{-c}\mathrm{d}x + \int_1^\infty x^{-(1+c)}\mathrm{d}x < \infty.$$

So, I_4 is a well-defined integral.

The following technique is called integration about a branch point. Let $G = \{z : z \neq 0 \text{ and } 0 < \arg(z) < 2\pi\}$ and define a branch of the logarithm on G by $\ell(re^{i\theta}) = \log(r) + i\theta$ where $0 < \theta < 2\pi$. Then for $z \in G$ put $f(z) = \exp(-c \cdot \ell(z))$ (i.e. f is a branch of z^{-c}).

Fix $0 < r < 1 < R < \infty$ and $\delta > 0$. Then let $L_1 = [r + \delta i, R + \delta i]$ and $L_2 = [R - \delta i, r - \delta i]$, and connect those paths together with the curves γ_r, γ_R defined as arcs of the circles $|z| = r$ and $|z| = R$ respectively. To visualize this, here is another drawing I'm shamelessly plagiarizing from Conway.



Now the only singularity of $f(z)(1+z)^{-1}$ in G is $z = -1$. Since this singularity is a simple pole, we can calculate by proposition V.2.4 that:
$$\mathrm{Res}(f(z)(1+z)^{-1}; -1) = f(-1) = e^{-\pi ci}.$$

It thus follows that:
$$\int_{L_1} \frac{f(z)}{1+z}\mathrm{d}z + \int_{\gamma_R} \frac{f(z)}{1+z}\mathrm{d}z + \int_{L_2} \frac{f(z)}{1+z}\mathrm{d}z + \int_{\gamma_r} \frac{f(z)}{1+z}\mathrm{d}z = 2\pi i e^{-\pi ci}.$$

Next note that $\int_{L_1} \frac{f(z)}{1+z}\mathrm{d}z = \int_0^1 \frac{f(r+t(R-r)+i\delta)}{1+r+t(R-r)+i\delta}(R-r)\mathrm{d}t = \int_r^R \frac{f(t+i\delta)}{1+t+i\delta}\mathrm{d}t.$

Now define g on the compact set $[r, R] \times [0, \frac{\pi}{2}]$ by $g(t, \delta) := \frac{f(t+i\delta)}{1+t+i\delta}$ when $\delta > 0$ and $g(t, 0) := \frac{t^{-c}}{1+t}$. By swapping which branch of the logarithm we are using to define f, we can easily see that g is a continuous function. Hence g is uniformly continuous and in particular we know that $\frac{f(t+i\delta)}{1+t+i\delta} \to \frac{t^{-c}}{1+t}$ uniformly as $\delta \to 0$. Consequently:
$$\int_r^R \frac{t^{-c}}{1+t}\mathrm{d}t = \lim_{\delta\to 0^+} \int_{L_1} \frac{f(z)}{1+z}\mathrm{d}z.$$

By similar reasoning, we can show that:

$$\int_{L_2} \frac{f(z)}{1+z}\mathrm{d}z = \int_R^r \frac{f(t-i\delta)}{1+t-i\delta}\mathrm{d}t \to \int_R^r \frac{e^{-2\pi ci}t^{-c}}{1+t}\mathrm{d}t \text{ as } \delta \to 0^+.$$

Therefore $2\pi i e^{-\pi ci} - (1-e^{-2\pi ic})\int_r^R \frac{t^{-c}}{1+t}\mathrm{d}t = \lim_{\delta\to 0^+}\left(\int_{\gamma_R} \frac{f(z)}{1+z}\mathrm{d}z + \int_{\gamma_r} \frac{f(z)}{1+z}\mathrm{d}z\right).$

Also, as $\delta \to 0^+$ both of the contour integrals in the limit above approach being contour integrals over a full circle as opposed to an arc. So:

$$\lim_{\delta\to 0^+}\left(\int_{\gamma_R} \frac{f(z)}{1+z}\mathrm{d}z + \int_{\gamma_r} \frac{f(z)}{1+z}\mathrm{d}z\right) = \int_0^{2\pi} \frac{\exp(-c\log(R)-ci\theta)}{1+Re^{i\theta}} Rie^{i\theta}\mathrm{d}z - \int_0^{2\pi} \frac{\exp(-c\log(r)-ci\theta)}{1+re^{i\theta}} rie^{i\theta}\mathrm{d}z$$

Now note that if $\rho > 0$ and $\rho \neq 1$, then:

$$\left|\int_0^{2\pi} \frac{\exp(-c\log(\rho)-ci\theta)}{1+\rho e^{i\theta}} \rho i e^{i\theta}\mathrm{d}z\right| \leq \int_0^{2\pi} \frac{\rho^{-c}}{|1-\rho|}\cdot \rho \mathrm{d}\theta = \frac{2\pi\rho^{-c+1}}{|1-\rho|}$$

The latter expression goes to 0 as $\rho \to 0^+$ or as $\rho \to \infty$ because $0 < -c+1 < 1$. Therefore, by taking $r \to 0^+$ and $R \to \infty$ we get that:

$$2\pi i e^{-\pi ci} - (1-e^{-2\pi ic})I_4 = 0$$

In other words, $\boxed{I_4 = \int_0^\infty \frac{x^{-c}}{1+x}\mathrm{d}x = \frac{2\pi i e^{-\pi ci}}{1-e^{-2\pi ci}} = \frac{2\pi}{e^{\pi ci}-e^{-\pi ci}} = \frac{\pi}{\sin(\pi c)}.}$ ■

- $I_5 = \int_0^\infty \frac{x^2}{x^4+x^2+1}\mathrm{d}x.$

  Since $\frac{x^2}{x^4+x^2+1}$ is a symmetric function we can say that $I_5 = \frac{1}{2}\int_{-\infty}^\infty \frac{x^2}{x^4+x^2+1}\mathrm{d}x$. Next, note that if $x^4+x^2+1 = 0$ then $x^2 = e^{2\pi i/3}$ or $e^{4\pi i/3}$. In turn, the roots of x^4+x^2+1 are $e^{\pi i/3}$, $e^{2\pi i/3}$, $-e^{\pi i/3}$, $-e^{2\pi i/3}$. Hence, if we let $\gamma_r(t) = re^{it}$ for $t \in [0,\pi]$, then we have that:

  $$\int_{-r}^r \frac{x^2}{x^4+x^2+1}\mathrm{d}x + \int_{\gamma_r} \frac{z^2}{z^4+z^2+1}\mathrm{d}z = 2\pi i\cdot(\mathrm{Res}(\tfrac{z^2}{z^4+z^2+1}; e^{\pi i/3}) + \mathrm{Res}(\tfrac{z^2}{z^4+z^2+1}; e^{2\pi i/3}))$$

  Next we calculate the residues.

  - $\mathrm{Res}(\frac{z^2}{z^4+z^2+1}; e^{\pi i/3}) = \frac{(e^{\pi i/3})^2}{2e^{\pi i/3}(e^{\pi i/3}-e^{2\pi i/3})(e^{\pi i/3}+e^{2\pi i/3})}$

    $= \frac{e^{\pi i/3}}{2(e^{2\pi i/3}-e^{4\pi i/3})} = \frac{1}{2(e^{\pi i/3}+1)} = \frac{1}{2(\frac{3}{2}+i\frac{\sqrt{3}}{2})}$

    $= \frac{1}{3+i\sqrt{3}} = \frac{3-i\sqrt{3}}{12} = \frac{1}{4} - i\frac{\sqrt{3}}{12}$

  - $\mathrm{Res}(\frac{z^2}{z^4+z^2+1}; e^{2\pi i/3}) = \frac{(e^{2\pi i/3})^2}{2e^{2\pi i/3}(e^{2\pi i/3}-e^{\pi i/3})(e^{2\pi i/3}+e^{\pi i/3})}$

    $= \frac{(e^{2\pi i/3})}{2(e^{4\pi i/3}-e^{2\pi i/3})} = \frac{1}{2(e^{2\pi i/3}-1)}$

    $= \frac{1}{2(\frac{-3}{2}+i\frac{\sqrt{3}}{2})} = \frac{1}{-3+i\sqrt{3}} = \frac{-3-i\sqrt{3}}{12}$

    $= \frac{-1}{4} - i\frac{\sqrt{3}}{12}$

  Therefore, $\int_{-r}^r \frac{x^2}{x^4+x^2+1}\mathrm{d}x + \int_{\gamma_r} \frac{z^2}{z^4+z^2+1}\mathrm{d}z = 2\pi i \cdot -i\frac{\sqrt{3}}{6} = \pi\frac{\sqrt{3}}{3}.$

Finally we now have that $2I_5 = \pi\frac{\sqrt{3}}{3} - \lim\limits_{r\to\infty} ir\int_0^\pi \frac{r^2e^{i2t}}{r^4e^{i4t}+r^2e^{i2t}+1}e^{it}\mathrm{d}t$. But note that:

$$\left|ir\int_0^\pi \frac{r^2e^{i2t}}{r^4e^{i4t}+r^2e^{i2t}+1}e^{it}\mathrm{d}t\right| \leq r^3\int_0^\pi \frac{1}{|r^4e^{i4t}+r^2e^{i2t}+1|}\mathrm{d}t \leq \frac{\pi r^3}{r^4-r^2-1} \to 0 \text{ as } r\to\infty$$

Therefore, we've proven that $\boxed{I_5 = \int_0^\infty \frac{x^2}{x^4+x^2+1}\mathrm{d}x = \frac{1}{2}\cdot\pi\frac{\sqrt{3}}{3} = \frac{\pi\sqrt{3}}{6}.}$ ■

Before I go home and take a nap before the final grind, here is another few exercises.

Exercise V.1.15: Let f be analytic in $G = \{z : 0 < |z-a| < r\}$ except that there is a sequence of poles $\{a_n\}_{n\in\mathbb{N}}$ in G with $a_n \to a$. Show that for any $\omega \in \mathbb{C}$ there is a sequence $\{z_n\}$ in G with $a = \lim_{n\to\infty} z_n$ and $\omega = \lim_{n\to\infty} f(z_n)$.

Without loss of generality we can assume $\omega = 0$. After all, if $\omega \neq 0$ then we can just repeat all the following logic but with the function $f(z) - \omega$.

Suppose that there exists $\rho > 0$ and $\delta > 0$ such that $|f(z)| > \delta$ for any $z \in A(a;0,\rho)$. Then by defining $g(z) = 1/f(z)$ when z is not a pole and $g(a_n) = 0$ for all $n \in \mathbb{N}$, we have that g is an continuous function $A(a;0,\rho)$. Furthermore, g is certainly analytic outside it's zero set. And since all the poles of f and in turn the zeros of g are isolated points, we can actually conclude g is analytic everywhere on $A(a;0,\rho)$.

As a reminder from my notes from last spring, if h is continuous on $B_r(a_k)$ and complex differentiable on $B_R(a_k) - \{a_k\}$, then we actually have that h is complex differentiable on $B_R(a_k)$. This is also easy to prove now though by just noting that technically h has a singularity at a_k which we trivially know is removable.

Now since a is an isolated singularity of g, we know it must be either removable, a pole, or essential. That said, we can show as follows that none of those three possibilities work.

- g can't have a removable singularity at a since that would contradict that the zeros of an analytic function can't have a limit points in the function's domain.
- g also can't have a pole at a since that would require that $\lim_{z\to a} g(z) = \infty$ and yet we know that $g(a_n) \to 0$ as $a_n \to a$.
- Thirdly, g can't have an essential singularity because we know that $g(z) < \delta^{-1}$ for all $z \in A(a;0,\rho)$.

Thus, we've arrived at a contradiction and can conclude that no such ρ and δ exist. ■

Exercise V.2.4: Suppose that f has a simple pole at $z = a$ and let g be analytic in an open set containing a. Show that $\mathrm{Res}(fg;a) = g(a)\mathrm{Res}(f;a)$.

Express $f(z) = \sum_{n=-1}^\infty a_n(z-a)^n$ and $g(z) = \sum_{m=0}^\infty b_m(z-a)^m$. Then note that for some $r > 0$ we have that both sums converge absolutely on $A(a;0,r)$. Therefore, by Merten's theorem (see my math 140a notes) we have for all $z \in A(a;0,r)$ that:

$$\begin{aligned} f(z)g(z) &= \left(\sum_{m=0}^\infty a_{m-1}(z-a)^{m-1}\right)\left(\sum_{m=0}^\infty b_m(z-a)^m\right) \\ &= \sum_{m=0}^\infty \left(\sum_{k=0}^m a_{k-1}(z-a)^{k-1}b_{m-k}(z-a)^{m-k}\right) = \sum_{m=0}^\infty\left(\sum_{k=0}^m a_{k-1}b_{m-k}\right)(z-a)^{m-1}. \end{aligned}$$

In particular, this shows us that $\mathrm{Res}(fg;a) = a_{-1}b_0 = \mathrm{Res}(f;a)g(a)$. ■

If G is open and f is a function defined and analytic in G except for at poles, then we say f is a meromorphic function on G.

Recall from *page 426* that if a function f has zeros $a_1, \ldots, a_n$ (where the a_k repeat according to the multiplicity of the zeros), then we have that:

$$\frac{f'(z)}{f(z)} = \frac{g'(z)}{g(z)} + \sum_{k=1}^{n}(z-a_k)^{-1}$$

(where g satisfies that $f(z) = \left(\prod_{k=1}^{n}(z-a_k)\right)g(z)$ and g has no zeros).

Meanwhile, suppose f has a pole of order m at $z = a$. In other words:

$f(z) = (z-a)^{-m}g(z)$ where g is analytic and $g(a) \neq 0$.

Then $f'(z) = -m(z-a)^{-m-1}g(z) + (z-a)^{-m}g'(z)$. Hence after dividing by $f(z) = (z-a)^{-m}g(z)$ we get that:

$$\frac{f'(z)}{f(z)} = \frac{-m}{(z-a)} + \frac{g'(z)}{g(z)}$$

By repeating the prior reasoning as many times as necessary, we can say that f is meromorphic in G with poles $p_1, \ldots p_m$ and zeros $a_1, \ldots, a_n$ (both lists repeating according to multiplicity), then there exists a nonzero analytic function g on G such that:

$$\frac{f'(z)}{f(z)} = \frac{g'(z)}{g(z)} + \sum_{k=1}^{n}(z-a_k)^{-1} - \sum_{j=1}^{m}(z-p_j)^{-1}$$

The Argument Principle: Let f be meromorphic in G with poles $p_1, \ldots p_m$ and zeros $a_1, \ldots, a_n$ (both lists repeating according to multiplicity). If γ is a closed piecewise C^1 curve in G with $\gamma \approx_G 0$ and γ not passing through any of the poles or zeros of f, then:

$$\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)}\mathrm{d}z = \sum_{k=1}^{n} n(\gamma; a_k) - \sum_{j=1}^{m} n(\gamma; p_j).$$

Proof:
Write $\frac{f'(z)}{f(z)} = \frac{g'(z)}{g(z)} + \sum_{k=1}^{n}(z-a_k)^{-1} - \sum_{j=1}^{m}(z-p_j)^{-1}$.

By Cauchy's theorem we know that $\int_\gamma \frac{g'(z)}{g(z)}\mathrm{d}z = 0$. Hence, the rest of proof is obvious. ■

Side note: Why is this equation called the argument principle?

Suppose it were possible to define $\log(f(z))$. Then as $\log(f(z))$ would be a primitive for $f'(z)/f(z)$, we'd have that as z goes around γ we have that $\log(f(z))$ would change by $2\pi i K$ where:

$$K = \sum_{k=1}^{n}(z-a_k)^{-1} - \sum_{j=1}^{m}(z-p_j)^{-1}$$

This would then imply that $\mathrm{Im}(\log(f(z))) = \arg(f(z))$ changes by $2\pi K$. Or to put that in other words, the piecewise C^1 curve $f \circ \gamma$ would satisfy that $n(f \circ \gamma; 0) = K$.

That all said, there almost certainly does not exist a branch of $\log(f(z))$ on $G - \bigcup\{a_k\} - \bigcup\{p_j\}$. After all, the latter set is not simply connected (as will be proven later on *page* ____). Hence, we can't apply our theorem on *page 416*, Also, if such a branch $\ell(z)$ of $\log(f(z))$ did exist, then we'd always have that $\int_\gamma f'(z)/f(z) = 0$.

We can fix our reasoning to still show that $\arg(f(z))$ changes by $2\pi K$. To do this, for each $w \in \{\gamma\}$ pick $\varepsilon_w > 0$ such that:

$$B_{\varepsilon_w}(w) \subseteq G - \bigcup\{a_k\} - \bigcup\{p_j\}.$$

Then by the compactness of $\{\gamma\}$ we can find a single $\varepsilon > 0$ such that $B_\varepsilon(w) \subseteq G - \bigcup\{a_k\} - \bigcup\{p_j\}$ for any $w \in \{\gamma\}$. Furthermore, if the domain of γ is $[a, b]$ then we can use the uniform continuity of γ to get a partition $a = t_0 < t_1 < \ldots < t_k = b$ such that $\gamma(t) \in B_\varepsilon(\gamma(t_{j-1}))$ for all $t \in [t_{j-1}, t_j]$ and $1 \leq j \leq k$.

Now it is possible via the theorem on *page 416* to construct a branch ℓ_j of $\log(f(z))$ on $B_\varepsilon(\gamma(t_j))$ for each $1 \leq j \leq k$. Furthermore, since the jth and $(j+1)$th balls both contain $\gamma(t_j)$, we can choose $\ell_1, \ldots, \ell_k$ such that $\ell_j(\gamma(t_j)) = \ell_{j+1}(\gamma(t_j))$ for all $1 \leq j < k$.

Letting γ_j equal the function γ with it's domain restricted to $[t_{j-1}, t_j]$, we have that:

$$2\pi i K = \int_\gamma \tfrac{f'}{f} = \sum_{j=1}^{k} \int_{\gamma_j} \tfrac{f'}{f} = \sum_{j=1}^{k} \left(\ell_j(\gamma(t_j)) - \ell_j(\gamma(t_{j-1}))\right)$$

In turn, we have that:

$$\begin{aligned}2\pi K &= \sum_{j=1}^{k} \left(\mathrm{Im}(\ell_j(\gamma(t_j))) - \mathrm{Im}(\ell_j(\gamma(t_{j-1})))\right) \\ &= \sum_{j=1}^{k} \left(\arg(f \circ \gamma(t_j)) - \arg(f \circ \gamma(t_{j-1}))\right) = \arg(f \circ \gamma(b)) - \arg(f \circ \gamma(a)).\end{aligned}$$

One other note I want to make is that this theorem is purely a generalization of *theorem IV.7.2 on pages 426-427*.

<u>(Conway) Theorem V.3.6:</u> Let f be meromorphic in G with poles $p_1, \ldots p_m$ and zeros $a_1, \ldots, a_n$ (both lists repeating according to multiplicity). If g is analytic in G and γ is a piecewise C^1 curve in G with $\gamma \approx_G 0$ and γ not passing through any a_k or p_j, then:

$$\frac{1}{2\pi i} \int_\gamma g(z) \frac{f'(z)}{f(z)} \mathrm{d}z = \sum_{k=1}^{n} g(a_k) n(\gamma; a_k) - \sum_{j=1}^{m} g(p_j) n(\gamma; p_j).$$

Proof:
Like before we know there exists $h : G \to \mathbb{C}$ that is analytic on G with no zeros or poles and satisfies that:

$$\frac{f'(z)}{f(z)} = \frac{h'(z)}{h(z)} + \sum_{k=1}^{n} (z - a_k)^{-1} - \sum_{j=1}^{m} (z - p_j)^{-1}.$$

In turn, since $g(z)\frac{h'(z)}{h(z)}$ is analytic on G we can conclude by Cauchy's theorem and Cauchy's integral formula that:

$$\frac{1}{2\pi i}\int_\gamma g(z)\frac{f'(z)}{f(z)}\mathrm{d}z = \frac{1}{2\pi i}\left(\sum_{k=1}^{n}\int_\gamma \frac{g(z)}{z-a_k}\mathrm{d}z - \sum_{j=1}^{m}\int_\gamma \frac{g(z)}{z-p_j}\mathrm{d}z\right)$$
$$= \sum_{k=1}^{n} g(a_k)n(\gamma;a_k) - \sum_{j=1}^{m} g(p_j)n(\gamma;p_j). \blacksquare$$

We already know from the open mapping theorem that if f is an injective analytic function on G then f has an analytic inverse on $f(G)$. [See *(Conway) proposition 2.20 on page 331* as well as the *open mapping theorem on page 429*.]

I guess I technically need to to explain why f' is nonzero (so that we can apply proposition 2.20). Suppose $a \in G$ and $\beta = f(a)$ and $f'(a) = 0$. Then we'd have that $f(z) - \beta$ has a zero at $z = a$ with multiplicity ≥ 2. But then by *Theorem IV.7.4 on page 428*, we know that f can't be injective on some small neighborhood of a in G. This contradicts our initial premise.

Using theorem V.3.6 we now give a formula for the inverse of f.

(Conway) Proposition V.3.7: Let f be analytic on an open set containing $\overline{B_R(a)}$ and suppose that f is injective on $B_R(a)$. If $\Omega = f(B_R(a))$ and $\gamma(t) = a + Re^{it}$ for $t \in [0, 2\pi]$ then $f^{-1}(\omega)$ is defined for each $\omega \in \Omega$ by the formula:

$$f^{-1}(\omega) = \frac{1}{2\pi i}\int_\gamma \frac{zf'(z)}{f(z)-\omega}\mathrm{d}z$$

Proof:
Given any $\omega \in \Omega$ there is a unique $\xi \in B_R(a)$ such that $f(\xi) = \omega$. Then we know that the only zero of $f(z) - \omega$ in $B_R(a)$ is ξ. Futhermore, we know that $f(z) - \omega$ has no poles on $B_r(a)$. Therefore, after letting $g(z) = z$ and applying theorem V.3.6 we have that:

$$\frac{1}{2\pi i}\int_\gamma \frac{zf'(z)}{f(z)-\omega}\mathrm{d}z = \xi \cdot n(\gamma;\xi) = \xi \cdot 1 = f^{-1}(\omega). \blacksquare$$

Here is the last theorem in this class.

Rouché's Theorem: Suppose that f and g are meromorphic in a neighborhood of $\overline{B_R(a)}$ with no zeros or poles on the circle $\gamma(t) = a + Re^{it}$ (where $t \in [0, 2\pi]$). If Z_f, Z_g are the number of zeros of f and g inside γ and P_f, P_g are the number of poles of f and g inside γ (each counted according to their multiplicites) and if $|f(z) + g(z)| < |f(z)| + |g(z)|$ on $\{\gamma\}$ then $Z_f - P_f = Z_g - P_g$.

Proof:
To start off, by our hypothesis we know that $\left|\frac{f(z)}{g(z)} + 1\right| = \frac{|f(z)+g(z)|}{|g(z)|} < \left|\frac{f(z)}{g(z)}\right| + 1$ on $\{\gamma\}$.
Now set $\lambda = f(z)/g(z)$ and note that if $\lambda(z)$ is ever a nonnegative real number, then this inequality becomes $\lambda + 1 < \lambda + 1$ (which is a contradiction). Hence, λ must map $\{\gamma\}$ onto $\Omega = \mathbb{C} - [0, \infty)$.

Now let ℓ be a branch of the logarithm on Ω. Then $\ell(f(z)/g(z))$ is a well-defined primitive for $\frac{(f/g)'}{(f/g)}$ in a neighborhood of $\{\gamma\}$.

But also note that $\frac{(f/g)'}{(f/g)} = \frac{\frac{f'g-fg'}{g^2}}{\frac{f}{g}} = \frac{f'g-fg'}{fg} = \frac{f'}{f} - \frac{g'}{g}$. Therefore, as $n(\gamma;a) = 1$ for each pole or zero a of g or f, we can conclude by the argument principle that:

$$0 = \frac{1}{2\pi i}\int_\gamma \frac{(f/g)'}{(f/g)} = \frac{1}{2\pi i}\int_\gamma \frac{f'}{f} - \frac{1}{2\pi i}\int_\gamma \frac{g'}{g} = (Z_f - P_f) - (Z_g - P_g). \blacksquare$$

Rouché's theorem also gives the following alternate proof of the fundamental theorem of algebra:

Suppose $p(z) = z^n + a_1 z^{n-1} + \ldots + a_n$ and assume $a_n \neq 0$. Then we know that $\frac{p(z)}{z^n}$ approaches 1 as $|z| \to \infty$. Hence, there is a sufficiently large $R > 0$ such that:

$$\left|\frac{p(z)}{z^n} - 1\right| < 1 \leq \left|\frac{p(z)}{z^n}\right| + |-1| \text{ whenever } |z| = R.$$

Then since the constant function -1 has no zeros or poles, we must have by Rouché's theorem that $\frac{p(z)}{z^n}$ has the same number of zeros and poles. In particular, since $\frac{p(z)}{z^n}$ has n poles, we must have that $\frac{p(z)}{z^n}$ has n zeros.

Meanwhile, if $a_n = 0$ then we just factor out the zero of p at $z = 0$ until we are left with a polynomial whose constant term is nonzero. ■

Finally, note that the proof Rouche's theorem mostly works if γ is replaced with any other curve homotopic to 0 in $\overline{B_R(a)}$. The only difficulty is that we might not have that $\frac{1}{2\pi i}\int_\gamma \frac{f'}{f} - \frac{1}{2\pi i}\int_\gamma \frac{g'}{g} = (Z_f - P_f) - (Z_g - P_g)$ if $n(\gamma;a) \neq 1$ for some zero or pole of f or g.

Exercise V.3.10 (A variation on Brouwer's fixed point theorem): Let $D = \overline{B_1(0)}$ and suppose f is analytic on some neighborhood of D.

1. Suppose $|f(z)| \leq 1$ with $f(z) \neq z$ when $|z| = 1$. Then there exists a unique w with $|w| < 1$ such that $f(w) = w$.

   Let $g(z) = f(z) - z$ and $h(z) = z$. By assumption we know g and h have no zeros on $\{z : |z| = 1\}$. Also, clearly g and h have no poles on D. Thirdly:

   $$|g(z) + h(z)| = |f(z)| \leq 1 = |h(z)| < |g(z)| + |h(z)| \text{ for all } z \text{ with } |z| = 1.$$

   Thus we can conclude by Rouché's theorem that $f(z) - z$ and z have the same number of zeros in D°. Or in other words, $f(z) - z$ has only one zero in D°. ■

2. Corollary: If $|f(z)| \leq 1$ when $|z| = 1$ and f doesn't have at least one fixed point in ∂D then we must have that f has a unique fixed point in D°.

12/8/2025

Before I switch to cramming for math 200a tomorrow, I thought I'd relax by doing some function analysis again.

Math 241a Notes:

Let $K = \mathbb{R}$ or $\mathbb{C}$. Then recall from *pages 442-443* that if $\mathcal{Y}$ is an n-dimensional Hausdorff topological K-vector space then any bijective linear map from K^n (equipped with the standard euclidean norm) to $\mathcal{Y}$ is a homeomorphism. Here are two quick corollaries from that:

- If V, W are both n-dimensional Hausdorff K-vector spaces and $T : V \to W$ is a bijective linear map then T is a homeomorphism.

  Why?
  Let $v_1, \ldots, v_n$ be a basis for V and set $u_i = T(v_i)$ for each i. Now, if $e_1, \ldots, e_n$ is the standard basis of K^n, then there exists bijective linear maps $S_1 : K^n \to V$ and $S_2 : K^n \to W$ such that $S_1(e_i) = v_i$ and $S_2(e_i) = u_i$ for each i. As both S_1 and S_2 are homeomorphisms and $T = S_2 \circ S_1^{-1}$, it follows that T is also a homeomorphism.

- If V is an n-dimensional Hausdorff topological K-vector space when equipped with the topology $\mathcal{T}$ or the topology $\mathcal{T}'$, we have that $\mathcal{T} = \mathcal{T}'$.

  Why?
  We know from the last bullet point that $\mathrm{Id} : (V, \mathcal{T}) \to (V, \mathcal{T}')$ is a homeomorphism. As a result, a set $U \subseteq V$ is open with respect to $\mathcal{T}$ iff it is open with respect to $\mathcal{T}'$.

Given the facts above, if V is any finite dimensional K-vector space then we often just assume V is equipped with it's unique Hausdorff vector space topology. To get this topology, we can employ the following simple construction:

Let $T : V \to K^n$ be any bijective linear map. Then let $\mathcal{T}$ be the weak topology:

$$\mathcal{T} = \{T^{-1}(U) : U \subseteq K^n \text{ is open}\}.$$

Recall from my math 240b homework that a net $\langle v_\alpha \rangle \in V$ converges to v with respect to $\mathcal{T}$ iff $T(v_\alpha) \to T(v)$. Using that fact it is easy to show that $(V, \mathcal{T})$ is a topological vecotr space.

- If $(v_\alpha, w_\alpha) \to (v, w)$ in $V \times V$ then $(T(v_\alpha), T(w_\alpha)) \to (T(v), T(w))$ in $K^n \times K^n$. Since K^n is a topological vecotr space, we thus know know that $T(v_\alpha + w_\alpha) = T(v_\alpha) + T(w_\alpha) \to T(v) + T(w) = T(v + w)$. Hence, $v_\alpha + w_\alpha \to v + w$.
- If $c_\alpha \to c$ in K and $v_\alpha \to v$ in V then we know $T(v_\alpha) \to T(v)$ in K^n. Hence since K^n is a topological vector space we know that $T(c_\alpha v_\alpha) = c_\alpha T(v_\alpha) \to cT(v) = T(cv)$. And this proves that $c_\alpha v_\alpha \to cv$.

Finally, to see that $(V, \mathcal{T})$ is Hausdorff it suffices to note that T separates points and K^n is Hausdorff. ■

Corollary 2.2.8: Let G be a compact group and V be a finite dimensional K-vector space. Let $\pi : G \to \mathrm{GL}(V)$ be a strong-operator-continuous representation. Then there is a (positive definite) inner product $\langle\ ,\ \rangle$ on V which is G-invariant (meaning that $\langle \pi(g)v, \pi(g)w\rangle = \langle v, w\rangle$ for all $v, w \in V$ and $g \in G$).

Proof:

Let $\langle\ ,\ \rangle_0$ be any (positive definite) inner product on V. Note that one certainly exists since we can construct one as follows:

Let $u_1, \ldots, u_n$ be a basis for V. Then if $v = \sum_{i=1}^n a_i u_i$ and $w = \sum_{i=1}^n b_i u_i$ where each a_i, b_i is in $\mathbb{R}$ or $\mathbb{C}$, we define $\langle v, w\rangle_0$ by taking the standard inner product of $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ in $\mathbb{R}^n$ or $\mathbb{C}^n$. This is easily seen to define an inner product on V.

Next, we define $\langle v, w\rangle := \int_G \langle \pi(g)v, \pi(g)w\rangle_0 \mathrm{d}\mu(g)$ where μ is a Haar measure on G. Note that $\langle \pi(g)v, \pi(g)w\rangle_0 \in L^1(\mu(g))$ for any $v, w \in V$ because of the following reasoning:

We can equivalently view V as having the inner product topology defined by $\langle\ ,\ \rangle_0$. Then it is clear that $\langle\ ,\ \rangle_0$ is continuous on $V \times V$. Also note that for any net $\langle g_i\rangle$ in G converging to g we have that $\pi(g_i)v \to \pi(g)v$ and $\pi(g_i)w \to \pi(g)w$ (since π is strong operator continuous). Therefore, $\langle \pi(g_i)v, \pi(g_i)w\rangle_0 \to \langle \pi(g)v, \pi(g)w\rangle_0$ as $g_i \to g$. Hence, we've proven that $g \mapsto \langle \pi(g)v, \pi(g)w\rangle_0$ is a continuous function from G to K.

Having proven continuity it's now obvious that $g \mapsto \langle \pi(g)v, \pi(g)w\rangle_0$ is measurable on G. Also, by extreme value theorem we know that this map from G to $\mathbb{C}$ is bounded by some constant $C > 0$. And as $\mu(G) < \infty$ it is clear that:

$$\int_G |\langle \pi(g)v, \pi(g)w\rangle_0| \,\mathrm{d}\mu \leq C\mu(G) < \infty.$$

You can now easily verify that $\langle\ ,\ \rangle$ is a positive definite inner product that is also G-invariant. ■

Recall that since G is a compact group we know that G is unimodular (see *page 433*) and thus μ is both a left and right Haar measure. I mention this because when proving that our inner product is in fact G-invariant, you use the right-invariance of μ.

12/9/2025

Ok. I have two days to cram as much algebra as possible!! My first order of business will be to finish taking notes on the week of class I missed. As a side note: for the rest of my notes for this class all rings will be assumed to be commutative and unital.

Math 200a Notes:

To start off, you can show by similar means to *problem 4 on the 9th problem* that $\mathbb{Z}[i]$ is a Euclidean domain. For details see my math 100b notes.

Lemma: If A is an integral domain and $a \in A$, then a being prime implies that a is irreducible.

Proof:

Suppose a is prime and $a = bc$. Then trivially $a \mid bc$ since $a = 1bc$. So, we know that either $a \mid b$ or $a \mid c$. Without loss of generality we'll assume $a \mid b$. Then we know that $\langle b \rangle \subseteq \langle a \rangle$. Yet since $a = bc$ we also know that $b \mid a$ and thus $\langle a \rangle \subseteq \langle b \rangle$. Hence, we've proven that $\langle a \rangle = \langle b \rangle$.

By a lemma on *page 458* we thus know that $bc = a = bu$ for some $u \in A^\times$. Finally, as $a \neq 0$ we know that $b \neq 0$ either. So, we can cancel it and get that $c = u$. ■

Theorem: Let D be an integral domain. The following two statements are equivalent:

1. Every irreducible element of D is prime and every nonzero-nonunit element in D can be factored into a product of irreducible elements.
2. D is a U.F.D.

$(1 \Longrightarrow 2)$

Suppose $p_1 \cdots p_n = q_1 \cdots q_m$ where each p_i and q_j is irreducible. Then it suffices to show that $m = n$ and that there is some $\sigma \in S_n$ such that $\langle p_i \rangle = \langle q_{\sigma(i)} \rangle$ for each i. So, we proceed by induction on n.

Firstly, suppose $u \in D^\times$ and $u = q_1 \cdots q_m = 1$. In turn q_m is a unit with inverse $u^{-1}q_1 \cdots q_{m-1}$. But this contradicts that q_m is irreducible. So, we can conclude that there is no product of irreducibles equaling a unit.

Next, suppose $n > 0$ and that $p_1 \cdots p_n = q_1 \cdots q_m$. Then we know that $p_n \mid q_1 \cdots q_m$. And because p_n is prime (since it is irreducible), we know that $p_n \mid q_i$ for some i. In other words, $\langle q_i \rangle \subseteq \langle p_n \rangle$. Yet because q_i is irreducible, we also know that $\langle q_i \rangle$ is maximal among principal proper ideals. In turn, because $\langle p_n \rangle$ is a principal proper ideal it must be that $\langle p_n \rangle = \langle q_i \rangle$ and so $q_i = p_n u'$ for some $u' \in D^\times$.

If $n = 1$, then after canceling out p_n we're left with $1 = u'q_1 \cdots q_{i-1}q_{i+1} \cdots q_m$. By the first observation I made, this is only possible if $m - 1 = 0$. So, $m = 1 = n$ and $p_1 = u'q_1$.

Meanwhile, if $n > 1$, then after canceling out p_n we're left with:

$$p_1 \cdots p_{n-1} = u'q_1 \cdots q_{i-1}q_{i+1} \cdots q_m.$$

Now the desired conclusion is clear by applying our induction hypothesis.

$(2 \Longrightarrow 1)$

It suffices to show that if p is irreducible then p is prime. So suppose $p \mid bc$. Then we know that $pa = bc$ for some $a \in D$. Also, we let $a = p_1 \cdots p_n$, $b = \ell_1 \cdots \ell_m$, and $c = q_1 \cdots q_k$ be factorizations of a, b, and c into products of irreducibles.

Now $p \cdot p_1 \cdots p_n = \ell_1 \cdots \ell_m q_1 \cdots q_k$. By the uniqueness of the factorization, we know that either $p = u\ell_i$ or $p = uq_i$ for some $u \in D^\times$ and integer i. And from there it's clear that p divides either b or c respectively. ■

Lemma: If D is a P.I.D. then all irreducible elements are prime.

Proof:

If a is irreducible then we know that $a \neq 0$ and that $\langle a \rangle$ is maximal among proper principal ideals. Yet because D is a P.I.D. we know all ideals are principal. Thus a is prime since $\langle a \rangle \in \mathrm{Max}(D) \subseteq \mathrm{Spec}(D)$. ■

Corollary: If D is a P.I.D. then $\mathrm{Spec}(D) = \{0\} \cup \mathrm{Max}(D)$.

Proof:

Suppose $\mathfrak{p} \in \mathrm{Spec}(D)$. Then $\mathfrak{p} = \langle a \rangle$ for some $a \in D$ since D is a P.I.D.. Without loss of generality we may assume $\mathfrak{p} \neq \{0\}$. Then as we also know that $\mathfrak{p} \neq D$ since $\mathfrak{p}$ is a prime, we must have that a is prime. In turn, a is irreducible and thus $\mathfrak{p}$ is maximal among all proper principal ideals. So all ideals are principal and we know $\mathfrak{p}$ is maximal. ■

Given a nonzero ring A, we define the Krull Dimension of that ring to be:

$$\dim(A) \coloneqq \sup\{n \in \mathbb{Z}_{\geq 0} : \exists \mathfrak{p}_0, \ldots, \mathfrak{p}_n \in \mathrm{Spec}(A) \text{ } s.t. \text{ } \mathfrak{p}_0 \subsetneqq \mathfrak{p}_1 \subsetneqq \cdots \subsetneqq \mathfrak{p}_n\}$$

Note that if F is an integral domain, then $\dim(F) = 0$ if and only if F is a field. Meanwhile, if D is an integral domain then we know $\dim(D) = 1$ iff $\mathrm{Spec}(D) = \{0\} \cup \mathrm{Max}(D)$ and $\{0\} \notin \mathrm{Max}(D)$. In particular, this means that if D is a P.I.D. then $\dim(D) \leq 1$ with $\dim(D) \neq 1$ iff D is a field.

Lemma: $\dim(A[x]) \geq \dim(A) + 1$.

Proof:

Suppose that $\mathfrak{p}_0 \subsetneqq \mathfrak{p}_1 \subsetneqq \cdots \subsetneqq \mathfrak{p}_n$ where each $\mathfrak{p}_j$ is a prime ideal in A. Then recall from part (b) of *problem 1 of the 9th problem set* that $\mathfrak{p}_j[x] \in \mathrm{Spec}(A[x])$ for each j and we clearly have that $\mathfrak{p}_0[x] \subsetneqq \mathfrak{p}_1[x] \subsetneqq \cdots \subsetneqq \mathfrak{p}_n[x]$. Our goal now is to find any proper ideal properly containing $\mathfrak{p}_n[x]$ That way, we know that $\mathfrak{p}_n[x]$ is not maximal and we can use the theorem on *pages 456-457* to find an ideal in $\mathrm{Max}(A[x]) \subseteq \mathrm{Spec}(A[x])$ which properly contains $\mathfrak{p}_n[x]$.

Consider the set $\widetilde{\mathfrak{p}} \coloneqq \{f(x) \in A[x] : f(0) \in \mathfrak{p}_n\}$. Importantly, $\widetilde{\mathfrak{p}}$ is an ideal in $A[x]$ since it is the kernel of the evaluation at zero homomorphism $A[x] \to A$ composed with the natural projection homomorphism $A \to A/\mathfrak{p}_n$. $\widetilde{\mathfrak{p}}$ is also a proper subset of $A[x]$ since it doesn't include any constant polynomial for which the constant isn't in $\mathfrak{p}_n$. Finally, if $a \in \mathfrak{p}_n$ and $b \in A - \mathfrak{p}_n$ then $bx + a \in \widetilde{\mathfrak{p}}$ and $bx + a \notin \mathfrak{p}_n[x]$. Hence, $\widetilde{\mathfrak{p}}$ is a proper ideal in $A[x]$ properly containing $\mathfrak{p}_n[x]$. ■.

Corollary: $A[x]$ is a P.I.D. if and only if A is a field.

Proof:

We already showed the $(\Longleftarrow)$ implication on *page 459* when we noted that $A[x]$ is an E.D. when A is a field. To show the other direction, note that if $A[x]$ is a P.I.D. then we must have that $\dim(A[x]) = 1$ and therefore $\dim(A) = 0$. Also, since $A[x]$ is a domain we know that A is a domain. So, because A is a domain with $\dim(A) = 0$, we know A is a field. ■

A ring A is called Noetherian if every chain $\mathcal{C}$ of ideals (simply ordered by $\subseteq$) has a maximal element inside $\mathcal{C}$.

Lemma: The following are equivalent:

(a) A is Noetherian,

(b) Every nonempty family of ideals in A has a maximal element,

(c) If we have a sequence $\{\mathfrak{a}_n\}_{n\in\mathbb{N}}$ of ideals in A such that $\mathfrak{a}_1 \subseteq \mathfrak{a}_2 \subseteq \cdots$ then there must exist some $N \in \mathbb{N}$ with $\mathfrak{a}_N = \mathfrak{a}_{N+1} = \cdots$.

(d) Every ideal in A is generated by a finite set.

$(a \implies b)$
By assumption we know every chain in that family has an upper bound. Hence, we can apply Zorn's lemma.

$(b \implies c)$
We know the family $\{\mathfrak{a}_n\}_{n\in\mathbb{N}}$ must have a maximal element, say $\mathfrak{a}_N$. Then as that ideal is maximal and $\mathfrak{a}_N \subseteq \mathfrak{a}_{N+k}$ for any $k \in \mathbb{N}$, we conclude that $\mathfrak{a}_N = \mathfrak{a}_{N+k}$ for all $k \in \mathbb{N}$.

$(c \implies d)$
Suppose to the contrary that an ideal $\mathfrak{a} \lhd A$ is not finitely generated. Then we can find by induction a sequence $(a_n)_{n\in\mathbb{N}}$ such that $\langle a_1 \rangle \subsetneqq \langle a_1, a_2 \rangle \subsetneqq \cdots \subseteq \mathfrak{a}$.

Specifically, for each $k \in \mathbb{N}$ pick $a_k \in \mathfrak{a} - \langle a_1, \ldots, a_{k-1} \rangle$. We can always do this because $\mathfrak{a}$ is not finitely generated.

But now we have a sequence of ideals contradicting (c).

$(d \implies a)$
Suppose $\mathcal{C}$ is a chain of ideals in A and let $\mathfrak{a} := \bigcup_{\mathfrak{b}\in\mathcal{C}} \mathfrak{b}$. By the lemma at the top of *page 455* we know $\mathfrak{a}$ is an ideal. Also, by part (c) we know that $\mathfrak{a}$ is finitely generated by elements $x_1, \ldots, x_n$. But now as $\mathcal{C}$ is simply ordered by inclusion and each x_i is in some $\mathfrak{b} \in \mathcal{C}$, we know there must contain some $\mathfrak{b}' \in \mathcal{C}$ containing all of the x_i. So, $\mathfrak{b}' \subseteq \mathfrak{a} = \langle x_1, \ldots, x_n \rangle \subseteq \mathfrak{b}'$. This proves that $\mathfrak{a} \in \mathcal{C}$. ■

By applying part (d) of the above lemma, we can clearly see that every P.I.D. is Noetherian.

Lemma: If D is Noetherian and an integral domain, then every nonzero-nonunit element of D can be factored into a product of irreducible elements.

Proof:
Suppose to the contrary that there exists $a \in D - (\{0\} \cup D^\times)$ such that a cannot be written as a product of irreducibles. Then let:

$\Sigma := \{\langle a \rangle : a \in D - (\{0\} \cup D^\times)$ can't be written as a product of irreducibles$\}$.

Since D is Noetherian, we know that Σ must contain a maximal element. Hence, there exists $a \in a \in D - (\{0\} \cup D^\times)$ such that a cannot be written as a product of irreducibles and if b also can't be written as a product of irreducibles and $\langle a \rangle \subseteq \langle b \rangle$ then we must have that $\langle a \rangle = \langle b \rangle$.

But note that a is not irreducible itself since the trivial product with just a in it is not a product of irreducibles. Thus, there must exist elements $b, c \in D - (\{0\} \cup D^{\times})$ such that $a = bc$. In turn, we know that $\langle a \rangle \subseteq \langle b \rangle$. At the same time, we can't have that both b or c are able to be written as a product of irreducibles since that would contradict that a can't be written as such a product. So without loss of generality we know that b can't be written as a product of irreducibles.

By the maximality of a we thus know that $\langle a \rangle = \langle b \rangle$. Hence, $bc = a = ub$ where $u \in D^{x}$. Finally, after canceling out b we get that $c = u$. This contradicts that c isn't a unit. ■

Corollary: Every P.I.D. is a U.F.D.

Proof:

We showed a few pages ago that every irreducible element in a P.I.D. is a prime element. Also a P.I.D. is Noetherian integral domain and we showed in that last lemma that all nonzero-nonunit elements in a Noetherian integral domain can be factored as a product of irreducible elements. Hence, our claim now follows from the theorem on *page 486*. ■

Hilbert's Basis Theorem: If A is a Noetherian ring then so is $A[x]$.

Proof:

Suppose $\mathfrak{a} \lhd A[x]$. Then we want to show that $\mathfrak{a}$ is finitely generated. So without loss of generality we can assume $\mathfrak{a} \neq \{0\}$.

Claim 1: Let $\mathrm{lt}(\mathfrak{a}) := \{\mathrm{lt}(f) : f \in \mathfrak{a}\}$ (where $\mathrm{lt}(a_n x^n + \ldots + a_0) = a_n$ when $a_n \neq 0$ and $\mathrm{lt}(0) = 0$). We claim $\mathrm{lt}(\mathfrak{a}) \lhd A$.

Proof:

Suppose $a, a' \in \mathrm{lt}(\mathfrak{a})$. Then let $f(x)$ and $g(x)$ be the polynomials in $\mathfrak{a}$ such that $\mathrm{lt}(f) = a$ and $\mathrm{lt}(g)$. By multiplying either f or g by x^k for some integer k we can guarentee that $\deg(f) = \deg(g)$. Then $\mathrm{lt}(f + g) = a + a'$. Also, as $\mathfrak{a}$ is an ideal, we know that $f + g \in \mathfrak{a}$. So, we've shown that $\mathrm{lt}(\mathfrak{a})$ is closed under addition.

Similarly, suppose $c \in A$ and let $f(x) \in \mathfrak{a}$ be such that $\mathrm{lt}(f) = a$. If $ca = 0$ then we trivially know that $ca \in \mathrm{lt}(\mathfrak{a})$ because the zero polynomial is in $\mathfrak{a}$. Meanwhile, if $ca \neq 0$ Then we have that $\mathrm{lt}(cf) = ca$ and $cf \in \mathfrak{a}$. Hence, $ca \in \mathrm{lt}(\mathfrak{a})$.

Claim 2: For all n let $\mathrm{lt}_n(\mathfrak{a}) := \{\mathrm{lt}(f) : f \in \mathfrak{a}$ and $\deg(f) = n\} \cup \{0\}$. We also claim that $\mathrm{lt}_n(\mathfrak{a}) \lhd A$.

The proof of this is mostly the same as that of claim 1. The only difference is that we don't need to multiply f or g by x^k to make the degrees of the polynomials equal.

Since A is Noetherian, we know that $\mathrm{lt}(\mathfrak{a})$ is finitely generated. Hence, there exists $f_1, \ldots, f_n \in \mathfrak{a}$ such that $\mathrm{lt}(\mathfrak{a}) = \langle \mathrm{lt}(f_1), \ldots, \mathrm{lt}(f_n) \rangle$. (Also, without loss of generality assume none of the f_i are redundant. This is mostly just to make sure we don't have that $f_i = 0$ for some i...).

Claim 3: For all $f(x) \in \mathfrak{a}$ there exists $r(x) \in \mathfrak{a}$ such that:

1. $f(x) = f_1(x)q_1(x) + \ldots + f_n(x)q_n(x) + r(x)$,

2. $\deg(r) < \max(\deg(f_1), \ldots, \deg(f_n))$.

Proof:

We proceed by strong induction on the degree of f in $\mathfrak{a}$.

If $\deg(f) < \max(\deg(f_1), \ldots, \deg(f_n))$ then we just set $q_1(x) = \cdots = q_n(x) = 0$ and $r(x) = f(x)$. Then it's clear that both (1) and (2) hold. Meanwhile, suppose $\deg(f) = m \geq \max(\deg(f_1), \ldots, \deg(f_n))$. In other words:

$$f(x) = ax^m + \text{[lower degree stuff]}.$$

Then as $a = \mathrm{lt}(f) \in \langle \mathrm{lt}(f_1), \ldots, \mathrm{lt}(f_n)\rangle$ we know that $a = \sum_{i=1}^n c_i \mathrm{lt}(f_i)$ for some $c_1, \ldots, c_n \in A$. Also, since $\deg(f) \geq \deg(f_i)$ for each i we know there exists a positive integer k_i such that $ax^m = \sum_{i=1}^n c_i \mathrm{lt}(f_i) x^{\deg(f_i)} x^{k_i}$. It now follows that:

$$\deg\left(f(x) - \sum_{i=1}^n c_i x^{k_i} f_i(x)\right) < m$$

Also, since $f(x) - \sum_{i=1}^n c_i x^{k_i} f_i(x)$ is in $\mathfrak{a}$, we can use our induction hypothesis to find $q_1'(x), \ldots, q_n'(x) \in A[x]$ as well as $r(x) \in \mathfrak{a}$ such that:

- $f(x) - \sum_{i=1}^n c_i x^{k_i} f_i(x) = f_1(x)q_1'(x) + \ldots + f_n(x)q_n'(x) + r(x)$,
- $\deg(r) < \max(\deg(f_1), \ldots, \deg(f_n))$.

Now we're done showing this claim because:

$$f(x) = f_1(x)(q_1'(x) + c_1 x^{k_1}) + \ldots + f_1(x)(q_n'(x) + c_n x^{k_n}) + r(x).$$

With that, we just need to show that all of the possibilities for $r(x) \in \mathfrak{a}$ where $\deg(r)$ is less than a certain number N are generated by a finite subset S of $\mathfrak{a}$. After all, we'll then know that $\mathfrak{a} = \langle S \cup \{f_1, \ldots, f_n\}\rangle$. To accomplish that, we consider the ideals $\mathrm{lt}_k(\mathfrak{a})$ for $k < N$.

Since A is Noetherian, we know for each k that there exists polynomials $g_k^{(1)}, \ldots, g_k^{(\ell_k)}$ in $\mathfrak{a}$ all with degree k such that $\mathrm{lt}_k(\mathfrak{a}) = \langle \mathrm{lt}(g_k^{(1)}), \ldots, \mathrm{lt}(g_k^{(\ell_k)})\rangle$.

<u>Claim 4:</u> Suppose $f \in \mathfrak{a}$ and $\deg(f) < N$. Then $f \in \langle \bigcup_{k=1}^N \{g_k^{(1)}, \ldots, g_k^{(\ell_k)}\}\rangle$.

Proof:

We again proceed by strong induction on the degree of f. Note that if $\deg(f) = 0$ then there is nothing to show. Meanwhile suppose $\deg(f) = k$. Then pick $c_1, \ldots, c_{\ell_k} \in A$ such that $\mathrm{lt}(f) = c_1 \mathrm{lt}(g_k^{(1)}) + \ldots + c_{\ell_k} \mathrm{lt}(g_k^{(\ell_k)})$. Then, as f and all the g_k^i have the same degree, we know that:

$$\deg\left(f - \sum_{i=1}^{\ell_k} c_i g_k^{(i)}\right) < k$$

By strong induction we can conclude that:

$$f - \sum_{i=1}^{\ell_k} c_i g_k^{(i)} = \sum_{j=0}^{k-1} \sum_{i=1}^{\ell_j} c_{i,j} g_j^{(i)}$$

And now our claim is obvious.

We now conclude $\mathfrak{a} = \langle \{f_1, \ldots, f_n\} \cup \bigcup_{k=1}^N \{g_k^{(1)}, \ldots, g_k^{(\ell_k)}\}\rangle$. So, $\mathfrak{a}$ is finitely generated.

■

Note that we can also show the converse of Hilbert's basis theorem. In other words, A is a Noetherian ring if $A[x]$ is Noetherian.

Proof:

Suppose A isn't a Noetherian ring. Then there exists a sequence $\{\mathfrak{a}_n\}_{n\in\mathbb{N}}$ of ideals in A such that $\mathfrak{a}_1 \subsetneqq \mathfrak{a}_2 \subsetneqq \cdots$. In turn, $\{\mathfrak{a}_n[x]\}_{n\in\mathbb{N}}$ is also a sequence of ideals in $A[x]$ such that $\mathfrak{a}_1[x] \subsetneqq \mathfrak{a}_2[x] \subsetneqq \cdots$. This proves that $A[x]$ isn't a Noetherian ring. ■

Also note that if A is Noetherian and $\mathfrak{a} \lhd A$, then we have that $A/\mathfrak{a}$ is Noetherian.

Proof:

By the correspondance theorem we know that any ideal of $A/\mathfrak{a}$ is of the form $\mathfrak{b}/\mathfrak{a}$ where $\mathfrak{b} \lhd A$ with $\mathfrak{a} \subseteq \mathfrak{b}$. Since A is Noetherian, we know that $\mathfrak{b}$ is finitely generated. In turn, we also know that $\mathfrak{b}/\mathfrak{a}$ is finitely generated. ■

12/10/2025

Today I'm preparing for the math 200a final and also taking some miscellaneous notes on topics that came up while I was doing practice problems.

Recall how if n is an integer and p is a prime then we defined:

$$\nu_p(n) = \max\{k \in \mathbb{Z}_{\geq 0} : p^k \mid n\}.$$

We can extend ν_p to being defined on $\mathbb{Q}$ as follows. Given any rational number ${}^m/_n \neq 0$ we define $\nu_p({}^m/_n) = \nu_p(m) - \nu_p(n)$.

To prove this is well-defined, suppose $\frac{m_1}{n_1} = \frac{m_2}{n_2}$. Then after canceling common prime factors of m_i and n_i, we can find pairs of integers m_i', n_i' such that for each i:

$$\frac{m_i'}{n_i'} = \frac{m_i}{n_i},\ \gcd(m_i, n_i) = 1, \text{ and } \nu_p(m_i') - \nu_p(n_i') = \nu_p(m_i) - \nu_p(n_i)$$

But now it is easy to check that $m_1' = m_2'$ and $n_1' = n_2'$. After all, we know that $\frac{m_1'}{n_1'} = \frac{m_2'}{n_2'} \iff m_1'n_2' = m_2'n_1'$. Then by comparing the prime factorings of both sides of the latter equation and noting that m_i' and n_i' are coprime for each i, we can show that $m_1' = \pm m_2'$ and $n_1' = \pm n_2'$. Yet the sign of m_1' doesn't effect $\nu_p(m_1')$. ■

Note that since $\nu_p(n_1 n_2) = \nu_p(n_1)\nu_p(n_2)$ when $n_1, n_2 \in \mathbb{Z}$, we can easily show that $\nu_p(rs) = \nu_p(r) + \nu_p(s)$ for all $r, s \in \mathbb{Q} - \{0\}$. We can also easily see that $\nu_p(r^{-1}) = -\nu_p(r)$ for all $r \in \mathbb{Q}$.

Problem 2 From Fall 2024 Midterm: Suppose G is a finite group, p is a prime integer, and H, K are 2 subgroups of G such that $G = HK$.

(a) Prove that there exists $P \in \mathrm{Syl}_p(G)$ such that $P \cap H \in \mathrm{Syl}_p(H)$ and $P \cap K \in \mathrm{Syl}_p(K)$.

Pick $Q_H \in \mathrm{Syl}_p(H)$ and $Q_k \in \mathrm{Syl}_p(K)$. By Sylow's second theorem, there exists $P_1, P_2 \in \mathrm{Syl}_p(G)$ such that $Q_H \subseteq P_1$ and $Q_k \subseteq P_2$.

We claim that $Q_H = P_1 \cap H$.

To see why, note that $P_1 \cap H$ must be a p-group containing Q_H. Furthermore, if $P_1 \cap H$ properly contained Q_H then that would contradict that Q_H is a Sylow p-subgroup.

Similarly, we have that $Q_K = P_2 \cap K$.

But now note by Sylow's second theorem plus the fact $G = HK$ that there must exist $h \in H$ and $k \in K$ such that $P_1 = hkP_2k^{-1}h^{-1}$. Hence, we may let $P := h^{-1}P_1h = kP_2k^{-1}$ and know that $P \in \mathrm{Syl}_p(G)$.

Finally, $P \cap H = h^{-1}(P_1 \cap H)h = h^{-1}Q_Hh \in \mathrm{Syl}_p(H)$ and
$$P \cap K = k(P_2 \cap K)k^{-1} = kQ_Kk^{-1} \in \mathrm{Syl}_p(K).$$

By the way I got the final three sentences of this proof from Hagan.

(b) Suppose $P \in \mathrm{Syl}_p(G)$, $P_H := P \cap H \in \mathrm{Syl}_p(H)$, and $P_K := P \cap K \in \mathrm{Syl}_p(K)$. Then prove that $P = P_HP_K$.

To start off, we know that $|P_H| = p^{r_1}$, $|P_K| = p^{r_2}$, and $|P_H \cap P_K| = p^{r_3}$ where $r_3 \leq \min(r_1, r_2)$. It follows that $|P_HP_K| = |P_H||P_K|/|P_H \cap P_K| = p^\ell$ where $\ell = r_1 + r_2 - r_3 \geq 0$. Also note that $P_H \subseteq P$ and $P_K \subseteq P$ implies that $P_HP_K \subseteq P$. It follows that $p^\ell \leq |P| = p^{\nu_p(|G|)}$. Hence, we can conclude that $\ell \leq \nu_p(|G|)$. And from there it is clear that:
$$0 \leq \nu_p(\tfrac{|G|}{|P_HP_K|})$$

At the same time, note that:
$$\begin{aligned}\nu_p(\tfrac{|G|}{|P_HP_K|}) &= \nu_p(\tfrac{|H|}{|P_H|} \cdot \tfrac{|K|}{|P_K|} \cdot \tfrac{|P_H \cap P_K|}{H \cap K}) \\ &= \nu_p(\tfrac{|H|}{|P_H|}) + \nu_p(\tfrac{|K|}{|P_K|}) + \nu_p(\tfrac{|P_H \cap P_K|}{H \cap K}) = 0 + \nu_p(\tfrac{|P_H \cap P_K|}{H \cap K})\end{aligned}$$

Additionally, note that $P_H \cap P_K = (P \cap H) \cap (P \cap K) = P \cap (H \cap K)$. Hence, it follows that $P_H \cap P_K$ is p-subgroup of $H \cap K$ and therefore $\nu_p(\frac{|P_H \cap P_K|}{H \cap K}) \leq 0$.

With that we know that $\nu_p(\frac{|G|}{|P_HP_K|}) = 0$. Yet, we also know that $\nu_p(\frac{|G|}{|P|}) = 0$. It follows that $\nu_p(\frac{|P|}{|P_HP_K|}) = 0$, and this proves that $|P| = |P_HP_K|$. By invoking one last time that $P_HP_K \subseteq P$ we now know that $P = P_HP_K$. ■

Problem 5(b) from a past final: Suppose A is a (commutative unital) Noetherian ring and $\phi : A \to A$ is a surjective ring homomorphism. Then ϕ is an isomorphism.

Proof:

We need to show ϕ is injective, and to do that it suffices to show $\ker(\phi) = \{0\}$. Luckily note that $\ker(\phi^n) \subseteq \ker(\phi^{n+1})$ for all integers $n \geq 0$ (where we consider ϕ^0 to be the identity map). Thus since A is Noetherian, there must exist some smallest nonnegative integer N such that $\ker(\phi^{N+j}) = \ker(\phi^N)$ for all $j \in \mathbb{N}$.

Suppose $N > 0$. Then we can find $a \in A$ such that $\phi^{N-1}(a) = b \neq 0$ and $\phi(b) = \phi^N(a) = 0$. Yet also note that because ϕ is surjective, we can find $c \in A$ such that $\phi(c) = a$. In turn, we have that $\phi^N(c) = \phi^{N-1}(\phi(a)) = b \neq 0$ but $\phi^{N+1}(c) = \phi(\phi^N(c)) = \phi(b) = 0$. This contradicts that $\ker(\phi^{N+1}) = \ker(\phi^N)$. Hence, we conclude that we can't have that $N > 0$.

But now in particular we must have that $\{0\} = \ker(\phi^0) = \ker(\phi^1)$. ■

Another miscellaneous note I want to make is that if A, A' are both unital rings and $\phi : A \to A'$ is a ring homomorphism such that $1_{A'} \in \operatorname{im}(\phi)$ then we must have that $\phi(1_A) = 1_{A'}$.

After all, suppose $\phi(b) = 1_{A'}$ where b is any element in A. Then:

$$\phi(1_A) = \phi(1_A)1_{A'} = \phi(1_A)\phi(b) = \phi(1_Ab) = \phi(b) = 1_{A'}.$$

Here's some homework problems from the past that I never finished.

Set 6 Problem 6: Suppose G is a group. For all $x, y \in G$, let $[x, y] := xyx^{-1}y^{-1}$ and ${}^xy := xyx^{-1}$. Then Hall's equation asserts that:

$$[[x, y], {}^yz][[y, z], {}^zx][[z, x], {}^xy] = 1.$$

To prove this, first note that:

$$\begin{aligned}[[a, b], {}^bc] &= (aba^{-1}b^{-1})(bcb^{-1})(bab^{-1}a^{-1})(bc^{-1}b^{-1}) \\ &= (aba^{-1})c(ab^{-1}a^{-1})(bc^{-1}b^{-1}) = {}^ab \cdot c \cdot {}^a(b^{-1}) \cdot {}^b(c^{-1})\end{aligned}$$

Also note that ${}^b(a^{-1}) \cdot {}^ba = bab^{-1} \cdot ba^{-1}b^{-1} = 1$. Therefore:

$$\begin{aligned}&[[x, y], {}^yz][[y, z], {}^zx][[z, x], {}^xy] \\ &= ({}^xy \cdot z \cdot {}^x(y^{-1}) \cdot {}^y(z^{-1}))({}^yz \cdot x \cdot {}^y(z^{-1}) \cdot {}^z(x^{-1}))({}^zx \cdot y \cdot {}^z(x^{-1}) \cdot {}^x(y^{-1})) \\ &= ({}^xy \cdot z \cdot {}^x(y^{-1}))(x \cdot {}^y(z^{-1}))(y \cdot {}^z(x^{-1}) \cdot {}^x(y^{-1})) \\ &= (xyx^{-1}zxy^{-1}x^{-1})(xyz^{-1}y^{-1})(yzx^{-1}z^{-1}xy^{-1}x^{-1}) \\ &= (xyx^{-1}zxy^{-1})(yz^{-1})(zx^{-1}z^{-1}xy^{-1}x^{-1}) \\ &= (xyx^{-1}zx)(x^{-1}z^{-1}xy^{-1}x^{-1}) = 1\end{aligned}$$

Next consider the lower central series $\gamma_1(G) = G$ and $\gamma_{i+1}(G) = [G, \gamma_i(G)]$ for all i.

Note that $[H_1, H_2] = [H_2, H_1]$ for any subgroups $H_1, H_2 < G$ since $([h_1, h_2])^{-1} = [h_2, h_1]$. So this definition is equivalent to the one in class.

(a) Suppose $H, K, L \lhd G$ and prove that $[[H, K], L] < [[K, L], H][[L, H], K]$.

To start off, as $H, K, L \lhd G$ we know that $[H, K]$, $[K, L]$, and $[L, H]$ are normal subgroups of G. In turn, we also know that $[[H, K], L]$, $[[K, L], H]$, and $[[L, H], K]$ are normal subgroups of G. And finally, this tells us that $[[K, L], H][[L, H], K]$ is a normal subgroup of G.

(See the lemma at the bottom of *page 378*. Also, I think I forgot to ever prove that if $N_1 \lhd G$ and $N_2 \lhd G$ then $N_1N_2 \lhd G$. Fortunately, the proof is incredibly simple. $xN_1N_2x^{-1} = xN_1x^{-1}xN_2x = N_1N_2$ for all $x \in G$.)

Consider $\overline{G} := \frac{G}{[[K,L],H][[L,H],K]}$. Then if $\pi : G \to \overline{G}$ is the natural projection homomorphism, let $\overline{H} = \pi(H)$, $\overline{K} = \pi(K)$, and $\overline{L} = \pi(L)$.

We claim for all $\overline{h} \in \overline{H}$, $\overline{k} \in \overline{K}$, and $\overline{l} \in \overline{L}$ that $[\overline{h}, \overline{k}]$ and $\overline{l}$ commute (where $\overline{x} = \pi(x) = x[[K,L],H][[L,H],K]$).

Note that $[[k,l],{}^{l}h][[l,h],{}^{h}k] \in [[K,L],H][[L,H],K]$ whenever $h \in H$, $k \in K$, and $l \in L$. This is because H and K are normal subgroups.

In turn, we can apply Hall's equation from the last page to get that:

$$\overline{[[h,k],{}^{k}l]} = \overline{[[h,k],{}^{k}l][[k,l],{}^{l}h][[l,h],{}^{h}k]} = \overline{1}$$

Therefore, we have that $[\overline{h}, \overline{k}]$ and $\overline{{}^{k}l}$ commute.

Finally, note that because L is a normal subgroup we know that conjugation by k is an automorphism on L. Hence, for any $l_1 \in L$ there exists $l_0 \in L$ with $kl_0k^{-1} = l$. Hence, substituting in l_0 for l in the above reasoning we've shown that $[\overline{h}, \overline{k}]$ and $\overline{l}$ commute for all $\overline{h} \in \overline{H}$, $\overline{k} \in \overline{K}$, and $\overline{l} \in \overline{L}$.

Going a step further, we claim that any element in $[\overline{H}, \overline{K}]$ commutes with any $\overline{l}$ in $\overline{L}$.

Suppose $x_1, \ldots, x_n$ are all commutators of $\overline{H}$ and $\overline{K}$. Then since each x_i individually commutes with $\overline{l}$, it's clear that:

$$x_1 \cdots x_{n-1} x_n \overline{l} = x_1 \cdots x_{n-1} \overline{l} x_n = \cdots = \overline{l} x_1 \cdots x_n$$

But now we've shown that $[[\overline{H}, \overline{K}], \overline{L}] = \{1\}$. And since surjective group homomorphisms pass in and out of the brackets in commutator subgroups, we know that $[[\overline{H}, \overline{K}], \overline{L}] = \pi([[H,K],L])$. Hence, $\overline{x} = \overline{1}$ in $\overline{G}$ for all $x \in [[H,K],L]$. And this proves that $[[H,K],L] < [[K,L],H][[L,H],K]$.

(b) Prove for every positive integers m and n that $[\gamma_m(G), \gamma_n(G)] \subseteq \gamma_{m+n}(G)$.

We shall proceed by induction on m.

Note by definition that $[\gamma_1(G), \gamma_n(G)] = [G, \gamma_n(G)] = \gamma_{n+1}(G)$. Thus our base case of $m = 1$ holds trivially.

Next, suppose $m > 1$ and that $[\gamma_k(G), \gamma_n(G)] \subseteq \gamma_{k+n}(G)$ for all $n \in \mathbb{N}$ and $k < m$. Then by part (a) we have for any $n \in \mathbb{N}$ that:

$$\begin{aligned}
[\gamma_m(G), \gamma_n(G)] = [[G, \gamma_{m-1}(G)], \gamma_n(G)] &\subseteq [[\gamma_{m-1}(G), \gamma_n(G)], G][[\gamma_n(G), G], \gamma_{m-1}(G)] \\
&\subseteq [\gamma_{m+n-1}(G), G][\gamma_{n+1}(G), \gamma_{m-1}(G)] \\
&\subseteq \gamma_{m+n}(G)\gamma_{m+n}(G) = \gamma_{m+n}(G). \blacksquare
\end{aligned}$$

Problem 3 From Fall 2024 Midterm:

(a) Let G be a finite group and suppose $P, Q \in \mathrm{Syl}_p(G)$ are distinct. Then show that $P \cap N_G(Q) = P \cap Q$.

We know $\mathrm{Syl}_p(N_G(Q)) = \{Q\}$. Also, $P \cap N_G(Q)$ is a p-group in $N_G(Q)$. Hence, by Sylow's second theorem we must have that $P \cap N_G(Q) \subseteq Q$. Since $P \cap N_G(Q) \subseteq P$ as well we know that $P \cap N_G(Q) \subseteq P \cap Q$. And as $Q \subseteq N_G(Q)$ we trivially know that $P \cap Q \subseteq P \cap N_G(Q)$.

(b) Suppose $P \in \mathrm{Syl}_p(G)$ and consider the action of P on $\mathrm{Syl}_p(G)$ by conjugation. Prove that the P-orbit of $Q \in \mathrm{Syl}_p(G)$ (which I'll hereafter denote $P \cdot Q$) has $[P : P \cap Q]$ many elements.

By the orbit-stabilizer theorem we have that $|P \cdot Q| = [P : P_Q]$ where P_Q is the stabilizer of Q. But note that $x \in P$ is in P_Q if and only if $xQx^{-1} = Q$. Hence $P_Q = P \cap N_G(Q) = P \cap Q$.

(c) Let $s_p = |\mathrm{Syl}_p(G)|$. Then suppose $p^e \mid (s_p - 1)$ and $p^{e+1} \nmid (s_p - 1)$ (where e is some integer). Prove that there are distinct $P, Q \in \mathrm{Syl}_p(G)$ such that $[P : P \cap Q] \leq p^e$.

Suppose $[P : P \cap Q] > p^e$ for all distinct pairs $P, Q \in \mathrm{Syl}_p(G)$. This means by part (b) that if we fix any $P \in \mathrm{Syl}_p(G)$ then the orbit of every $Q \in \mathrm{Syl}_p(G) - \{P\}$ has more than p^e elements. Also note that if $\mathrm{Syl}_p(G)/P$ denotes the set of all P-orbits of the action $P \curvearrowright \mathrm{Syl}_p(G)$ described in part (b), then:

$$s_p = \sum_{P \cdot Q \in \mathrm{Syl}_p(G)/P} |P \cdot Q|$$

This hints at how we can derive a contradiction. Firstly, note that $P \cdot P = \{P\}$. Hence, we can say that:

$$s_p - 1 = \sum_{\substack{P \cdot Q \in \mathrm{Syl}_p(G)/P \\ Q \neq P}} |P \cdot Q|$$

Next note for any $Q \in \mathrm{Syl}_p(G) - \{P\}$ that $|P \cdot Q| = [P : P \cap Q]$ is a power of p. Thus, in order for $|P \cdot Q|$ to have more than p^e elements we must have that $|P \cdot Q| = p^{e+1}p^{k_Q}$ where k_Q is some nonnegative integer. As a result, we can now factor out a p^{e+1} term from the sum above to get that:

$$s_p - 1 = p^{e+1}\Big(\sum_{\substack{P \cdot Q \in \mathrm{Syl}_p(G)/P \\ Q \neq P}} p^{k_Q} \Big)$$

But now we've shown that p^{e+1} divides $s_p - 1$. This is a contradiction. ■

Problem 2 From Fall 2025 Midterm: Suppose G is a finite group and p is a prime divisor of $|G|$. Let P be a Sylow p-subgroup of G and let $x, y \in C_G(P) := \{g \in G : \forall a \in P, ga = ag\}$. Prove that if x and y are conjugate in G then they are conjugate in $N_G(P)$.

Write $y = gxg^{-1}$ where $g \in G$. Then note that since $xa = ax$ for all $a \in P$ we know that $P \subseteq C_G(x)$. Furthermore, since $ya = pa$ for all $a \in P$ we know $gxg^{-1}a = agxg^{-1}$ for all $a \in P$. Equivalently, this means that $xg^{-1}ag = g^{-1}agx$ for all $a \in P$. So, we can conclude that $g^{-1}Pg \subseteq C_G(x)$.

Now we must have that both $P, g^{-1}Pg \in \mathrm{Syl}_p(C_G(x))$. Hence, by Sylow's second theorem there exists $h \in C_G(x)$ such that $hPh^{-1} = g^{-1}Pg$. In turn we know that $P = ghPh^{-1}g^{-1}$. Hence $gh \in N_G(P)$. Also note that $ghxh^{-1}g^{-1} = gxg^{-1} = y$ since $h \in C_G(x)$. ■

Set 7 Problem 3: Suppose G is a finite group and H is a nontrivial subgroup of G.

(a) Show that there exists a function $f : \mathrm{Syl}_p(H) \to \mathrm{Syl}_p(G)$ such that for all $\overline{P} \in \mathrm{Syl}_p(H)$ we have that $\overline{P} = f(\overline{P}) \cap H$. Deduce that $|\mathrm{Syl}_p(H)| \leq |\mathrm{Syl}_p(G)|$.

By Sylow's second theorem, for each $\overline{P} \in \mathrm{Syl}_p(H)$ we can choose some $f(\overline{P}) \in \mathrm{Syl}_p(G)$ such that $\overline{P} \subseteq f(\overline{P})$. Then as $f(\overline{P}) \cap H$ is a p-subgroup in H containing $\overline{P}$, we must have that $f(\overline{P}) \cap H = \overline{P}$. This proves that the function f we want exists.

To show the other inequality, we just note that f is injective. After all, if we know that $f(\overline{P}) = f(\overline{Q})$ then $\overline{P} = f(\overline{P}) \cap H = f(\overline{Q}) \cap H = \overline{Q}$.

(b) Suppose G does not have a non-trivial normal p-subgroup. Then suppose $\overline{P}$ is a non-trivial p-subgroup of G and prove that $|\mathrm{Syl}_p(N_G(\overline{P}))| < |\mathrm{Syl}_p(G)|$.

Note that because $\overline{P}$ is a non-trivial p-subgroup which isn't normal, we know that $\{1\} \subsetneqq N_G(\overline{P}) \subsetneqq G$. Now construct a function $f : \mathrm{Syl}_p(N_G(\overline{P})) \to \mathrm{Syl}_p(G)$ as in part (a). Since we already know f is injective, it suffices to now show that f isn't also surjective.

Suppose for the sake of contradiction that f is a bijection. Then define:

$$O_p(G) \coloneqq \bigcap_{P \in \mathrm{Syl}_p(G)} P.$$

It's easy to see that $O_p(G)$ is a normal p-subgroup of G. Therefore, by assumption we know that $O_p(G) = \{1\}$.

Yet also note that because f is a bijection we know that every $P \in \mathrm{Syl}_p(G)$ is uniquely identified with a group $Q \in \mathrm{Syl}_p(N_G(\overline{P}))$ such that $Q = N_G(\overline{P}) \cap P$. It follows that $O_p(G) \cap N_G(\overline{P}) = O_p(N_G(\overline{P})) = \bigcap_{Q \in \mathrm{Syl}_p(N_G(\overline{P}))} Q$.

Finally, note that that for every $Q \in \mathrm{Syl}_p(N_G(\overline{P}))$ we must have that $\overline{P} \subseteq Q$. After all, we know by Sylow's second theorem that there exists $g \in N_G(\overline{P})$ such that $\overline{P} \subseteq gQg^{-1}$. Equivalently, $\overline{P} = g^{-1}\overline{P}g \subseteq g^{-1}gQg^{-1}g = Q$. This proves that:

$$\overline{P} \subseteq \bigcap_{Q \in \mathrm{Syl}_p(N_G(\overline{P}))} Q \subseteq O_p(G).$$

That contradicts that $O_p(G)$ is trivial since we know that $\overline{P}$ isn't. ■

12/14/2025

Math 241a Notes:

Suppose V is a finite dimensional vector space and $\pi : G \to \mathrm{GL}(V)$ is a representation. Then π is called irreducible if the only $\pi(G)$-invariant subspaces are $\{0\}$ and V. π is called completely reducible if $V = \bigoplus V_i$ where V_i is a $\pi(G)$-invariant irreducible subspace.

Also if $V = \mathbb{C}^n$ or $\mathbb{R}^n$, then I shall denote $\mathrm{GL}(V)$ as $\mathrm{GL}_n(\mathbb{C})$ or $\mathrm{GL}_n(\mathbb{R})$ respectively. Similarly, I shall denote $U(V)$ as $U(n)$.

Proposition 2.2.11: If G is a group and $\pi : G \to U(n)$ is a unitary representation, then:

(i) every $\pi(G)$-invariant subspace has a $\pi(G)$-invariant orthogonal complement.

Proof:
Suppose V is invariant and $w \in V^\perp$. Then as $\pi(g)$ is unitary (which means $\pi(g)^* = \pi(g)^{-1}$) for each $g \in G$, we know:

$$\langle \pi(g)w, v\rangle = \langle w, \pi(g)^* v\rangle = \langle w, \pi(g^{-1})v\rangle = 0.$$

It follows that $V^\perp$ is G-invariant.

(ii) π is completely reducible.

Proof:
We can prove this by induction. If $\mathbb{C}^n$ isn't irreducible then we can write $\mathbb{C}^n = V \oplus V^\perp \cong \mathbb{C}^k \oplus \mathbb{C}^{n-k}$ where both V and $V^\perp$ are G-invariant. Then we just repeat this reasoning on the smaller subspaces. ■

Proposition 2.2.12: If G is a compact group, V is a finite dimensional real or complex Hausdorff topological vector space, and $\pi : G \to \mathrm{GL}(V)$ is a (strong operator) continuous representation, then π is completely reducible.

Proof:
Using *corollary 2.2.8 on page 485*, let $\langle\ ,\ \rangle$ be a G-invariant inner product on V. Then π is a unitary representation with respect to this inner product. So, we can apply the prior proposition. ■

Let $\mathcal{X}$ be a real or complex vector space and let $A \subseteq \mathcal{X}$ be convex.

- Given any $x, y \in \mathcal{X}$ we let $[x, y] := \{ty + (1-t)x : 0 \leq t \leq 1\}$. Also, we let $(x, y) := \{ty + (1-t)x : 0 < t < 1\}$.
- We say $x \in A$ is an extreme point if for any $y, z \in A$ we have that $x \in [y, z]$ iff $x = y$ or $x = z$. We denote the set of such points as $\mathrm{ex}(A)$.
- We say $\emptyset \neq B \subseteq A$ is an extreme set if for any $y, z \in A$ we have that:

$$(y, z) \cap B \neq \emptyset \Longrightarrow [y, z] \subseteq B.$$

Given a set $E \subseteq \mathcal{X}$ where $\mathcal{X}$ is a topological real or complex vector space, we define $\overline{\text{conv}}(A)$ to be the smallest closed convex set containing A. This is well-defined because arbitrary intersections of closed convex sets are closed and convex.

Clearly, if one has a convex polyhedron in $\mathbb{R}^3$, then the faces of the polyhedron are extreme sets and the extreme points are precisely the vertices.

Exercise 2.2.10: Let X be a compact Hausdorff space and let $M(X)$ denote the set of Radon probability measures on X. Then $\text{ex}(M(X)) = \{\delta_x : x \in X\}$ where δ_x is the Dirac delta measure at x.

Proof:
Let μ_0 and μ_1 be probability measures on X, and suppose that $\delta_x \in [\mu_0, \mu_1]$. Hence, there exists $t \in [0,1]$ such that $t\mu_1 + (1-t)\mu_0 = \delta_x$. If $t = 0$ or $t = 1$, there is nothing to show. So suppose $t \in (0,1)$. As $\delta_x(\{x\}^\complement) = 0$, we know that $t\mu_1(\{x\}^\complement) = -(1-t)\mu_0(\{x\}^\complement)$. That said, we also must have that $\mu_1(\{x\}^\complement) \geq 0$ and $\mu_0(\{x\}^\complement) \geq 0$. In turn, the left side of our equation must be nonnegative and the right side must be nonpositive. The only way this works out is if $t\mu_1(\{x\}^\complement) = 0 = -(1-t)\mu_0(\{x\}^\complement)$. And since $t \neq 0$ and $-(1-t) \neq 0$, we can conclude that $\mu_1(\{x\}^\complement) = 0 = \mu_0(\{x\}^\complement)$. And now it is clear that $\mu_0 = \delta_x = \mu_1$ since all three have total measure 1. This proves that $\delta_x \in \text{ex}(M(X))$ for any Dirac delta measure δ_x.

To show the converse, we first introduce a lemma. Suppose ν is a Borel Radon probability measure on X. Then $\nu(E) \in \{0,1\}$ for all sets $E \in \mathcal{B}_X$ if and only if ν is a Dirac delta measure.

Proof:
The $(\Longleftarrow)$ claim is obvious. To show the other claim, you could just use the reasoning on *pages 444-445*. However, I wrote a different proof before realizing that.

Let $\mathcal{F}$ be the set of all compact subsets of X with measure 1. This collection is partially ordered by inclusion, and by Zorn's lemma we can conclude that there is a minimal set F in $\mathcal{F}$.

Suppose $\mathcal{F}_0$ is a chain in $\mathcal{F}$ and let $K' = \bigcap_{K \in \mathcal{F}_0}$. I claim that $\mu(K') = 1$. This will be a compact subset of X since it is a closed subset of X. We also claim $\mu(K') = 1$. After all, if not then by the outer regularity of μ plus the fact that $\mu(E) \in \{0,1\}$ for all sets $E \in \mathcal{B}_X$ we know there exists an open set $U \supseteq K'$ with $\mu(U) = 0$. Next, by the compactness of X we know there are finitely many sets $K_1 \supseteq K_2 \supseteq \cdots \supseteq K_n$ in $\mathcal{F}_0$ such that $X = U \cup \bigcup_{j=1}^n K_j^\complement$. Finally, we know that $K' \neq K_n$ since $\mu(K') \neq \mu(K_n)$. So, there must exist $K \in \mathcal{F}_0$ with $K' \subseteq K \subsetneq K_n$. But in turn we must have that $K \subseteq U$. This implies that $\mu(K) = 0$, which is a contradiction as $K \in \mathcal{F}_0$ means that $\mu(K) = 1$.

We conclude that $K' \in \mathcal{F}$. And clearly K' is a bound to $\mathcal{F}_0$.

Finally, suppose there exists distinct $x, y \in F$. Then we know that $\{x\}$ is a proper compact subset of F. Hence, $\mu(\{x\}) = 0$. Also, by similar arguments to before we know that there is an open set $V \supseteq \{x\}$ such that $\mu(V) = 0$. And, by intersecting V with a neighborhood of x not containing y (which we know exists since X is T_1), we can assume $y \notin V$. But now $F - V$ is a compact subset of X properly contained in F such that $\mu(F - V) = 1$. This contradicts the minimality of F. Hence, we conclude that there does not exist two distinct elements in F.

That said, F isn't empty since otherwise we'd have that $\mu(F) = 0$. So, we conclude that F is a singleton $\{x\}$ and $\mu = \delta_x$.

Now suppose for the sake of contradiction that ν is any measure in $M(X)$ that isn't a Dirac delta measure. Then by our prior lemma we know that there exists a set $E \subseteq X$ such that $0 < \nu(E) < 1$. In turn, we now know there exists well-defined probability measures $\mu_0(A) := (\nu(E))^{-1}\nu(A \cap E)$ and $\mu_1(A) := (\nu(X - E))^{-1}\nu(A - E)$ which are distinct from ν. Finally, by setting $t = \nu(X - E)$ we have that $0 < t < 1$ and $\nu(E) = 1 - t$. And then for all $A \in \mathcal{B}_X$ we have that:

$$\begin{aligned}\nu(A) = \nu(A \cap E) + \nu(A - E) &= \nu(E)\mu_0(A) + \nu(X - E)\mu_1(A) \\ &= (1 - t)\mu_0(A) + t\mu_1(A)\end{aligned}$$

This shows that $\nu \in [\mu_0, \mu_1]$ but $\nu \neq \mu_0$ and $\nu \neq \mu_1$. So, ν is not an extreme point of $M(X)$ if ν is not a Dirac delta measure. ■

Lemma 2.3.5: Suppose $\mathcal{X}$ is a locally convex topological real vector space, $A \subseteq \mathcal{X}$ is closed and convex, and $x \in \mathcal{X} - A$. Then there exists $f \in \mathcal{X}^*$ and $c \in \mathbb{R}$ with $f(y) < c < f(x)$ for all $y \in A$.

Proof:

Let U be an open neighborhood of $0 \in \mathcal{X}$ such that $(x + U) \cap A = \emptyset$. By local convexity we can restrict U to a convex open subset containing 0. And by the reasoning on *page 230-232*, we can further restrict U to also ensure that U is balanced.

Next note that because U is balanced, we can equivalently say that $x \notin U + A$. But we want slightly more wiggle room so we'll instead consider the set $\frac{1}{2}U + A$. A fact we will use later is that because because U is convex, we have that $\frac{1}{2}U + \frac{1}{2}U \subseteq U$.

Note that $\frac{1}{2}U + A$ is open since $\frac{1}{2}U + A = \bigcup_{a \in A}(a + \frac{1}{2}U)$. It's also convex because $t(a_1 + \frac{1}{2}u_1) + (1-t)(a_0 + \frac{1}{2}u_0) = (ta_1 + (1-t)a_0) + \frac{1}{2}(tu_1 + (1-t)u_0) \in A + \frac{1}{2}U$ for all $t \in [0, 1]$, $a \in A$, and $u \in U$ since both U and A are convex. Going a step further, we can assume without loss of generality that $0 \in \frac{1}{2}U + A$.

To see why, note that if $0 \notin \frac{1}{2}U + A$ then we can translate our entire vector space by some fixed $\frac{1}{2}u + a \in U + A$. Then after doing the later reasoning, we will have a linear functional f and $c \in \mathbb{R}$ such that $f(y - (a + \frac{1}{2}u)) < c < f(x - (a + \frac{1}{2}u))$ for all $y \in A$. In turn $f(y - x) < c - f(x - (a + \frac{1}{2}u)) < 0$ and thus:

$$f(y) < c - f(x - (a + \tfrac{1}{2}u)) + f(x) < f(x) \text{ for all } y \in A$$

where $c - f(x - (a + \frac{1}{2}u)) + f(x)$ is another fixed constant in $\mathbb{R}$.

But now if p is the Minkowski functional associated to $\frac{1}{2}U + A$, we can do the reasoning on *page 233* and in the second claim on *page 319* to see that p satisfies the triangle inequality and is continuous. And while $p(cy) \neq |c|p(y)$ necessarily if c is negative since $\frac{1}{2}U + A$ isn't balanced, we do at least have that $p(cy) = cp(y)$ if $c \geq 0$. Hence, we know that p is a well-defined sublinear functional on $\mathcal{X}$.

Now it's obvious that $p(x) \geq 1$ and that $p(y) \leq 1$ for all $y \in A$. What's less obvious is that these inequalities are strict.

- To see that $p(x) > 1$, suppose to the contrary that $x \in c(\frac{1}{2}U + A)$ for all $c > 1$. Equivalently, this means that $cx \in \frac{1}{2}U + A$ for all $c < 1$. But now as $cx - x \to 0$ as $c \to 1$ and $\frac{1}{2}U$ is a neighborhood of 0 in $\mathcal{X}$, we know that eventually $cx - x \in \frac{1}{2}U$. So, we can pick c close enough to 1 such that $cx - x = \frac{1}{2}u'$ for some $u' \in U$. At the same time, as $cx \in \frac{1}{2}U + A$ we know there exists $u \in U$ and $a \in A$ such that $cx = \frac{1}{2}u + a$. Hence, we get a contradiction as:
$$x = \tfrac{1}{2}u - \tfrac{1}{2}u' + a \in \tfrac{1}{2}U + \tfrac{1}{2}U + A \subseteq U + A.$$

- To see that $p(y) < 1$ for any fixed $y \in A$, note again that because $cy - y \to 0$ as $c \to 1$ and $\frac{1}{2}U$ is a neighborhood of 0, we know that there is some $\varepsilon_y > 0$ such that $cy - y \in \frac{1}{2}U$ when $c < 1 + \varepsilon_y$. In turn, $cy \in \frac{1}{2}U + y \subseteq \frac{1}{2}U + A$ when $c < 1 + \varepsilon_y$. And finally, we have that $y \in c(\frac{1}{2}U + A)$ if $c > (1 + \varepsilon_y)^{-1}$ where the latter is strictly less than 1.

Finally, we actually create our linear functional. Let $\mathcal{M} = \{cx : c \in \mathbb{R}\}$ and then define $g : \mathcal{M} \to \mathbb{R}$ by $g(cx) = cp(x)$. Then g is a linear functional on the subspace $\mathcal{M}$. Also since $p(cx) \geq 0 > g(cx)$ when $c < 0$ and we know from the sublinearity of p that $g(cx) = p(cx)$ when $c \geq 0$, we can conclude that $g \leq p$ on $\mathcal{M}$. So, by the real Hahn-Banach theorem we know there exists a linear functional $f : \mathcal{X} \to \mathbb{R}$ with $f(y) \leq p(y)$ for all $y \in \mathcal{X}$ and $f(cx) = g(cx)$ for all $c \in \mathbb{R}$.

Note that $|f(y)| = \max(-f(y), f(y)) = \max(f(-y), f(y)) \leq \max(p(-y), p(y))$ and that p is continuous, meaning that $p(-y) \to 0$ and $p(y) \to 0$ as $y \to 0$. Hence, we can conclude that f is continuous. Also, $f(x) = p(x) > 1 > p(y) \geq f(y)$ for all $y \in A$. ■

Krein-Millman Theorem: Let $\mathcal{X}$ be a topological vector space whose topology is defined by a sufficient family of seminorms. If $A \subseteq \mathcal{X}$ is compact and convex, then $\overline{\operatorname{conv}}(\operatorname{ex}(A)) = A$.

Proof:

Without loss of generality, we may assume $\mathcal{X}$ is a real vector space.

Claim: If B is a closed convex extreme subset of A, then $B \cap \operatorname{ex}(A) \neq \emptyset$.

To prove this we use Zorn's lemma. Let $\mathcal{F}$ be the collection of all closed convex extreme subsets of A. Also partially order $\mathcal{F}$ by inclusion. Then we claim $\mathcal{F}$ has a minimal element.

Let $\mathcal{F}_0$ be a chain in $\mathcal{F}$ and set $C = \bigcap_{B \in \mathcal{F}_0} B$. Then C is not empty by the finite intersection property of A (since A is compact). Also C is closed and convex since it is the intersection of closed convex sets. Finally, suppose $y, z \in A$ satisfy that $(y, z) \cap C \neq \emptyset$. Then for any $B \in \mathcal{F}_0$ we know $(y, z) \cap B \neq \emptyset$. In turn, $[y, z] \subseteq B$ for all $B \in \mathcal{F}_0$. And this proves that $[y, z] \subseteq C$. All of this shows that $C \in \mathcal{F}$.

Now let D be a minimal set in $\mathcal{F}$. If D is a singleton $\{x\}$, then we will be done as $x \in B \cap \mathrm{ex}(A)$.

Suppose for the sake of contradiction that x, y are distinct elements of D. Then by lemma 2.3.5, there exists $f \in \mathcal{X}^*$ such that $f(x) < f(y)$. Since D is compact, we know that $M = \max\{f(z) : z \in D\}$ exists. So, let $E = \{z \in D : f(z) = M\}$. Then E is a proper subset of D as $x \notin E$. We also claim that E is an extreme set, thus contradicting that minimality of D.

E is compact since it is a closed subset of D. Also note that E is convex because if $z_0, z_1 \in D$ and $t \in [0,1]$ then:

$$f(tz_1 + (1-t)z_0) = tf(z_1) + (1-t)z_0 = tM + (1-t)M = M.$$

Finally, suppose $z_0, z_1 \in A$ and $tz_1 + (1-t)z_0 \in E$ for some $t \in (0,1)$. As $D \supseteq E$ is an extreme set we must have that $z_0, z_1 \in D$. And now as $M = f(tz_1 + (1-t)z_0) = tf(z_1) + (1-t)f(z_0)$ and both $f(z_0) \leq M$ and $f(z_1) \leq M$, we must have that $f(z_0) = M = f(z_1)$. So $[z_0, z_1] \subseteq E$.

Now it's clear that $\overline{\mathrm{conv}}(\mathrm{ex}(A)) \subseteq A$ (since A is a closed convex set containing $\mathrm{ex}(A)$). But suppose for the sake of contradiction that there exists $x \in A$ with $x \notin \overline{\mathrm{conv}}(\mathrm{ex}(A))$. By lemma 2.3.5. again we can find a linear functional $f \in \mathcal{X}^*$ such that $f(y) < \alpha < f(x)$ for all $y \in \overline{\mathrm{conv}}(\mathrm{ex}(A))$ (where $\alpha \in \mathbb{R}$). And since A is compact we know like before that $M = \max\{f(x) : x \in A\}$ exists.

By identical reasoning to before we know that $B = \{x \in A : f(x) = M\}$ is an extreme set. So by our claim, we have that $B \cap \mathrm{ex}(A) \neq \emptyset$. Yet this is a contradiction because $\mathrm{ex}(A) \subseteq \overline{\mathrm{conv}}(\mathrm{ex}(A))$ is disjoint from B. ■

Obvious Corollary: If A is a compact convex subset of a topological vector space $\mathcal{X}$ whose topology is generated by a sufficient family of seminorms, then $\mathrm{ex}(A) \neq \emptyset$.

A small lemma worth noting is that if $\mathcal{X}$ is a topological vector space and $A \subseteq \mathcal{X}$ is convex, then so is $\overline{A}$.

To see this, suppose $x, y \in \overline{A}$. Then we know that there are nets $\langle x_i \rangle_{i \in I}$ and $\langle y_J \rangle_{j \in J}$ contained in A and converging to x and y respectively. In turn, by considering the product net $\langle x_i, y_j \rangle_{I \times J}$ we have for any $t \in [0,1]$ that $ty_j + (1-t)x_i \to ty + (1-t)x$. And since $ty_j + (1-t)x_i \in A$ for all $(i,j) \in I \times J$ we have shown that $ty + (1-t)x \in \overline{A}$. So, $\overline{A}$ is convex.

Consequently, we always have that $\overline{\mathrm{conv}(E)} \supseteq \overline{\mathrm{conv}}(E)$ for any set $E \subseteq \mathcal{X}$. And this lets us rephrase the Krein Millman theorem in a slightly more useful way. If $\mathcal{X}$ is as stated in the theorem and $A \subseteq \mathcal{X}$ is compact and convex, then $\overline{\mathrm{conv}(\mathrm{ex}(A))} = A$.

12/22/2025

For this section assume that all vector spaces are Banach spaces.

Suppose $\mathcal{X}, \mathcal{Y}$ are Banach spaces. Then a bounded linear operator $T : \mathcal{X} \to \mathcal{Y}$ is called compact if $\overline{T(\mathcal{X}_1)}$ is compact in $\mathcal{Y}$. (See *page 469* for a reminder of what $\mathcal{X}_1$ means...)

Note: if T has finite rank (meaning $T(\mathcal{X})$ has finitely many dimensions), then T is compact.

Why?

Since $T(\mathcal{X})$ is a finite dimensional subspace, we know by (Rudin) Theorem 1.21 on *page 442* that $T(\mathcal{X})$ is a closed set. Hence, $C := \overline{T(\mathcal{X}_1)}$ is a closed subset of $T(\mathcal{X})$. Furthermore, $C \subseteq \{y \in T(\mathcal{X}) : \|y\| \leq \|T\|_{\mathrm{op}}\}$. So, if we consider any bijective linear isometric map between $\mathbb{C}^n$ (or $\mathbb{R}^n$) and $T(\mathcal{X})$, then we will get that C is homeomorphic to a closed and bounded subset of $\mathbb{C}^n$ (or $\mathbb{R}^n$). By Heine-Borel we thus have that C is compact. ■

As a side note, you can use similar reasoning to show that any closed and bounded set in a finite dimensioned normed vector space is compact.

Lemma 3.1.3: Suppose $\{T_n\}_{n\in\mathbb{N}}$ is a sequence of compact maps in $B(\mathcal{X}, \mathcal{Y})$ and $\|T_n - T\|_{\mathrm{op}} \to 0$ as $n \to \infty$. Then T is also compact.

Proof:

It suffices to prove that $T(\mathcal{X}_1)$ is totally bounded since that will imply $\overline{T(\mathcal{X}_1)}$ is totally bounded (and we already have completeness just from the fact it's a closed set in a complete metric space $\mathcal{Y}$). Note that for all $x, y \in \mathcal{X}_1$, we have that:

$$\begin{aligned}\|Tx - Ty\| &\leq \|Tx - T_n x\| + \|T_n x - T_n y\| + \|T_n y - Ty\| \\ &\leq \|T - T_n\|_{\mathrm{op}}\|x\| + \|T_n x - T_n y\| + \|T_n - T\|_{\mathrm{op}}\|y\| \\ &\leq 2\|T - T_n\|_{\mathrm{op}} + \|T_n x - T_n y\|.\end{aligned}$$

Given any $\varepsilon > 0$, fix n large enough so that $\|T_n - T\|_{\mathrm{op}} < \varepsilon/4$. Then using the fact that T_n is a compact operator, pick $x_1, \ldots, x_m \in \mathcal{X}_1$ such that any $y \in T_n(\mathcal{X}_1)$ is within $\varepsilon/2$ from some $T_n x_i$. It then follows that any $y \in T(\mathcal{X}_1)$ is within ε from some Tx_i. ■

As an application of the above points, suppose $\mathcal{H}$ is a Hilbert space with orthonormal basis $\{e_i\}_{i\in I}$ and $T \in B(\mathcal{H})$ is given by a diagonal matrix $\left[\ \lambda_i\delta_{i,j}\ \right]$ (in other words $Te_i = \lambda e_i$ for all $i \in I$). Then T is compact iff $\{i \in I : |\lambda_i| > \varepsilon\}$ is finite for all $\varepsilon > 0$.

Lemma: If S is a linear operator on $\mathcal{H}$ given by a diagonal matrix $\left[\ \mu_i\delta_{i,j}\ \right]$ where the μ_i are bounded, then $\|S\|_{\mathrm{op}} = \sup_{i\in I}|\mu_i|$.

Proof:

We can use *example 1.2.1 on page 284*. Specifically, recall that $\mathcal{H}$ is unitarily isomorphic to $\ell^2(I)$ by a natural map U. Furthermore, S is unitarily equivalent to multiplication by the element $\mu \in \ell^\infty(I)$ where $\mu = \{\mu_i\}_{i\in I}$.

In other words, $S = U^{-1}M_\mu U$.

Therefore, we have that $\|S\|_{\mathrm{op}} = \|M_\mu\|_{\mathrm{op}} = \|\mu\|_{\mathrm{op}} = \sup_{i\in I}|\mu_i|$. ■

$(\Longleftarrow)$

If the latter is true then the set of i for which $\lambda_i \neq 0$ must be countable. Hence we can enumerate those i as $\{i_n\}_{n\in\mathbb{N}} \subseteq I$. Next, for each n we define T_n by letting $T_n e_{i_k} = \lambda_{i_k}$ for all $k \leq n$ and $T_n e_i = 0$ for all other $i \in I$. Then each T_n is bounded with finite rank, and is thus compact. Also, $\|T - T_n\|_{\mathrm{op}} = \sup_{k>n}|\lambda_{i_k}| \to 0$ as $n \to \infty$. So, T is compact.

$(\Longrightarrow)$
Suppose that there is some $\varepsilon > 0$ such that $S = \{i \in I : \lambda_i \geq \varepsilon\}$ is an infinite set. Then for all $i, j \in S$ we have that $\|Te_i - Te_j\|^2 = |\lambda_i|^2 + |\lambda_j|^2 \geq 2\varepsilon^2$. So if we pick a nonrepeating sequence $\{e_{i_n}\}_{n\in\mathbb{N}}$ where each $i_n \in S$, then this sequence has no subsequential limits. This proves that $\overline{T(\mathfrak{X}_1)}$ is not compact. ■

Recall from my math 240b notes that if $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces, $p \in [1, \infty]$, K is a $(\mathcal{M} \otimes \mathcal{N})$-measurable function on $X \times Y$, and there exists $C > 0$ such that $\int |K(x, y)| \mathrm{d}\mu(x) \leq C$ for a.e. y and $\int |K(x, y)| \mathrm{d}\nu(y)$ for a.e. x, then we have that $(T_K f)(x) \coloneqq \int_Y K(x, y) f(y) \mathrm{d}\nu(y)$ is a linear operator in $B(L^p(\nu), L^p(\mu))$ such that $\|T_K\|_{\mathrm{op}} \leq C$.
(This was Folland theorem 6.18).

The following theorem goes into more depth on this linear operator.

Theorem 3.1.5: Let X and Y be LCH spaces with σ-finite Borel measure μ and ν. Also assume μ and ν are finite on compact sets. If $K \in C_c(X \times Y)$ and $p \in [1, \infty]$ then the integral operator $T_K : L^p(\nu) \to L^p(\mu)$ is compact.

Proof:
Firstly, recall (Folland) proposition 7.22 on *pages 183-184* to see that K is $(\mathcal{M} \otimes \mathcal{N})$-measurable. Furthermore, let $U \subseteq X$ and $V \subseteq Y$ be precompact open sets such that $\mathrm{supp}(K) \subseteq U \times V$. Then given any $\widetilde{K} \in C_c(X, Y)$ with $\mathrm{supp}(\widetilde{K}) \subseteq U \times V$, we have that:

$$\int |\widetilde{K}(x, y)| \mathrm{d}\nu(y) \leq \|\widetilde{K}\|_u \nu(\overline{V}) \text{ and } \int |\widetilde{K}(x, y)| \mathrm{d}\mu(x) \leq \|\widetilde{K}\|_u \mu(\overline{U})$$

Therefore, for all $\widetilde{K} \in C_c(X, Y)$ with $\mathrm{supp}(\widetilde{K}) \subseteq U \times V$, we have that:

$$\|T_{\widetilde{K}}\|_{\mathrm{op}} \leq \|\widetilde{K}\|_u \cdot \max(\mu(\overline{U}), \nu(\overline{V})).$$

But also note by (Folland) proposition 7.21 (also on *page 183*) that there is a sequence of functions $\{K_n\}_{n\in\mathbb{N}}$ in $C_c(X)$ converging uniformly to K and satisfying for all $n \in \mathbb{N}$ that:

- $\mathrm{supp}(K_n) \subseteq U \times V$
- there exists $m \in \mathbb{N}$ such that $K_n(x, y) = \sum_{i=1}^m \phi(x)\psi(y)$ where $\phi_i \in C_c(X)$ and $\psi_i \in C_c(Y)$ for all $i \in \{1, \ldots, n\}$.

Importantly, note that $(T_{K_n} f)(x) = \sum_{i=1}^n (\int_X \psi_i f \mathrm{d}\mu)\phi_i(x)$. It thus follows that each T_{K_n} is a bounded linear operator with finite rank. Additionally,

$$\|T_K - T_{K_n}\|_{\mathrm{op}} = \|T_{(K-K_n)}\|_{\mathrm{op}} \leq \|K - K_n\|_u \cdot \max(\mu(\overline{U}), \nu(\overline{V})) \to 0 \text{ as } n \to \infty.$$

By lemma 3.1.3 we thus know that T_K is compact. ■

There are some other theorems for determining when an integral operator $T_K f(x) = \int K(x, y) f(y) \mathrm{d}\nu(y)$ is a well-defined bounded map.

Example 1.2.14: Let $p \in [1, \infty)$. Then suppose (X, μ) and (Y, ν) are σ-finite measure spaces and $K \in L^p(X \times Y, \mu \times \nu)$. If q is the conjugate exponent of p, then we have that $T_K : L^q(Y) \to L^p(X)$ defined by $(T_K f)(x) = \int_Y K(x, y) f(y) \mathrm{d}\nu(y)$ is a bounded linear map with $\|T_K\|_{\text{op}} \leq \|K\|_{L^p(X \times Y)}$.

This is because for all $f \in L^q(Y)$:

$$\begin{aligned} \|T_K f\|_p^p &= \int | \int K(x,y) f(y) \mathrm{d}\nu(y)|^p \mathrm{d}\mu(x) \\ &\leq \int \left(\int |K(x,y) f(y)| \mathrm{d}\nu(y) \right)^p \mathrm{d}\mu(x) \\ &\leq \int \left(\int |K(x,y)|^p \mathrm{d}\nu(y) \right)^{p/p} \cdot \|f\|_q^p \mathrm{d}\mu(x) = \|f\|_q^p \iint |K(x,y)|^p \mathrm{d}\nu(y) \mathrm{d}\mu(x) \\ &= \|f\|_q^p \|K\|_{L^p(X \times Y)}^p \end{aligned}$$

12/23/2025

Before continuing on with Zimmer, I'd like to do some miscellaneous book keeping as well as an exercise from Folland. As a quick reminder, a function $f : X \to (-\infty, \infty]$ is lower semicontinuous if $f^{-1}(\{x : f(x) > a\})$ is open for all $a \in \mathbb{R}$. See my notes from math 240c for more information and theorems.

(Folland) Exercise 7.30: Let X and Y be LCH spaces and suppose μ and ν are Radon measures on X and Y respectively (not necessarily σ-finite). If f is a nonnegative lower semicontinuous function on $X \times Y$ then $x \mapsto \int f_x \mathrm{d}\nu$ and $y \mapsto \int f^y \mathrm{d}\mu$ are Borel measurable and:

$$\int f \mathrm{d}\mu \mathbin{\widehat{\times}} \nu = \iint f \mathrm{d}\mu \mathrm{d}\nu = \iint f \mathrm{d}\nu \mathrm{d}\mu.$$

The proof of this fact is identical to the proof I wrote down of *Proposition 7.25 on page 217*. Just replace χ_U with any other nonnegative LSC function. ■

I was thinking about this exercise specifically because I want to never worry about whether I can swap orders of summation ever again. Notice that if X and Y are sets (that may be uncountable) equipped with the discrete topology and μ and ν are the counting measures on X and Y, then both μ and ν are Radon measures. Furthermore, by the aforementioned proposition 7.25 we can see that $\mu \mathbin{\widehat{\times}} \nu$ is the counting measure on $X \times Y$. Finally, any real function $X \times Y$ will be LSC since the product topology on $X \times Y$ is the discrete topology. Thus by the prior exercise, if f is any nonnegative function then:

$$\sum_{(x,y) \in X \times Y} f(x,y) = \sum_{x \in X} \sum_{y \in Y} f(x,y) = \sum_{y \in Y} \sum_{x \in X} f(x,y).$$

And we said all this without making any assumption that μ and ν are σ-finite.

For this section always assume $\mathcal{H}$ is a Hilbert space.

Suppose $\mathcal{H}$ has an orthonormal basis $\{e_i\}_{i \in I}$. If $T \in B(\mathcal{H})$ then T is called a Hilbert-Schmidt operator if $\sum_{i,j} |T_{i,j}|^2 < \infty$. Equivalently, this means that $\sum_j \|Te_j\|^2 < \infty$.

(As a reminder, $[\ T_{i,j}\ ]$ is the matrix describing T. In other words, $T_{i,j} = \langle Te_j, e_i \rangle$.)

Lemma 3.1.7: If $\{e_i\}_{i\in I}$ and $\{f_j\}_{j\in J}$ are both orthonormal bases of $\mathcal{H}$ then:

$$\sum_{i_1,i_2} |\langle Te_{i_1}, e_{i_2}\rangle|^2 = \sum_{j_1,j_2} |\langle Tf_{j_1}, f_{j_2}\rangle|.$$

Proof:
By using a combination of Parseval's identity plus the fact that $|\overline{z}| = |z|$ for any $z \in \mathbb{C}$ as well as a few other identities, we have that:

$$\begin{aligned}
\sum_{i_1\in I}\Big(\sum_{i_2\in I} |\langle Te_{i_1}, e_{i_2}\rangle|^2\Big) = \sum_{i_1\in I} \|Te_{i_1}\|^2 &= \sum_{i_1\in I}\Big(\sum_{j_2\in J} |\langle Te_{i_1}, f_{j_2}\rangle|^2\Big) \\
&= \sum_{i_1\in I}\Big(\sum_{j_2\in J} |\langle e_{i_1}, T^* f_{j_2}\rangle|^2\Big) \\
&= \sum_{i_1\in I}\Big(\sum_{j_2\in J} |\overline{\langle T^* f_{j_2}, e_{i_1}\rangle}|^2\Big) \\
&= \sum_{j_2\in J}\Big(\sum_{i_1\in I} |\langle T^* f_{j_2}, e_{i_1}\rangle|^2\Big) \\
&= \sum_{j_2\in J} \|T^* f_{j_2}\|^2 \\
&= \sum_{j_2\in J}\Big(\sum_{j_1\in J} |\langle T^* f_{j_2}, f_{j_1}\rangle|^2\Big) \\
&= \sum_{j_1\in J}\Big(\sum_{j_2\in J} |\overline{\langle f_{j_1}, T^* f_{j_2}\rangle}|^2\Big) \\
&= \sum_{j_1\in J}\Big(\sum_{j_2\in J} |\langle f_{j_1}, T^* f_{j_2}\rangle|^2\Big) \\
&= \sum_{j_1\in J}\Big(\sum_{j_2\in J} |\langle Tf_{j_1}, f_{j_2}\rangle|^2\Big). \blacksquare
\end{aligned}$$

The last lemma the sum $\sum_{i,j} |T_{i,j}|^2$ is independent of the particular orthonormal basis chosen for $\mathcal{H}$. Hence, we define the Hilbert-Schmidt norm of T (denoted $\|T\|_2$) by:

$$\|T\|_2 = \big(\textstyle\sum_{i,j} |T_{i,j}|^2\big)^{1/2}$$

where $\left[\ T_{i,j}\ \right]$ is the matrix describing T with respect to any orthonormal basis.

Clearly T is Hilbert-Schmidt iff $\|T\|_2 < \infty$. Also note that $\|T\|_2 = \|T^*\|_2$.

Lemma 3.1.10: If T is Hilbert-Schmidt then $\|T\|_{\text{op}} \leq \|T\|_2$.

Proof:
Let $\{e_i\}_{i\in I}$ be any orthonormal basis for $\mathcal{H}$. Then given any $x \in \mathcal{H}$, after writing $x = \sum_i c_i e_i$ we have that:

$$\|Tx\|^2 = \|T(\textstyle\sum_i c_i e_i)\|^2 = \sum_j |\langle T(\sum_i c_i e_i), e_j\rangle|^2 = \sum_j |\sum_i c_i \langle Te_i, e_j\rangle|^2$$

By the Cauchy-Schwartz inequality on $\ell^2(I)$, we have that:

$$|\textstyle\sum_i c_i \langle Te_i, e_j\rangle|^2 \leq (\sum_i |c_i|^2)(\sum_i |\overline{\langle Te_i, e_j\rangle}|^2) = \|x\|^2 \sum_i |\langle Te_i, e_j\rangle|^2$$

So, $\|Tx\|^2 \leq \sum_j (\|x\|^2 \sum_i |\langle Te_i, e_j\rangle|^2) = \|x\|^2 \sum_{j,i} |\langle Te_i, e_j\rangle|^2 = \|x\|^2 \|T\|_2^2$. $\blacksquare$

Proposition 3.1.11: If T is Hilbert-Schmidt then T is compact.

Let $\{e_i\}_{i\in I}$ be an orthonormal basis. Since $\sum_{i\in I}\|Te_i\|^2$ is finite, we know that there is a countably many $i\in I$ such that $Te_i\neq 0$. So, let $\{i_k\}_{k\in\mathbb{N}}$ be an enumeration of all those i for which $Te_i\neq 0$.

Next, for each n define $T_n\in\mathbb{N}$ by $T_ne_{i_k}=Te_{i_k}$ if $k\leq n$ and $T_ne_i=0$ for all other i. Then note that each T_n has finite rank and is thus compact. Also note that each $T-T_n$ is Hilbert-Schmidt with $\|T-T_n\|_2^2=\sum_{k>n}\|Te_{i_k}\|^2$. It follows that $\|T-T_n\|_2\rightarrow 0$ as $n\rightarrow\infty$. By the last lemma this means that $\|T-T_n\|_{\mathrm{op}}\rightarrow 0$ as $n\rightarrow\infty$. So, by lemma 3.1.3 we know that T is compact. ■

As a side note, the converse is not true. After all, just pick a sequence $\{\lambda_i\}_{i\in\mathbb{N}}$ that isn't in $\ell^2(\mathbb{N})$ but which still converges to 0 as $i\rightarrow\infty$. Then let T be a linear map associated to the diagonal matrix $[\ \lambda_i\delta_{i,j}\ ]$. By the reasoning on *pages 502-503* we know that T is compact. That said, T won't be Hilbert-Schmidt.

Proposition 3.1.12: Let (X,μ) be a σ-finite measure space and $K\in L^2(X\times X)$. Then $T_k : L^2(X)\rightarrow L^2(X)$ (given by $(T_kf)(x) := \int K(x,y)f(y)\mathrm{d}\mu(y)$) is Hilbert-Schmidt. Furthermore, $\|T_K\|_2=\|K\|_2$.

Proof:

Note that $K_x(y) := K(x,y)$ is in $L^2(\mu)$ for a.e. $x\in X$ (by Fubini's theorem). Also, let $\{e_i\}_{i\in I}$ be an orthonormal basis of $L^2(X)$. Then:

$$\|T_K\|_2^2=\sum_{i\in I}\|T_Ke_i\|^2=\sum_{i\in I}\int_X|(T_Ke_i)(x)|^2\mathrm{d}\mu(x)$$

$$=\sum_{i\in I}\int_X|\langle\overline{K_x},e_i\rangle|^2\mathrm{d}\mu(x)=\int_X\|\overline{K_x}\|_2^2\mathrm{d}\mu(x)=\|K\|_2^2$$

We let $B_2(\mathcal{H})$ denote the set of all Hilbert-Schmidt operators in $B(\mathcal{H})$. As a side note, note that $B_2(\mathcal{H})$ equipped with the Hilbert-Schmidt norm is a Hilbert space. After all, if $\{e_i\}_{i\in I}$ is an orthonormal basis for $\mathcal{H}$ then there is an obvious norm preserving linear map from $B_2(\mathcal{H})$ to $\ell^2(I\times I)$. In turn, we can easily see that $\|\cdot\|_2$ is a well-defined norm on $B_2(\mathcal{H})$ and that $B_2(\mathcal{H})$ is a complete Banach space. Then going one step further, we can conclude that the inner product $\langle T,S\rangle_2=\sum_{i,j\in I}T_{i,j}\overline{S_{i,j}}$ is an inner product for $B_2(\mathcal{H})$.

Note that $TS^*e_i=T(\sum_{j\in I}\overline{S_{i,j}}e_j)=\sum_{j\in I}\overline{S_{i,j}}Te_j=\sum_{j\in I}\overline{S_{i,j}}\sum_{k\in I}T_{k,j}e_k=\sum_{j,k\in I}\overline{S_{i,j}}T_{k,j}e_k$.

In particular, this shows that $(TS^*)_{i,i}=\sum_{j\in I}T_{i,j}\overline{S_{i,j}}$. So, we can equivalently say that $\langle T,S\rangle_2=\sum_{i,i\in I}(TS^*)_{i,i}$.

Given a Banach Space $\mathcal{X}$, we let $B_c(\mathcal{X})$ denote the set of compact operators in $B(\mathcal{X})$. Note that $B_c(\mathcal{X})$ is a closed ideal of $B(\mathcal{X})$ considered as a Banach algebra.

Lemma: If $S\in B(\mathcal{X})$ and $T\in B_c(\mathcal{X})$ then $ST,TS\in B_c(\mathcal{X})$.

Proof:

Since $\overline{T(\mathcal{X}_1)}$ is compact, S is continuous, and $\mathcal{X}$ is Hausdorff, we know that $S(\overline{T(\mathcal{X}_1)})$ is closed and compact. And since $ST(\mathcal{X}_1)\subseteq S(\overline{T(\mathcal{X}_1)})$, we have that $\overline{ST(\mathcal{X}_1)}\subseteq S(\overline{T(\mathcal{X}_1)})$. So, $\overline{ST(\mathcal{X}_1)}$ is compact. This shows that ST is compact.

Meanwhile, note that $TS(\mathcal{X}_1) \subseteq T(\|S\|_{\text{op}}\mathcal{X}_1) = \|S\|_{\text{op}}T(\mathcal{X}_1) \subseteq \|S\|_{\text{op}}\overline{T(\mathcal{X}_1)}$. If $\|S\|_{\text{op}} = 0$ then we trivially have that $\overline{TS(\mathcal{X}_1)} = \{0\}$ is compact. Meanwhile, if $\|S\|_{\text{op}} \neq 0$ then we know that scalar multiplication by $\|S\|_{\text{op}}$ is a homeomorphism. So $\|S\|_{\text{op}}\overline{T(\mathcal{X}_1)}$ is closed and compact. In turn $\overline{TS(\mathcal{X}_1)} \subseteq \|S\|_{\text{op}}\overline{T(\mathcal{X}_1)}$ is also compact. ■

Since scalar multiplication is a linear map, the other lemma also proves that $B_c(\mathcal{X})$ is closed under scalar multiplication. Finally, note that if $T, S \in B_c(\mathcal{X})$ then so is $T + S$.

Proof:

Note that $(T+S)(\mathcal{X}_1) \subseteq T(\mathcal{X}_1) + S(\mathcal{X}_1) \subseteq \overline{T(\mathcal{X}_1)} + \overline{S(\mathcal{X}_1)}$. The latter set is compact since it's the image of the compact set $\overline{T(\mathcal{X}_1)} \times \overline{S(\mathcal{X}_1)}$ under vector addition. Since $\mathcal{X}$ is Hausdorff we also know the latter set is closed. So $\overline{(T+S)(\mathcal{X}_1)} \subseteq \overline{T(\mathcal{X}_1)} + \overline{S(\mathcal{X}_1)}$ is compact.

Finally, the fact $B_c(\mathcal{X})$ is closed is just *lemma 3.1.3*.

Let $\mathcal{X}$ be a Banach space and $T \in B(\mathcal{X})$. Then for $\lambda \in \mathbb{C}$ let $\mathcal{X}_{T,\lambda} \coloneqq \{x \in E : Tx = \lambda x\}$. Importantly note that $\mathcal{X}_{T,\lambda}$ is a closed subspace (as it's the kernel of the continuous map $T - \lambda\mathrm{Id}$). We say λ is an eigenvalue if $\mathcal{X}_{T,\lambda} \neq \{0\}$. Also, we say any $x \in \mathcal{X}_{T,\lambda}$ is an eigenvector with eigenvalue λ.

Lemma 3.2.4: Let $\mathcal{H}$ be a Hilbert space and $T \in B(\mathcal{H})$.

(i) If $T = T^*$ and W is T-invariant (i.e. $T(W) \subseteq W$), then $W^\perp$ is also T-invariant.

Why? Suppose $x \in W^\perp$. Then for any $y \in W$ we have that $\langle Tx, y\rangle = \langle x, Ty\rangle = 0$ since $Ty \in W$. So, $T(W^\perp)$.

(ii) If $T = T^*$ then for all $x \in \mathcal{H}$ we have that $\langle Tx, x\rangle \in \mathbb{R}$.

Why? $\overline{\langle Tx, x\rangle} = \langle x, Tx\rangle = \langle Tx, x\rangle$.

Consequently, all eigenvalues are real.

Suppose $Tx = \lambda x$ and $x \neq 0$. Then $\langle Tx, x\rangle = \lambda\|x\|^2$. Therefore:

$$\lambda = \frac{1}{\|x\|^2}\langle Tx, x\rangle \in \mathbb{R}.$$

(iii) $\|T\|_{\text{op}} = \sup\{|\langle Tx, y\rangle| : x, y \leq 1\}$.

Proof:

By the Cauchy-Schwartz inequality we know that $|\langle Tx, y\rangle| \leq \|Tx\|\|y\| \leq \|T\|_{\text{op}}$ if $\|x\|, \|y\| \leq 1$. On the other hand, suppose $T \neq 0$ and let $y \in \mathcal{H}$ with $Ty \neq 0$. After setting $x = Ty/\|Ty\|$ we have that $|\langle Ty, x\rangle| = \|Ty\|$. Then taking the supremum over all $y \in \mathcal{H}_1$ gives us $\|T\|_{\text{op}}$.

(iv) If $T = T^*$ then $\|T\|_{\text{op}} = \sup\{|\langle Tx, x\rangle| : \|x\| \leq 1\}$.

Proof:

Let $\alpha = \sup\{|\langle Tx, x\rangle| : \|x\| \leq 1\}$. Then it's clear that:

$$\alpha \leq \sup\{|\langle Tx, y\rangle| : \|x\|, \|y\| \leq 1\}.$$

If we can show the other inequality, we will then know by part (iii) that $\alpha = \|T\|_{\text{op}}$. Let $x, y \in \mathcal{H}_1$. Without loss of generality we can assume $\langle Tx, y\rangle = \mathbb{R}$ since if not we can scale y by some constant c such that $|c| = 1$ and $|\langle Tx, y\rangle| = \langle Tx, cy\rangle$.

Next note that $\langle T(x+y), x+y\rangle = \langle Tx, x\rangle + \langle Tx, y\rangle + \langle Ty, x\rangle + \langle Ty, y\rangle$. In turn, since $T = T^*$ and $\langle Tx, y\rangle = \overline{\langle Tx, y\rangle}$, we have that:

$$\langle T(x+y), x+y\rangle = \langle Tx, x\rangle + 2\langle Tx, y\rangle + \langle Ty, y\rangle.$$

Similarly, we have that $\langle T(x-y), x-y\rangle = \langle Tx, x\rangle - 2\langle Tx, y\rangle + \langle Ty, y\rangle$. Thus, by subtracting we get that:

$$4\langle Tx, y\rangle = \langle T(x+y), x+y\rangle - \langle T(x-y), x-y\rangle$$

Now note that for any $w \in \mathcal{H}$ we know that $\frac{1}{\|w\|^2}|\langle Tw, w\rangle| = |\langle T\frac{w}{\|w\|}, \frac{w}{\|w\|}\rangle| \leq \alpha$. Combining this fact with the prior result, we get that:

$$|\langle Tx, y\rangle| \leq \tfrac{\alpha}{4}(\|x+y\|^2 + \|x-y\|^2).$$

By the parallelogram law we know $\frac{\alpha}{4}(\|x+y\|^2 + \|x-y\|^2) = \frac{\alpha}{2}(\|x\|^2 + \|y\|^2)$. And finally since $\|x\|^2 + \|y\|^2 \leq 2$, we have that $|\langle Tx, y\rangle| \leq \alpha$.

(v) If $T = T^*$ and $\lambda \neq \beta$ then $\mathcal{H}_{T,\lambda} \perp \mathcal{H}_{T,\beta}$ (meaning that $\langle x, y\rangle = 0$ for all $x \in \mathcal{H}_{T,\lambda}$ and $y \in \mathcal{H}_{T,\beta}$).

Why? Suppose $Tx = \lambda x$ and $Ty = \beta y$. Then we know by part (ii) that $\lambda, \beta \in \mathbb{R}$. Hence:

$$\lambda\langle x, y\rangle = \langle Tx, y\rangle = \langle x, Ty\rangle = \beta\langle x, y\rangle.$$

Since $\lambda \neq \beta$, this implies that $\langle x, y\rangle = 0$. ■.

<u>Spectral Theorem For Compact Operators:</u> Suppose $\mathcal{H} \neq \{0\}$ is a Hilbert space and $T \in B(\mathcal{H})$ is compact and self-adjoint. Then $\mathcal{H}$ has an orthonormal basis consisting of eigenvectors for T. Furthermore, for each $\lambda \neq 0$ we have that $\dim(\mathcal{H}_{T,\lambda}) < \infty$. Also, for each $\varepsilon > 0$ we have that $\{\lambda : |\lambda| > \varepsilon$ and $\dim(\mathcal{H}_{T,\lambda}) > 0\}$ is a finite set.

Proof:

If $T = 0$ then the theorem is trivial. So assume $T \neq 0$. Then our first goal is to find a single eigenvector. By part (iv) of the prior lemma, pick a sequence $\{x_n\}_{n\in\mathbb{N}}$ in $\mathcal{H}_1$ with $|\langle Tx_n, x_n\rangle| \to \|T\|_{\text{op}}$. Next, by parts (ii) and (iv) of the prior lemma we know that $\langle Tx_n, x_n\rangle$ is contained in the compact set $[-\|T\|_{\text{op}}, +\|T\|_{\text{op}}] \subseteq \mathbb{R}$. So, we can pass to a subsequence and assume $\langle Tx_n, x_n\rangle \to \lambda$ where $\lambda = \pm\|T\|_{\text{op}}$. Then since T is compact, we can assume after passing to another subsequence that $Tx_n \to y$ for some $y \in \mathcal{H}$ as $n \to \infty$. Since $\|T\|_{\text{op}} \neq 0$ and $\|x_n\|$ is bounded, we also know that $y \neq 0$.

To see why, note by the Cauchy-Schwarz inequality that $|\langle Tx_n, x_n\rangle| \leq \|Tx_n\|\|x_n\|$. If $Tx_n \to 0$ then the latter quantity also goes to 0, thus contradicting that $|\langle Tx_n, x_n\rangle| \to \|T\|_{\text{op}} \neq 0$.

Next note that as $T = T^*$, we have that:

$$\begin{aligned}\|Tx_n - \lambda x_n\|^2 &= \|Tx_n\|^2 - \langle \lambda x_n, Tx_n\rangle - \langle Tx_n, \lambda x_n\rangle + \|\lambda x_n\|^2 \\ &= \|Tx_n\|^2 - 2\lambda\langle Tx_n, x_n\rangle + |\lambda|^2\|x_n\|^2 \leq 2\|T\|_{\text{op}} - 2\lambda\langle Tx_n, x_n\rangle\end{aligned}$$

The latter goes to 0 as $n \to \infty$. Hence, we've proven that $\|Tx_n - \lambda x_n\| \to 0$ as $n \to \infty$. Yet as $Tx_n \to y$ as well, we thus know that $\lambda x_n \to y$. Composing that limit with T and using the continuity of T, we get that $T(\lambda x_n) = \lambda Tx_n \to Ty$. And finally, as $\lambda Tx_n \to \lambda y$ as well we can conclude that $\lambda y = Ty$. So, y is an eigenvector of T with eigenvalue λ.

Now that we know that T has a nonzero eigenvector, we now show that T has an othonormal eigenbasis. Specifically, using Zorn's lemma we can find a maximal orthonormal set of eigenvectors. Then we let W be the closure of the span of that set.

The reason we can use Zorn's lemma is that we can let $\mathcal{F}$ be the collection of all orthonormal sets of eigenvectors. This will be partially ordered by inclusion and any chain will have an upper bound gotten by just taking the union of the sets in the chain.

Clearly $T(W) \subseteq W$. So by part (i) of the prior lemma we know that $T(W^\perp) \subseteq W^\perp$. In particular, this means that $T|_{W^\perp}$ is a linear operator in $B(W^\perp)$. Since, $W^\perp$ is a closed subset of $\mathcal{H}$, we know that the closure of $T(W_1^\perp)$ in the subspace topology of $W^\perp$ is also closed in $\mathcal{H}$. Adding in the fact that $T(\mathcal{H}_1) \supseteq T(W_1^\perp)$, we know that $T|_{W^\perp}$ is compact. At the same time, $(T^*)|_{W^\perp} = (T|_{W^\perp})^*$. After all, if we consider an orthonormal basis of $W^\perp$, then both maps are described by the same matrix.

But now note that if $W^\perp \neq \{0\}$, then we know by our prior reasoning that $T|_{W^\perp}$ has a nonzero eigenvector with eigenvalue $\pm\|T|_{W^\perp}\|_{\text{op}}$. This contradicts the maximality of the orthonormal set which we constructed W from. Hence, we know that $W^\perp = \{0\}$. And this proves the existence of an orthonormal eigenbasis for $\mathcal{H}$.

All the remaining claims about the eigenbasis then follow from the example on *pages 502-503*. (The matrix associated with T according to that eigenbasis is a diagonal matrix).

Corollary 3.2.5: Let $\{T_\alpha\}_{\alpha \in A}$ be a finite subset of $B(\mathcal{H})$ such that each T_α is compact and self-adjoint, and $T_\alpha T_\beta = T_\beta T_\alpha$ for all $\alpha, \beta \in A$. Then there is an orthonormal basis $\{e_j\}_{j \in J}$ of $\mathcal{H}$ such that each e_j is an eigenvector for every T_α.

Lemma: Suppose $B \subseteq A$ and for each $\beta \in B$ pick $\lambda_\beta \in \mathbb{C}$. Then after setting $W := \bigcap_{\beta \in B} \mathcal{H}_{T_\beta, \lambda_\beta}$, if $W \neq \{0\}$ then we can find an orthonormal basis of W consisting of vectors which are eigenvectors for T_α as well as all T_β where $\beta \in B$.

Proof:

Suppose $x \in W$. Then for any $\alpha \in A$ and $\beta \in B$ we have that:

$$T_\beta T_\alpha x = T_\alpha T_\beta x = \lambda_\beta T_\alpha x.$$

In other words, $x \in W \Longrightarrow T_\alpha x \in W$. Hence, $T_\alpha|_W \in B(W)$. Since W is finite-dimensioned, we can also conclude that W is closed. Then by identical reasoning as was used in the main theorem, we can conclude that $T_\alpha|_W$ is compact and self-adjoint. Hence, we can apply the main theorem to find an orthonormal eigenbasis of $T_\alpha|_W$ on W. And clearly, every vector in that basis is also an eigenvector for each T_β with eigenvalue λ_β.

Now our theorem is just a matter of induction on $|A|$. The base case of when $|A| = 1$ was our main theorem. As for the induction step, pick any $\alpha_0 \in A$ and let $A' = A - \{\alpha_0\}$. Then by induction let $\{f_i\}_{i \in I}$ be an orthonormal basis of $\mathcal{H}$ such that each f_i is an eigenvector for all T_α where $\alpha \in A'$. This tells us that $\mathcal{H}$ is the closure of the direct product of a bunch of vector subspaces W as in the lemma above. So, we just apply that lemma to each of those subspaces to get a new orthonormal basis of $\mathcal{H}$. ■

As a side note, I can't figure out how to extend this corollary to the case where A is infinite.

Suppose $\mathcal{H}$ is a Hilbert space and $T \in B(\mathcal{H})$. Then T is called normal if $TT^* = T^*T$.

- If T is unitary or self-adjoint, then T is normal.
- Suppose $T = T^*$. If $\mathcal{H}$ is a complex vector space then $p(T)$ is normal for any $p(x) \in \mathbb{C}[x]$. Also, if $\mathcal{H}$ is a real or complex vector space and $p(x) \in \mathbb{R}[x]$ then we have that $p(T)$ is self-adjoint.

To see why, first note that if $c_1, c_2 \in \mathbb{C}$ then $(c_1 A + c_2 B)^* = \overline{c_1} A^* + \overline{c_2} B^*$. After all, $\langle c_1 Ax + c_2 Bx, y \rangle = \langle x, \overline{c_1} A^* y + \overline{c_2} B^* y \rangle$ for all $x, y \in \mathcal{H}$.

As a side note, the $*$ operator is only conjugate linear as opposed to linear because the identification $x \mapsto \langle \cdot, x \rangle$ of $\mathcal{H}$ with $\mathcal{H}^*$ is conjugate linear. If we were instead considering adjoints as linear maps on $\mathcal{H}^*$, then the $*$ operation would be properly linear.

Also note that as $(AB)^* = B^* A^*$, we have that $(T^n)^* = (T^*)^n$ for all integers $n \geq 0$. Therefore, if $p(x) = a_n x^n + \ldots + a_1 x + a_0 \in \mathbb{C}[x]$, then upon letting $\overline{p}(x) = \overline{a_n} T^n + \ldots + \overline{a_1} T + \overline{a_0} \mathrm{Id}$ we have that:

$$p(T)^* = \overline{p}(T^*) = \overline{p}(T).$$

Now all the claims follow from the fact that $p(x)\overline{p}(x) = \overline{p}(x)p(x)$ and that $\overline{p}(x) = p(x)$ if $p(x) \in \mathbb{R}[x]$.

Lemma: If $T \in B(\mathcal{H})$, then $(T^*)^* = T$.

To see why, note for any $x, y \in \mathcal{H}$ that:

$$\langle Tx, y \rangle = \langle x, T^* y \rangle = \overline{\langle T^* y, x \rangle} = \overline{\langle y, (T^*)^* x \rangle} = \langle (T^*)^* x, y \rangle$$

Now note that given any $T \in B(\mathcal{H})$, we have that $T = T_1 + iT_2$ where $T_1 = \frac{1}{2}(T^* + T)$ and $T_2 = \frac{i}{2}(T^* - T)$. Furthermore, note that:

$$T_1^* = \tfrac{1}{2}((T^*)^* + T^*) = T_1 \text{ and } T_2^* = \tfrac{-i}{2}((T^*)^* - T^*) = \tfrac{i}{2}(T^* - (T^*)^*) = T_2.$$

Hence, both T_1 and T_2 are self-adjoint.

Exercise 3.17: $T = S_1 + iS_2$ where both S_1, S_2 are bounded self-adjoint operators, then $S_1 = T_1$ and $S_2 = T_2$.

To see why, note that $T_1 = \frac{1}{2}(T^* + T) = \frac{1}{2}((S_1 - iS_2) + (S_1 + iS_2)) = S_1$. Similarly, $T_2 = \frac{i}{2}(T^* - T) = \frac{i}{2}((S_1 - iS_2) - (S_1 + iS_2)) = S_2$.

Furthermore, T is normal iff $T_1T_2 = T_2T_1$.

To see why, note that:

- $TT^* = (T_1 + iT_2)(T_1 - iT_2) = T_1^2 - iT_1T_2 + iT_2T_1 + T_2^2$
- $T^*T = (T_1 - iT_2)(T_1 + iT_2) = T_1^2 + iT_1T_2 - iT_2T_1 + T_2^2$

If $T_1T_2 = T_2T_1$, then $TT^* = T_1^2 + T_2^2 = T^*T$. Meanwhile if $TT^* = T^*T$ then we have that $0 = (-iT_1T_2 + iT_2T_1) - (iT_1T_2 - iT_2T_1) = -2iT_1T_2 + 2iT_2T_1$. Or in other words, $T_1T_2 = T_2T_1$.

One more note is that if T is compact, then we have that both T_1 and T_2 are compact since $B_c(\mathcal{H})$ is an ideal. Therefore, we can generalize the Spectral Theorem for Compact Operators as well as corollary 3.2.5 to applying to compact normal operators (as opposed to just compact self-adjoint operators).

As a side note though, unitary operators won't be compact unless $\mathcal{H}$ is finite-dimensional.

To see why, note that if x is a nonzero eigenvector of the unitary map U with eigenvalue λ, then:

$$|\lambda|^2\|x\|^2 = \langle \lambda x, \lambda x\rangle = \langle Ux, Ux\rangle = \langle x, x\rangle = \|x\|^2$$

Thus, all eigenvalues of U have modulus 1. Yet this contradicts the spectral theorem for compact operators if U is compact and $\mathcal{H}$ is infinite dimensional.

1/1/2026

Let $\mathcal{X}$ be a Banach space and $T \in B(\mathcal{X})$. We say that $\lambda \in \mathbb{C}$ is in the spectrum of T if $T - \lambda\mathrm{Id} \notin \mathrm{Aut}(\mathcal{X})$ (i.e. $T - \lambda\mathrm{Id}$ doesn't have a continuous inverse). We denote the spectrum of T as $\sigma(T)$.

If λ is an eigenvalue, then $\lambda \in \sigma(T)$. Furthermore, if $\dim(\mathcal{X}) < \infty$ then we know from my math 100b notes that $\lambda \in \sigma(T)$ iff λ is an eigenvalue. In general though, if $\dim(\mathcal{X}) = \infty$ then $\sigma(T)$ contains numbers that aren't eigenvalues.

For example, let $\mathcal{H}$ be a separable Hilbert space with ornothonormal basis $\{e_n\}_{n\in\mathbb{N}}$. Then let $T \in B(\mathcal{H})$ be given by $Te_n = \lambda_n e_n$ where $\lambda_n \to 0$ as $n \to \infty$ but $\lambda_n \neq 0$ for any $n \in \mathbb{N}$. It follows that 0 is not an eigenvalue of T (which we know because $\ker(T) = 0$). That said $0 \in \sigma(T)$. After all, if T had a continuous inverse then we'd have that $T^{-1}e_n = \lambda_n^{-1}e_n$ for all n. Yet as $|\lambda_n^{-1}| \to \infty$ as $n \to \infty$, clearly T^{-1} wouldn't be bounded.

I just also remembered, by the open mapping theorem we know that if T is a surjective bounded linear map between two Banach spaces then T is an open map. In turn, if T had an inverse at all, then we'd automatically have that that inverse is continuous. Hence, we can actually go a step further and say that $\lambda \in \sigma(T)$ iff $T - \lambda\mathrm{Id}$ isn't a bijection.

Recall from my math 240b homework that if (X, μ) is a measure space and $f : X \to \mathbb{C}$ is a measurable function, then the essential range of f (denoted R_f) is the set of all $z \in \mathbb{C}$ such that $\mu(\{x \in X : |f(x) - z| < \varepsilon\}) > 0$ for all $\varepsilon > 0$. In Folland exercise 6.11, I proved that R_f is closed and that if $f \in L^\infty(X)$ then R_f is compact with $\|f\|_\infty = \max\{|z| : z \in R_f\}$.

Now recall *example 1.2.1 on page 284*. Let $\mathcal{Y} = L^p(X)$ where $p \in [1, \infty]$ and suppose $\varphi \in L^\infty(X)$. Then let T be the multiplication operator M_φ on $\mathcal{Y}$.

If $\lambda \in \mathbb{C}$ then $T - \lambda\mathrm{Id} = M_{\varphi - \lambda}$. Therefore, if $(T - \lambda\mathrm{Id})^{-1}$ exists it must equal $M_{1/(\varphi-\lambda)}$. Hence $\lambda \in \mathbb{C} - \sigma(T)$ iff $1/(\varphi(x) - \lambda) \in L^\infty(X)$. But the latter happens iff $\lambda \in \mathbb{C} - R_\varphi$. Therefore, $\sigma(M_\varphi) = R_\varphi$.

If $\mathcal{X}$ is a Banach space and $T \in B(\mathcal{X})$, we say that T is bounded below if there is some $k > 0$ such that $\|Tx\| \geq k\|x\|$ for all $x \in \mathcal{X}$.

Lemma 4.1.5: If $T \in B(\mathcal{X})$, then T is invertible iff T is bounded below and $T(\mathcal{X})$ is dense in $\mathcal{X}$.

$(\Longrightarrow)$
If T is invertible then trivially $T(\mathcal{X})$ is dense in $\mathcal{X}$. Also, we can let $k = 1/\|T^{-1}\|_{\mathrm{op}}$. Then as $\|x\| \leq \|T^{-1}\|_{\mathrm{op}}\|Tx\|$, we have that $k\|x\| \leq \|Tx\|$.

$(\Longleftarrow)$
T being bounded below guarentees that T is injective. Meanwhile, suppose $y \in \mathcal{X}$ and let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{X}$ such that $Tx_n \to y$ as $n \to \infty$. Since $\|Tx_n\|$ is Cauchy and $\|Tx_n - Tx_M\| \geq k\|x_n - x_m\|$, we also have that $\{x_n\}_{n\in\mathbb{N}}$. is Cauchy. Therefore, as $\mathcal{X}$ is complete we have that $x_n \to x$ for some $x \in \mathcal{X}$. Finally, as $T \in B(\mathcal{X})$, we have that $Tx_n \to Tx$. Hence, $Tx = y$ and we've proven that T is surjective.

This proves that T is invertible. And it is easy to see that $\|T^{-1}\|_{\mathrm{op}} \leq k^{-1}$. ■

Lemma 4.1.6: Suppose $\mathcal{H}$ is a Hilbert space and $T \in B(\mathcal{H})$. If T, T^* are bounded below, then T is invertible.

Proof:
It suffices by the prior lemma to show that $T(\mathcal{H})$ is dense in $\mathcal{H}$. Fortunately, recall from the definition of the adjoint of a linear map that $y \in \ker(T^*)$ iff $0 = \langle Tx, y\rangle$ for all $x \in \mathcal{H}$. In other words, we have that $T(\mathcal{H})^\perp = \ker(T^*)$. But since T^* is bounded below, we know that $\ker(T^*) = \{0\}$. It follows by the following exercise from Folland as well as the fact that the closure of a subspace is another subspace that $T(\mathcal{H})$ is dense in $\mathcal{H}$.

(Folland) Exercise 5.56: If E is a subset of a Hilbert space $\mathcal{H}$, then $(E^\perp)^\perp$ is the smallest closed subspace of $\mathcal{H}$ containing E.

Firstly, it's clear that $(E^\perp)^\perp$ is a closed subspace. Furthermore if $x \in E$. Then for any $y \in E^\perp$ we have that $\langle y, x\rangle = 0$. Hence, $(E^\perp)^\perp$ contains E.

Meanwhile, suppose F is any closed subspace of $\mathcal{H}$ containing E. To prove that $(E^\perp)^\perp \subseteq F$, we shall assume for the sake of contradiction that there exists $x \in (E^\perp)^\perp - F$. Then by (Folland) Theorem 5.8(a) in my math 240b notes, you can find $\lambda \in \mathcal{H}^*$ such that $F \subseteq \ker(\lambda)$ but $\lambda(x) \neq 0$. Yet note that there exists $y \in \mathcal{H}$ such that $\lambda(z) = \langle z, y\rangle$ for all $z \in \mathcal{H}$. In particular, as $E \subseteq F$ we must have that $\langle z, y\rangle = \lambda(z) = 0$ for all $z \in E$. So, $y \in E^\perp$. Yet this is a contradiction since $x \in (E^\perp)^\perp$, meaning that:

$$0 = \langle x, y\rangle = \lambda(x) \neq 0. \blacksquare$$

More generally, the second paragraph of this proof can be followed to show that if $\mathcal{H} = V \oplus V^\perp = W \oplus W^\perp$ and $V^\perp \subseteq W^\perp$, then $V \supseteq W$.

For the sake of convenience and readability, I'll use an I to represent the identity map on a Banach space from now on.

Proposition 4.1.7: Let $\mathcal{H}$ be a Hilbert space and $T \in \mathcal{H}$.

(a) If $T = T^*$ then $\sigma(T) \subseteq \mathbb{R}$.

Proof:
By lemma 4.1.6 it suffices to show that if $\operatorname{Im}(\lambda) \neq 0$ then $T - \lambda I$ and $(T - \lambda I)^* = T - \overline{\lambda} I$ are bounded below. So suppose $\|x\| = 1$. Then by the Cauchy-Schwartz inequality we have that:

$$\|(T - \lambda I)(x)\| \geq |\langle (T - \lambda I)x, x\rangle| = |\langle Tx, x\rangle - \lambda\langle x, x\rangle| = |\langle Tx, x\rangle - \lambda|$$

By *lemma 3.2.4(ii)*, we have that $\langle Tx, x\rangle \in \mathbb{R}$. Therefore, we can conclude that $\|(T - \lambda I)x\| \geq |\operatorname{Im}(\lambda)|$ Similarly, we have that $\|(T - \overline{\lambda} I)x\| \geq |-\operatorname{Im}(\lambda)|$. Hopefully from here it's clear that both linear maps are bounded below.

(b) If T is unitary then $\sigma(T) \subseteq S^1 = \{z \in \mathbb{C} : |z| = 1\}$.

Like before it suffices to show that if $|\lambda| \neq 1$ then $T - \lambda I$ and $(T - \lambda I) = T^{-1} - \overline{\lambda} I$ are bounded below. Fortunately, note by the reverse triangle inequality that:

$$\|(T - \lambda I)x\| \geq |\|Tx\| - |\lambda|\,\|x\|| = |1 - |\lambda|| \cdot \|x\|$$

Similarly, we have that $\|(T^{-1} - \overline{\lambda} I)x\| \geq \left|1 - |\overline{\lambda}|\right| \cdot \|x\|$. $\blacksquare$

Proposition 4.1.8: Let $\mathcal{X}$ be a Banach space, $T \in B(\mathcal{X})$ and $p(x) \in \mathbb{C}[x]$. Then:

$$\sigma(p(T)) = p(\sigma(T)).$$

Note that if T is the zero map, then $\sigma(T) = \{0\}$. It follows that the proposition holds when $p(x)$ is the constant zero polynomial. Meanwhile, if $c \neq 0$ then it's easy to see that

$\sigma(cT) = c\sigma(T)$. After all, $T - \lambda I$ is bijective iff $c(T - \lambda I) = cT - c\lambda I$ is bijective. Hence, it suffices to prove the proposition for the case where $p(x)$ is monic.

Suppose $\lambda \in \sigma(p(T))$. Then we know that $p(T) - \lambda I$ is not invertible. Also, if we let $\lambda_1, \ldots, \lambda_n$ be the roots of $p(x) - \lambda$ repeated according to multiplicity. Then we have that:

$$p(T) - \lambda I = \prod_{j=1}^{n}(T - \lambda_j I)$$

It follows that for at least one j that $(T - \lambda_j I)$ is not invertible, for otherwise we'd have that $(p(T) - \lambda I)^{-1} = \prod_{j=1}^{n}(T - \lambda_j I)^{-1}$. But now as $\lambda_j \in \sigma(T)$ and $p(\lambda_j) = \lambda$, we've shown that $\lambda \in p(\sigma(T))$. This proves that $\sigma(p(T)) \subseteq p(\sigma(T))$.

On the other hand, let $\lambda \in p(\sigma(T))$. Hence, for some $\alpha \in \sigma(T)$ we have that $p(\alpha) = \lambda$. In particular, after letting $\lambda_1, \ldots, \lambda_n$ be the roots (repeated according to multiplicity) of $p(x) - \lambda$, we have that $\alpha = \lambda_j$ for some j. And because $\lambda_j = \alpha \in \sigma(T)$, we know that $T - \lambda_j I$ must not be a bijection.

Remember by the open mapping theorem that any bijective bounded linear map between Banach spaces is automatically a homeomorphism.

Suppose the injectivity of $T - \lambda_j I$ fails. Then after writing:

$$p(T) - \lambda I = (T - \lambda_1 I)\cdots(T - \lambda_{j-1} I)(T - \lambda_{j+1} I)\cdots(T - \lambda_n I)(T - \lambda_j I),$$

we can also guarentee that $p(T) - \lambda I$ is also not injective. Hence $\lambda \in \sigma(p(T))$.

Meanwhile, suppose the surjectivity of $T - \lambda_j I$ fails. Then we can similarly write:

$$p(T) - \lambda I = (T - \lambda_j I)(T - \lambda_1 I)\cdots(T - \lambda_{j-1} I)(T - \lambda_{j+1} I)\cdots(T - \lambda_n I),$$

to show that $p(T) - \lambda I$ is not surjective. Hence $\lambda \in \sigma(p(T))$ in this case as well.

Either way, this proves that $p(\sigma(T)) \subseteq \sigma(p(T))$. ■

Let $\Omega \subseteq \mathbb{C}$ be an open set and let $\mathcal{X}$ be a Banach space. Then we say $f : \Omega \to \mathcal{X}$ is analytic if for all $z_0 \in \Omega$ there is some $r > 0$ with $\{z : |z - z_0| < r\} \subseteq \Omega$ and some sequence $\{x_n\}_{n \in \mathbb{N}} \subseteq \mathcal{X}$ such that for all $|z - z_0| < r$ we have that $f(z) = \sum_{n=0}^{\infty}(z - z_0)^n x_n$ where the sum converges absolutely.

By absolute convergence here I mean that $\sum_{n=0}^{\infty} \|(z - z_0)^n x_n\|$ converges. Note that any series converges absolutely also converges nonabsolutely in a Banach space.

There is more depth to go in here. I'll return to this topic on *page* ___.

1/6/2026

I want to start off today by doing an exercise from Folland. The purpose of this exercise is to generalize integration to functions with values in a separable Banach space.

Exercise 5.16: Define the following:

- let $(X, \mathcal{M}, \mu)$ be a measure space,
- let $\mathcal{Y}$ be a *separable* Banach space,
- let $L_{\mathcal{Y}}$ be the space of all $(\mathcal{M}, \mathcal{B}_{\mathcal{Y}})$-measurable maps from X to $\mathcal{Y}$ (where $\mathcal{B}_{\mathcal{Y}}$ is the Borel σ-algebra on $\mathcal{Y}$),
- let $F_{\mathcal{Y}}$ be the set of maps $f : X \to \mathcal{Y}$ of the form $f(x) = \sum_{j=1}^{n} y_j \chi_{E_j}(x)$ where $n \in \mathbb{N}$, $y_j \in \mathcal{Y}$, $E_j \in \mathcal{M}$, and $\mu(E_j) < \infty$.

If $f \in L_{\mathcal{Y}}$, then as $y \mapsto \|y\|$ is continuous we have that $x \mapsto \|f(x)\|$ is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$-measurable. So, we may define $\|f\|_1 = \int \|f(x)\| \mathrm{d}\mu(x)$. Finally, we define $L^1_{\mathcal{Y}} = \{f \in L_{\mathcal{Y}} : \|f\|_1 < \infty\}$.

(a) $L_{\mathcal{Y}}$ is a vector space, $F_{\mathcal{Y}}$ and $L^1_{\mathcal{Y}}$ are subspaces of it, $F_{\mathcal{Y}} \subseteq L^1_{\mathcal{Y}}$, and $\|\cdot\|_1$ is a seminorm on $L^1_{\mathcal{Y}}$ that becomes a norm if we identify two functions that are equal a.e.

Lemma: If $f, g : X \to \mathcal{Y}$ are $(\mathcal{M}, \mathcal{B}_{\mathcal{Y}})$-measurable functions then so is $f + g$. Furthermore, if $h : X \to \mathcal{Y}$ is $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$-measurable then so is hf.

Firstly, note that scalar multiplication and vector addition are continuous and thus $(\mathcal{B}_{\mathbb{C}\times\mathcal{Y}}, \mathcal{B}_{\mathcal{Y}})$- and $(\mathcal{B}_{\mathcal{Y}\times\mathcal{Y}}, \mathcal{B}_{\mathcal{Y}})$-measurable respectively. Next, since $\mathcal{Y}$ is a separable metric space, we know that $\mathcal{Y}$ is second countable. Therefore, by *theorem 7.20* we have that $\mathcal{B}_{\mathbb{C}\times\mathcal{Y}} = \mathcal{B}_{\mathbb{C}} \otimes \mathcal{B}_{\mathcal{Y}}$ and $\mathcal{B}_{\mathcal{Y}\times\mathcal{Y}} = \mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Y}}$.

Finally, the rest of the proof of this lemma follows via the same reasoning as a proposition on page 42 of my math 240a latex notes.

The claim that $L_{\mathcal{Y}}$ is a vector space easily follows from that lemma. Then everything else is obvious enough that I don't want to write out the proof.

(b) Let $\{y_n\}_{n\in\mathbb{N}}$ be a countable dense set in $\mathcal{Y}$. Given any $n \in \mathbb{N}$ and $\varepsilon > 0$, let:

$$B_n^{\varepsilon} := \{y \in \mathcal{Y} : \|y - y_n\| < \varepsilon\|y_n\|\}$$

Then $\bigcup_{n=1}^{\infty} B_n^{\varepsilon} \supseteq \mathcal{Y} - \{0\}$.

Proof:
Fix $y \in \mathcal{Y} - \{0\}$ and then pick $\alpha \in \mathbb{R}$ such that $\max(0, \|y\| - \varepsilon) < \alpha < \|y\|$.
Then since $\{y_n\}_{n\in\mathbb{N}}$ is a dense set in $\mathcal{Y}$, we can find some n such that $\|y - y_n\| < \varepsilon\alpha$ and $\alpha < \|y_n\|$. So, $\|y - y_n\| < \varepsilon\|y_n\|$, meaning $y \in B_n^{\varepsilon}$.

As a side note, if $y_n \neq 0$ for any n and $\varepsilon < 1$, then 0 won't be in $\bigcup_{n=1}^{\infty} B_n^{\varepsilon}$.

(c) If $f \in L^1_{\mathcal{Y}}$, there is a sequence $\{h_n\}_{n\in\mathbb{N}} \subseteq F_{\mathcal{Y}}$ with $h_n \to f$ pointwise and $\|h_n - f\|_1 \to 0$.

Proof:
Let $A_{n,j} := B_n^{1/j} - \bigcup_{m=1}^{n-1} B_m^{1/j}$ for each $n, j \in \mathbb{N}$. Then set $E_{n,j} := f^{-1}(A_{n,j})$ and consider $g_j(x) = \sum_{n=1}^{\infty} y_n \chi_{E_{n,j}}(x)$.

Notice that for any j the sequence of sets $\{E_{n,j}\}_{n\in\mathbb{N}}$ forms a partition of $X - f^{-1}(\{0\})$ consisting of measurable sets. Hence, each g_j is measurable and well-defined, and if $f(x) \neq 0$, we have that:

$$\|g_j(x)\| \leq \|g_j(x) - f(x)\| + \|f(x)\| < \tfrac{1}{j}\|g_j(x)\| + \|f(x)\|.$$

In turn, we have that $(1 - \frac{1}{j})\|g_j(x)\| \leq \|f(x)\|$. Hence, if $f(x) \neq 0$ then:

$$\|f(x) - g_j(x)\| < \tfrac{1}{j}\|g_j(x)\| \leq \tfrac{1}{j} \cdot \tfrac{\|f(x)\|}{1-\frac{1}{j}} < \tfrac{\|f(x)\|}{j-1}$$

By the side note in part (b) plus the main statement of part (b), we also know that when $j > 1$ then $g_j(x) = 0$ iff $f(x) = 0$. Hence, the above inequality also holds for $x \in f^{-1}(\{0\})$. Consequently:

- $g_j \to f$ pointwise as $j \to \infty$.
- Each $g_j \in L^1_{\mathcal{Y}}$ (where $j > 1$) with $\|g_j\|_1 \leq (1 - \frac{1}{j})^{-1}\|f\|_1$
- $\|g_j - f\|_1 \leq \frac{1}{j-1}\|f\|_1 \to 0$ as $j \to \infty$.

To finish off, note that the set where $f(x) \neq 0$ must be σ-finite since each $E_{n,j}$ has finite measure (or else $f \notin L^1_{\mathcal{Y}}$) and we already established before that $\{E_{n,j}\}_{n\in\mathbb{N}}$ forms a partition of $X - f^{-1}(\{0\})$ for each j. So, let us fix some sequence $\{F_n\}_{n\in\mathbb{N}}$ of disjoint sets of finite measure partitioning $X - f^{-1}(\{0\})$.

Next, note for each j that $\|g_j\|_1 = \sum_{n=1}^{\infty} \mu(E_{n,j})\|y_n\|$. In particular, this means that there is some $M_j > 0$ such that $\sum_{n>m_j} \mu(E_{n,j})\|y_n\| < \frac{1}{j}$ for all $m_j \geq M_j$.

Meanwhile, note for each j that $\bigcup_{n=1}^{j} F_n \subseteq \bigcup_{n=1}^{\infty} E_{n,j}$. And since $\mu(\bigcup_{n=1}^{j} F_n) < \infty$ we can thus find a $K_j > 0$ such that $\mu(\bigcup_{n=1}^{j} F_n - \bigcup_{n=1}^{k_j} E_{n,j}) < \frac{1}{j}$ for all $k_j \geq K_j$.

So, for all j let $h_j(x) = \sum_{n=1}^{N_j} y_n \chi_{E_{n,j}}(x)$ where $N_j = \max(M_j, K_j)$. Then it is clear that each h_j is in $F_{\mathcal{Y}}$ and that:

$$\|h_j - f\|_1 \leq \|h_j - g_j\|_1 + \|g_j - f\|_1 < \tfrac{1}{j} + \|g_j - f\|_1 \to 0 \text{ as } j \to \infty.$$

Furthermore, it's clear that $h_j \to f$ a.e. on each F_n individually. And since $h_j(x) = 0 = f(x)$ for all $x \notin \bigcup_{n=1}^{\infty} F_n$, we also know that $h_j \to f$ pointwise everwhere on $X - \bigcup_{n=1}^{\infty} F_n$.

(d) There is a unique linear map $\int : L^1_{\mathcal{Y}} \to \mathcal{Y}$ such that $\int y\chi_E = \mu(E)y$ for all $y \in \mathcal{Y}$ and $E \in \mathcal{M}$ with $\mu(E) < \infty$, and which also satisfies that $\|\int f\| \leq \|f\|_1$.

To start out, there is clearly a unique linear map $\int : F_{\mathcal{Y}} \to \mathcal{Y}$ satisfying that $\int y\chi_E = \mu(E)y$ for all $y \in \mathcal{Y}$ and $E \in \mathcal{M}$ with $\mu(E) < \infty$. Namely, given any $f(x) = \sum_{j=1}^{n} y_j\chi_{E_j}(x) \in F_{\mathcal{Y}}$ we define $\int f(x)\mathrm{d}\mu(x) = \sum_{j=1}^{n} \mu(E_j)y_j$.

The proof that $\int$ is well-defined here is identical to the proof that the normal Lebesgue integral is well-defined.

Furthermore, by the triangle inequality we can easily see that $\|\int f\| \leq \|f\|_1$ when $f \in F_{\mathcal{Y}}$. It follows that $\int$ is a continuous linear map on $F_{\mathcal{Y}}$ with an operator norm of at most 1. Finally, since we showed in part (c) that $F_{\mathcal{Y}}$ is dense in $L^1_{\mathcal{Y}}$ with respect to the $\|\cdot\|_1$-norm, we know by standard arguments there is a unique continuous extension of $\int$ to all of $L^1_{\mathcal{Y}}$. Namely, for any $f \in L^1_{\mathcal{Y}}$ we define $\int f = \lim_{n\to\infty} \int f_n$ where $\{f_n\}_{n\in\mathbb{N}}$ is any sequence in $F_{\mathcal{Y}}$ such that $\|f_n - f\|_1 \to 0$ as $n \to \infty$.

(e) The dominated convergence theorem: If $\{f_n\}_{n\in\mathbb{N}}$ is a sequence in $L^1_{\mathcal{Y}}$ such that $f_n \to f$ a.e. and there exists $g \in L^1(X)$ such that $\|f_n(x)\| \leq g(x)$ for all n and a.e. x, then $\int f_n \to \int f$.

Note that $\|f_n(x) - f(x)\| \leq \|f_n(x)\| + \|f(x)\| \leq 2g(x)$ for all n and a.e. x. Also, we already know from the problem statement that $\|f_n(x) - f(x)\| \to 0$ a.e. Hence, by the standard dominated convergence theorem we know that:

$$\|f_n - f\|_1 = \int \|f_n(x) - f(x)\| \mathrm{d}\mu(x) \to \int 0 = 0 \text{ as } n \to \infty.$$

In other words, $f_n \to f$ with respect to the $\|\cdot\|_1$-norm. It now follows by the continuity of the $\int$ operator that $\int f = \lim_{n\to\infty} \int f_n$.

(f) If $\mathcal{Z}$ is another separable Banach space, $T \in B(\mathcal{Y}, \mathcal{Z})$, and $f \in L^1_{\mathcal{Y}}$, then $T \circ f \in L^1_{\mathcal{Z}}$ and $\int T \circ f = T(\int f)$.

To start off, note that as T is $(\mathcal{B}_{\mathcal{Y}}, \mathcal{B}_{\mathcal{Z}})$-measurable, we know that $T \circ f$ is $(\mathcal{M}, \mathcal{B}_{\mathcal{Z}})$-measurable. Also, $T \circ f \in L^1_{\mathcal{Z}}$ since:

$$\|T \circ f\|_1 = \int \|T(f(x))\|_{\mathcal{Z}} \mathrm{d}\mu(x) \leq \|T\|_{\mathrm{op}} \int \|f(x)\|_{\mathcal{Y}} \mathrm{d}\mu(x) = \|T\|_{\mathrm{op}} \|f\|_1.$$

As for the other claim, note that $\int T \circ f = T(\int f)$ is clearly true when $f \in F_{\mathcal{Y}}$. Then since $F_{\mathcal{Y}}$ is dense in $L^1_{\mathcal{Y}}$ with respect to $\|\cdot\|_1$ and both $\int$ and T are continuous with respect to $\|\cdot\|_1$, we also have that:

$$T(\textstyle\int f) = \lim_{n\to\infty} T(\int f_n) = \lim_{n\to\infty} \int T \circ f_n = \int T \circ f$$

where $\{f_n\}_{n\in\mathbb{N}}$ is any sequence in $F_{\mathcal{Y}}$ with $\|f_n - f\|_1 \to 0$ as $n \to \infty$.

The integral established in the above problem is called the Bochner integral. It's not the only integral one can define for certain vector-valued functions but it is notable for basically being a generalization of the Lebesgue integral.

Math 200b Lecture Notes:

For a refresh on what last happened in math 200a, go to *page 491*. In this class assume all rings are unital and commutative unless stated otherwise.

We say a ring A is a finitely generated (f.g.) ring if there exists $a_1, \ldots, a_n \in A$ such that:

$$A = \{ \sum_{I=(i_1,\ldots,i_n)} c_I a_1^{i_1} \cdots a_n^{i_n} : c_I \in \mathbb{Z}1_A, \text{ each } i_j \in \mathbb{Z}^{\geq 0}, \text{ and all but finitely many } c_I = 0\}.$$

In other words, the ring homomorphism $\mathbb{Z}[x_1, \ldots, x_n] \to A$ defined by mapping $x_j \mapsto a_j$ and $N \mapsto N1_A$ is surjective.

Corollary: Every f.g. ring is Noetherian.

Proof:
We know there is a surjective ring homomorphism $f : \mathbb{Z}[x_1, \ldots, x_n]$ to A for some n. Futhermore we know $\mathbb{Z}[x_1, \ldots, x_n]$ is Noetherian by Hilbert's basis theorem. ($\mathbb{Z}$ is easily checked to be a P.I.D. and is thus Noetherian). Finally, if $\mathfrak{a}$ is any ideal in A, then we know $f^{-1}(\mathfrak{a})$ is an ideal in $\mathbb{Z}[x_1, \ldots, x_n]$. In turn, there are finitely many polynomials $p_1, \ldots, p_m \in \mathbb{Z}[x_1, \ldots, x_n]$ generating $f^{-1}(\mathfrak{a})$. So, $f(p_1), \ldots, f(p_m)$ generate $\mathfrak{a}$. ■

Suppose F is a field. Then A is a called an F-algebra if A is a unital commutative ring and there exists a ring homomorphism $f : F \to A$ such that $f(1_F) = f(1_A)$. Note that if A is an F-algebra then we also have that A is an F-vector space with $c \cdot a = f(c)a$ for all $c \in F$ and $a \in A$.

I go into more depth on F-algebras on *page 602*.

Let $f : F \to A$ be the ring homomorphism given to define A as an F-algebra. We say A is a f.g. F-algebra if $\exists a_1, \ldots, a_n$ such that:
$A = \{ \sum_{I=(i_1,\ldots,i_n)} c_I a_1^{i_1} \cdots a_n^{i_n} : c_I \in f(F),$ each $i_j \in \mathbb{Z}^{\geq 0}$, and all but finitely many $c_I = 0\}$.

In other words, like before the ring homomorphism $F[x_1, \ldots, x_n] \to A$ defined by mapping $x_j \mapsto a_j$ and $c \mapsto f(c)$ is surjective.

By identical reasoning to the corollary on the last page, we have that if A is a f.g. F-algebra then A is Noetherian.

Let D be a U.F.D. Our next goal will be to show that $D[x]$ is also a U.F.D., and to do that we will need to first build out some infrastructure.

Firstly, if $p \in D$ is irreducible then we define $\nu_p(a) := \max\{n \in \mathbb{Z}_{\geq 0} : p^n \mid a\}$ for all $a \in D - \{0\}$. Also for the sake of convenience we define $\nu_p(0) = \infty$.

Lemma: ν_p is well-defined and $\nu_p(ab) = \nu_p(a)\nu_p(b)$.

The lemma is trivially true if either a or b is zero. So assume $a, b \neq 0$.

ν_p is well-defined because D is a U.F.D. and so every nonzero element in D has a unique factorization in terms of irreducibles (we consider a unit as having a factorization with zero terms). Also note that since D is a U.F.D and p is irreducible, we also know that p is prime. Hence, if $a = p^{\nu_p(a)}x$ and $b = p^{\nu_p(b)}y$ then p doesn't divide xy. And since $ab = p^{\nu_p(a)+\nu_p(b)}xy$ this proves that $\nu_p(ab) = \nu_p(a)\nu_p(b)$. ■

Let $\mathcal{P}_0 \subseteq D$ contain precisely one element from each equivalence class of companions (see *page 458*) containing irreducible elements. Then for all $a \in D - (\{0\} \cup D^\times)$ we define $|a| := \prod_{p \in \mathcal{P}_0} p^{\nu_p(a)}$ and let $\text{sgn}(a)$ denote the element in $D^\times$ such that $a = \text{sgn}(a)|a|$. It's easy to see based on the last lemma that $\text{sgn}(ab) = \text{sgn}(a)\text{sgn}(b)$ and $|ab| = |a||b|$.

Lemma: For any $p \in \mathcal{P}_0$ we have that $\nu_p(a + b) \geq \min(\nu_p(a), \nu_p(b))$.

Proof:

Write $a = \text{sgn}(a) \prod_{p \in \mathcal{P}_0} p^{\nu_p(a)}$ and $b = \text{sgn}(b) \prod_{p \in \mathcal{P}_0} p^{\nu_p(b)}$. Then $p^{\min(\nu_p(a), \nu_p(b))} \mid (a + b)$.

For some motivation on why we'd define $\mathcal{P}_0$, note that we already implicitely picked out such a collection when working with irreducible elements of $\mathbb{Z}$. After all, when talking about prime integers we usually assume they are positive instead of negative.

Next we define a $\gcd$ function.

Lemma: $a \mid b$ iff $\nu_p(a) \leq \nu_p(b)$ for all $p \in \mathcal{P}_0$.

($\Longrightarrow$)
Write $b = ac$. Then $\nu_p(b) = \nu_p(a) + \nu_p(c) \geq \nu_p(a)$ for all $p \in \mathcal{P}_0$.

($\Longleftarrow$)
We can assume that $b \neq 0$. In turn, $a \neq 0$ since $\nu_p(a) \leq \nu_p(b) < \infty$ for all $p \in \mathcal{P}_0$. Finally, note that:

$$b = \mathrm{sgn}(b) \prod_{p \in \mathcal{P}_0} p^{\nu_p(b)} = (\mathrm{sgn}(b)\mathrm{sgn}(a)^{-1} \prod_{p \in \mathcal{P}_0} p^{\nu_p(b) - \nu_p(a)})a.$$

Now let $\gcd(a_1, \ldots, a_n) \coloneqq \prod_{p \in \mathcal{P}_0} p^{\min(\nu_p(a_1), \ldots, \nu_p(a_n))}$.

Lemma:

- $\gcd(a_1, \ldots, a_n) \mid a_i$ for all i.
- $d \mid a_1, \ldots, d \mid a_n \Longrightarrow d \mid \gcd(a_1, \ldots, a_n)$.
- $\gcd(ca_1, \ldots, ca_n) = |c| \gcd(a_1, \ldots, a_n)$
- $\gcd(a_1, \ldots, a_n) = d \Longrightarrow \gcd(\frac{a_1}{d}, \ldots, \frac{a_n}{d}) = 1$ (where $\frac{a_i}{d}$ is the unique element such that $d(\frac{a_i}{d}) = a_i$).

Hopefully all these things are obvious.

If $f(x) = a_n x^n + \ldots + a_0 \in D[x] - \{0\}$, let $\alpha(f) \coloneqq \gcd(a_0, \ldots, a_n)$. $\alpha(f)$ is called the content of f. Also, if $\alpha(f) = 1$ we say that f is primitive.

Observations:

- For all $f(x) \in D[x] - \{0\}$ we have that $f(x) = \alpha(f) \cdot f_{\prime}(x)$ (where f_{prim} is a primitive polynomial).
- $\alpha(cf) = |c|\alpha(f)$ for all $c \in D - \{0\}$ and $f \in D[x] - \{0\}$.
- Let $c_a : D[x] \to (D/\langle a \rangle)[x]$ be the map sending f to $f \pmod a$. (see *the problem on pages 459-461*). Then $c_p(f) = 0$ iff $p \mid \alpha(f)$.

Gauss's Lemma v1: If $f, g \in D[x]$ are primitive then fg is also primitive.

Suppose f and g are primitive but that $\alpha(fg) \neq 1$. Then there exists $p \in \mathcal{P}_0$ such that $p \mid \alpha(fg)$ and in turn we know that $c_p(fg) = 0$. But note that that means $c_p(f)c_p(g) = c_p(fg) = 0$. Also note that as p is prime (since D is a U.F.D.), we know that $\langle p \rangle$ is a prime ideal in D. Hence, $D/\langle p \rangle$ and in turn $(D/\langle p \rangle)[x]$ is an integral domain. That means either $c_p(f)$ or $c_p(g)$ must equal 0. And now we have a contradiction as either $p \mid \alpha(f)$ or $p \mid \alpha(g)$. ■

Gauss's Lemma v2: $\alpha(fg) = \alpha(f)\alpha(g)$.

$f = \alpha(f) f_{\mathrm{prim}}$ and $g = \alpha(g)g$, implies that $fg = \alpha(f)\alpha(g) f_{\mathrm{prim}} g_{\mathrm{prim}}$. Taking the contents of both sides, we get that $\alpha(fg) = |\alpha(f)\alpha(g)|\alpha(f_{\mathrm{prim}} g_{\mathrm{prim}})$. But now note by definition that $|\alpha(f)\alpha(g)| = \alpha(f)\alpha(g)$. Furthermore, by Gauss's lemma v1 we have that $\alpha(f_{\mathrm{prim}} g_{\mathrm{prim}}) = 1$. So, $\alpha(fg) = \alpha(f)\alpha(g)$. ■

Next let F be the field of fractions of D. For a reminder of what that is, see my math 100b notes.

Gauss's Leamm v3: Suppose $f(x) = g_1(x)g_2(x)$ where $g_1(x), g_2(x) \in F[x]$ and $f(x) \in D[x]$. Then there exists $c_1, c_2 \in F^\times$ such that $c_1g_1, c_2g_2 \in D[x]$ and $c_1c_2 = 1$.

Proof:

There exists $d_1, d_2 \in D - \{0\}$ such that $\widetilde{g}_1(x) := d_1g_1(x)$ and $\widetilde{g}_2(x) := d_2g_2(x)$ are in $D[x]$. In turn:

$$d_1d_2\alpha(f)f_{\text{prim}}(x) = d_1d_2f(x) = \widetilde{g}(x)_1\widetilde{g}_2(x) = \alpha(\widetilde{g}_1)\alpha(\widetilde{g}_2)\widetilde{g}_{1,\text{prim}}(x)\widetilde{g}_{2,\text{prim}}(x).$$

Taking the contents of both sides we get that $|d_1d_2|\alpha(f) = \alpha(\widetilde{g}_1)\alpha(\widetilde{g}_2)$. Thus, after writing $d_1d_2 = \text{sgn}(d_1d_2)|d_1d_2|$ we can say that:

$$\begin{aligned}\text{sgn}(d_1d_2)|d_1d_2|\alpha(f)f_{\text{prim}}(x) &= d_1d_2\alpha(f)f_{\text{prim}}(x)\\ &= \alpha(\widetilde{g}_1)\alpha(\widetilde{g}_2)\widetilde{g}_{1,\text{prim}}(x)\widetilde{g}_{2,\text{prim}}(x)\\ &= |d_1d_2|\alpha(f)\widetilde{g}_{1,\text{prim}}(x)\widetilde{g}_{2,\text{prim}}(x)\end{aligned}$$

After canceling out terms and rearranging units, this means that:

$$f_{\text{prim}}(x) = \text{sgn}(d_1d_2)^{-1}\widetilde{g}_{1,\text{prim}}(x)\widetilde{g}_{2,\text{prim}}(x).$$

So, $f(x) = \text{sgn}(d_1d_2)^{-1}\alpha(f)\widetilde{g}_{1,\text{prim}}(x)\widetilde{g}_{2,\text{prim}}(x)$ where $\text{sgn}(d_1d_2)^{-1}\alpha(f) \in D$, and $\widetilde{g}_{1,\text{prim}}, \widetilde{g}_{2,\text{prim}} \in D[x]$.

Finally, note that $\alpha(\widetilde{g}_i)\widetilde{g}_{i,\text{prim}} = d_ig_i(x)$. In turn, $\widetilde{g}_{i,\text{prim}}(x) = \frac{d_i}{\alpha(\widetilde{g}_i)}g_2(x)$. Hence, let:

$$c_1 = \tfrac{\alpha(f)\text{sgn}(d_1d_2)^{-1}d_1}{\alpha(\widetilde{g}_1)} \text{ and } c_2 = \tfrac{d_2}{\alpha(\widetilde{g}_2)}.$$

Then $c_1g_1(x) = \text{sgn}(d_1d_2)^{-1}\alpha(f)\widetilde{g}_{1,\text{prim}}(x) \in D[x]$, $c_2g_2(x) = \widetilde{g}_{2,\text{prim}}(x) \in D[x]$, and $c_1c_2 = \frac{\alpha(f)\text{sgn}(d_1d_2)^{-1}d_1d_2}{\alpha(\widetilde{g}_1)\alpha(\widetilde{g}_2)} = \frac{\alpha(f)|d_1d_2|}{\alpha(\widetilde{g}_1)\alpha(\widetilde{g}_2)} = 1$. ■

Corollary: Suppose D is a U.F.D., $F := Q(D)$ (i.e. F is the field of fractions of D), $f(x) \in D[x]$, $g_1, \ldots, g_n \in F[x]$, and $f(x) = \prod_{i=1}^n g_i(x)$. Then there exists constants $c_1, \ldots, c_n \in F^\times$ such that $\prod_{i=1}^n c_i = 1$ and $c_ig_i(x) \in D[x]$ for all i.

Proof:

To start off, use the prior lemma to get two constants $a_1, a_2 \in F^\times$ such that $a_1\prod_{i=1}^{n-1} g_i(x)$ and $a_2g_n(x)$ are in $D[x]$ and $a_1a_2 = 1$. Next, use an inductive hypothesis to find constants $b_1, \ldots, b_{n-1}$ such that $b_1a_1g_1(x) \in D[x]$, $b_ig_i(x) \in D[x]$ for all $i \in \{2, \ldots, n-1\}$, and $\prod_{i=1}^{n-1} b_i = 1$. Now we can easily see that:

$c_1 = b_1a_1$, $c_n = a_2$ and $c_i = b_i$ for $i \in \{2, \ldots, n-1\}$ satisfy our corollary. ■

I guess I should note that in the above proofs I and my professor were implicitely assuming $f \neq 0$. That said, the case where $f = 0$ is trivial since in that case we know one $g_i = 0 \in D[x]$ as well (since D and thus $D[x]$ is an integral domain). In turn, we can just multiply all the other g_i by whatever values makes them in $D[x]$.

The big use case of the prior corollary is that it tells us that if we can factor an element of $D[x]$ using polynomials in $F[x]$, then we can also factor that element using polynomials in $D[x]$. This is especially useful because $F[x]$ is a P.I.D. and thus automatically a U.F.D. (see *page 489*). Hence, we know any polynomial in $D[x] \subseteq F[x]$ can be factored uniquely into a product of irreducible polynomials in $F[x]$.

The last task we need to do before proving our desired theorem is that we need to connect irreducibility in $F[x]$ to irreducibility in $D[x]$. As it turns out, we can make this connection sometimes (such as when f is primitive) and rely on a different trick when f is a constant.

Lemma: For all $a \in D$:

1. a is irreducible in D iff a is irreducible in $D[x]$.
2. a is prime in D iff a is prime in $D[x]$.

Proof:
First we show claim 2. Note that a is prime in D iff $D/\langle a\rangle$ is an integral domain (and $\langle a\rangle \neq \{0\}$). But that happens iff $(D/\langle a\rangle)[x] \cong D[x]/\langle a\rangle$ is an integral domain (and $\langle a\rangle \neq \{0\}$), which in turn happens iff $\langle a\rangle \lhd D[x]$ is prime and nonzero, and that is true iff a is prime in $D[x]$.

Next we show claim 1. To start off, since $D[x]^\times = D^\times$ and every non-unit of D is in $D[x]$, it is clear that irreducibility in $D[x]$ implies irreducibility in D. As for the other direction, note that as D is a U.F.D., we have that a being irreducible in D implies a is prime in D. Then by claim 2 we know that a is prime in $D[x]$, and since $D[x]$ is a domain this means that a is irreducible in $D[x]$. $\blacksquare$.

Lemma: If $f(x) \in D[x]$ is primitive and $g(x) \in D[x]$, then $f \mid g$ in $F[x]$ iff $f \mid g$ in $D[x]$.

$(\Longleftarrow)$
This is trivial.

$(\Longrightarrow)$
Suppose $q(x) \in F[x]$ satisfies that $g(x) = f(x)q(x)$. Then pick $d \in D - \{0\}$ such that $\widetilde{q}(x) := dq(x)$ is in $D[x]$. In turn:

$$d\alpha(g)g_{\text{prim}}(x) = dg(x) = f(x)\widetilde{q}(x) = f(x)\alpha(\widetilde{q})\widetilde{q}(x)$$

Taking the contents of both sides we get that $|d|\alpha(g) = \alpha(\widetilde{q})$. Thus as $d = |d|\mathrm{sgn}(d)$ and $\mathrm{sgn}(d) \in D^\times$, we can perform manipulations like in the proof of guass's lemma v3 to say that $g_{\text{prim}}(x) = \mathrm{sgn}(d)^{-1}f(x)\widetilde{q}_{\text{prim}}(x)$.

Thus $g(x) = f(x) \cdot (\alpha(g)\mathrm{sgn}(d)^{-1}\widetilde{q}_{\text{prim}}(x))$ where the latter polynomial is in $D[x]$. This proves that $f \mid g$ in $D[x]$. $\blacksquare$

Two comments:

- If f is not primitive then this lemma is not true. For example $2x \mid x$ in $\mathbb{Q}[x]$ but $2x \nmid x$ in $\mathbb{Z}[x]$.
- Suppose $f(x)$ is a primitive polynomial in $D[x]$. The obvious corollary of the above lemma is that $f(x)$ is a prime element in $F[x]$ iff $f(x)$ is a prime element in $D[x]$. But note also that as $F[x]$ is a U.F.D., we have that any irreducible element in $F[x]$ is also prime in $F[x]$. Also, as $D[x]$ is a domain we have that any prime element in $D[x]$ is irreducible in $D[x]$. So, we know (if $f(x)$ is primitive) that being irreducible in $F[x]$ implies being irreducible in $D[x]$.

Lemma: If $f(x) \in D[x] - D$ then $f(x)$ being irreducible in $D[x]$ implies $f(x)$ is irreducible in $F[x]$.

Proof:

Suppose $f(x) = g(x)h(x)$ where $g(x), h(x) \in F[x]$. By Gauss's lemma v3 we can find constants $c_1, c_2 \in F^\times$ such that $f(x) = c_1 g(x) \cdot c_2 h(x)$ and $c_1 g(x), c_2 h(x) \in D[x]$. But now by the irreducibility of f in $D[x]$, we know that either $c_1 g(x)$ or $c_2 h(x)$ must be in $D^\times$. In turn, either $g(x)$ or $h(x)$ must be a constant polynomial. ■

Finally, we are ready to prove our desired theorem.

Theorem: If D is a U.F.D. then so is $D[x]$.

Proof:

By a theorem on *page 486*, it suffices to show:

1. Every $f(x) \in D[x] - (D^\times \cup \{0\})$ can be factored into a product of irreducible elements;
2. Every irreducible element of $D[x]$ is prime.

We shall start by showing the first requirement.

Case 1: Suppose $\deg(f) = 0$. Then since D is a U.F.D., we know that $f = \prod_{i=1}^{n} p_i$ where all $p_i \in D$ are irreducible. By one of the prior lemmas, we then know all p_i are also irreducible in $D[x]$.

Case 2: Suppose $\deg(f) > 0$ and then write $f(x) = \alpha(f) f_{\text{prim}}(x)$.

Since $f_{\text{prim}}(x) \in F[x] - F$ and $F[x]$ is a U.F.D., we can write $f_{\text{prim}}(x) = \prod_{i=1}^{n} g_i(x)$ where each $g_i(x) \in F[x]$ is irreducible. By the corollary to Gauss's lemma v3, we can actually assume each $g_i(x) \in D[x]$.

Moreover, note that as $1 = \alpha(f_{\text{prim}}) = \prod_{i=1}^{n} \alpha(g_i)$ and the only unit by definition that $\alpha(h)$ can be is 1 for any $h(x) \in D[x]$, we have that $\alpha(g_i) = 1$ for all i. Hence, all g_i are primitive and we can conclude by a prior lemma that they are irreducible in $D[x]$.

Therefore, we have written $f(x) = \alpha(f) \prod_{i=1}^{n} g_i(x)$ where $g_i(x)$ is irreducible in $D[x]$ for all i. To finish proving case 2, we just need to apply case 1 to $\alpha(f)$.

Next we show every irreducible element of $D[x]$ is prime.

Case 1: Suppose $\deg(f) = 0$. Then as $f \in D$, we have that f is irreducible in D. In turn, as D is a U.F.D. we have that f is prime in D, which in turn means that f is prime in $D[x]$.

Case 2: Suppose $\deg(f) > 0$ and then write $f(x) = \alpha(f) f_{\text{prim}}(x)$.

Since $f(x)$ is irreducible in $D[x]$ and $f_{\text{prim}}(x)$ is not a unit (since it has degree > 0), we must have that $\alpha(f) \in D[x]^\times = D^\times$. But by definition the only unit $\alpha(f)$ can be is 1. Hence, we've proven that f is primitive.

Now as $f(x)$ is irreducible in $D[x]$ and not a constant polynomial, we know by a prior lemma that $f(x)$ is irreducible in $F[x]$. In turn, $f(x)$ is prime in $F[x]$ since $F[x]$ is a U.F.D. And finally, by one of the prior lemmas we can conclude that $f(x)$ is prime in $D[x]$. ■

Before moving on to other topics, here are two irreducibility criterion.

<u>$\pmod p$ Irreducibility Criterion:</u> Suppose D is a U.F.D. and $F = Q(D)$ (i.e. F is the field of fractions of D). If $f(x) = a_n x^n + \ldots + a_0 \in D[x]$, $p \in D$ is irreducible, $p \nmid a_n$, and $f \pmod p$ is irreducible in $(D/\langle p \rangle)[x]$, then f is irreducible in $F[x]$.

Proof:
Suppose f is not irreducible in $F[x]$. Then $f(x) = g_1(x)g_2(x)$ for some $g_1, g_2 \in F[x]$ with degrees > 0. But note by Gauss's lemma v3 that we can actually assume $g_1, g_2 \in D[x]$.

Now in particular we have that $a_n = \mathrm{lt}(g_1)\mathrm{lt}(g_2)$. After applying the mod p ring homomorphism we have that $f \pmod p = (g_1 \pmod p)(g_2 \pmod p)$. In turn:

$$0 \pmod p \neq a_n \pmod p = (\mathrm{lt}(g_1) \pmod p)(\mathrm{lt}(g_2) \pmod p).$$

This shows that $p \nmid \mathrm{lt}(g_i)$ for both i and in turn that $\deg(g_i \pmod p) > 0$ for both i. But now since $(D/\langle p \rangle)[x]$ is an integral domain (since p being irreducible in D a U.F.D. means p is prime in D, hence making it so that $D/\langle p \rangle$ is an integral domain), we know neither g_1 nor g_2 is a unit in $(D/\langle p \rangle)[x]^\times = (D/\langle p \rangle)^\times$. This contradicts the assumption that $f \pmod p$ is irreducible. ■

<u>Eisenstein's Irreducibility Criterion:</u> Suppose D is a U.F.D. and $F = Q(D)$ (i.e. F is the field of fractions of D). If $f(x) = a_n x^n + \ldots + a_0 \in D[x]$, $p \in D$ is irreducible, $p \nmid a_n$, $p \mid a_{n-1}, \ldots, a_0$, and $p^2 \nmid a_0$, then f is irreducible in $F[x]$.

Proof:
Suppose f is not irreducible in $F[x]$. Then $f(x) = g_1(x)g_2(x)$ for some $g_1, g_2 \in F[x]$ with degrees > 0. Also like before, by Gauss's lemma v3 we can assume $g_1, g_2 \in D[x]$. By applying the mod p ring homomorphism, we thus get that:

$$f \pmod p = (g_1 \pmod p)(g_2 \pmod p).$$

But note by assumption that $f \pmod p = a_n x^n \pmod p$.

<u>Definition:</u> If D is an integral domain and $p \in D$ is prime, then we know that $D/\langle p \rangle$ is an integral domain as well. We call the field of fractions of $D/\langle p \rangle$ the <u>residue field</u> of $\langle p \rangle$ and denote it $k(p) = Q(D/\langle p \rangle)$.

Consider the equation $a_n x^n \pmod p = (g_1 \pmod p)(g_2 \pmod p)$ as being a factorization in $k(p)[x]$. Since $k(p)$ is a field, we know that $k(p)[x]$ is a P.I.D. and in turn a U.F.D. Also, x is an irreducible polynomial in $k(p)[x]$. So, by considering the unique factorization of $a_n x^n \pmod p$ in terms of irreducibles, and also by noting that all coefficients of each $g_i \pmod p$ are in $D/\langle p \rangle$, we can conclude that:

$g_1 \pmod p = c_1 x^{k_1}$ and that $g_2 \pmod p = c_2 x^{k_2}$ for some $c_1, c_2 \in D/\langle p \rangle$.

One more observation to make is that like in the proof of the prior criterion, we know that p can't divide the leading term of either g_i. In particular, this means that $\deg(g_i \pmod p) = \deg(g_i) > 0$. In particular, this means that $k_1, k_2 > 0$.

But now we've proven that p divides the constant terms of both g_1 and g_2. Since those terms get multiplied together to give the constant term of f, we've shown that p^2 divides the constant term of f. This contradicts the hypothesis of the criterion. ■

1/10/2026

Finishing up week 1 math 200b notes:

There are two big motivations for modules. Firstly, we can study modules for the sake of generalizing linear algebra to situations where our scalars are in a ring instead of a field. Secondly, we can say that modules are where rings act. Hence, modules are to rings like group actions are to groups.

Suppose A is a nonzero unital ring. For now we'll keep assuming A is commutative (although the study of left-modules and right-modules for noncommutative rings is also important). We say M is an A-module if:

(1.) $(M, +)$ is an abelian group;

(2.) There is a map $\cdot : A \times M \to M$ taking $(a, m) \mapsto a \cdot m$ with the following properties:

i. $1_A \cdot m = m$ for all $m \in M$;
ii. $a \cdot (m_1 + m_2) = (a \cdot m_1) + (a \cdot m_2)$ for all $a \in A$ and $m_1, m_2 \in M$;
iii. $(a_1 + a_2) \cdot m = (a_1 \cdot m) + (a_2 \cdot m)$ for all $a_1, a_2 \in A$ and $m \in M$.
iv. $(a_1 a_2) \cdot m = a_1 \cdot (a_2 \cdot m)$ for all $a_1, a_2 \in A$ and $m \in M$.

(If A is not commutative, then these same conditions define a left A-module. Meanwhile, M is a right A-module if property iv. is replaced with $(a_1 a_2) \cdot m = a_2 \cdot (a_1 \cdot M)$. But in that case, we typically write $m \cdot a$ instead of $a \cdot m$.)

Examples:

- If F is a field then M is an F-module iff M is an F-vector space.
- If M is an A-module and $N \subseteq M$ satisfies that N is a subgroup of $(M, +)$ and $a \cdot n \in N$ for all $a \in A$ and $n \in N$, we call N a submodule of M.
- If A is a unital commutative ring, then A is itself an A-module. The submodules of A (as an A-module) are precisely the ideals of A.
- If $\{M_i\}_{i \in I}$ is a family of A-modules, then so is $\prod_{i \in I} M_i$ (where we define all operations pointwise).

I want to return now to the topic of integrals of vector-valued function that I brought up on *pages 514-517*. Specifically, my goal will to be develop a few more integrals that work in more general vector spaces.

Firstly, I'll be establishing the Pettis integral. For a textbook I'll be following grandpa Rudin.

Here are some topological vector space facts:

Lemma 3.8(c) If $(X, \mathcal{T})$ is a compact topological space and if some sequence $\{f_n\}_{n\in\mathbb{N}}$ of continuous real or complex valued functions separates points on X, then X is metrizable.

By replacing each f_n with $\frac{|f_n|}{1+|f_n|}$ we can assume without loss of generality that $f_n(X) \subseteq [0,1)$ for all n. Next, we define $d(x,y) := \sum_{n=1}^{\infty} 2^{-n}|f_n(x) - f_n(y)|$. It is easy to verify that d satisfies the triangle inequality and the symmetric property. Plus, we can see as an easy consequence of the fact that $\{f_n\}_{n\in\mathbb{N}}$ separates points that $d(x,y) = 0$ iff $x = y$.

Finally, we need to show that the metric topology $\mathcal{T}_d$ induced by d is equal to $\mathcal{T}$. To start off, note that the infinite series defining d converges uniformly on $X \times X$ via the Weierstrass M-test. It follows that $d : X \times X \to \mathbb{R}$ is a continuous map. Consequently $x \mapsto d(x,y)$ is a continuous function. So, $\{x \in X : d(x,y) < r\} \in \mathcal{T}$ for all $y \in X$ and $r > 0$ and this proves that $\mathcal{T}_d \subseteq \mathcal{T}$.

To show the other inclusion, we bring in the following lemma (which I proved in my math 240b homework):

Folland Exercise 4.38: Suppose $(X, \mathcal{T})$ is a compact Hausdorff space and $\mathcal{T}'$ is another topology on X.

- If $\mathcal{T}'$ is strictly finer than $\mathcal{T}$ on X then $(X, \mathcal{T}')$ is Hausdorff but not compact.
- If $\mathcal{T}'$ is strictly coarser than $\mathcal{T}$ on X then $(X, \mathcal{T}')$ is compact but not Hausdorff.

Note that $(X, \mathcal{T}_d)$ is a Hausdorff topology since it's a metric space. Since any open cover in $\mathcal{T}_d$ is an open cover in $\mathcal{T}$ and $\mathcal{T}_d \subseteq \mathcal{T}$, we can also easily see that $(X, \mathcal{T}_d)$ is a compact topology. But now by the cited exercise, the only way for $(X, \mathcal{T})$ to be a compact space is if $\mathcal{T} = \mathcal{T}'$. ■.

Theorem 3.16: If $\mathcal{X}$ is a separable topological vector space and $K \subseteq \mathcal{X}^*$ is weak*-compact, then K is metrizable in the weak* subspace topology on K.

Proof:

Let $\{x_n\}_{n\in\mathbb{N}}$ be a countable dense set in $\mathcal{X}$ and put $f_n(\lambda) = \lambda x_n$ for all $\lambda \in \mathcal{X}^*$. By the definition of the weak* topology we know each f_n is weak* continuous on $\mathcal{X}^*$. Also note that if $\lambda, \eta \in \mathcal{X}^*$ satisfy that $f_n(\lambda) = f_n(\eta)$ for all n then we know that $\lambda = \eta$ on a dense subset of $\mathcal{X}$. It follows that $\lambda = \eta$ on all of $\mathcal{X}$ and in turn we've shown that $\{f_n\}_{n\in\mathbb{N}}$ separates points on $\mathcal{X}^*$.

To finish, just apply lemma 3.8(c). ■

Next, I want to generalize *(Zimmer) lemma 2.3.5 from pages 499-500* slightly.

Theorem 3.4: Suppose A and B are disjoint nonempty convex sets in a real or complex topological vector space $\mathcal{X}$.

(a) If A is open there exists $\lambda \in \mathcal{X}^*$ and $c \in \mathbb{R}$ such that $\mathrm{Re}(\lambda(x)) < c \leq \mathrm{Re}(\lambda y)$ for all $x \in A$ and $y \in B$.

(b) If A is compact, B is closed, and $\mathcal{X}$ is Hausdorff locally convex, then there exists $\lambda \in \mathcal{X}^*$ and $c_1, c_2 \in \mathbb{R}$ such that $\mathrm{Re}(\lambda(x)) < c_1 < c_2 < \mathrm{Re}(\lambda(y))$ for all $x \in A$ and $y \in B$.

Proof:
To start off, we can assume without loss of generality that $\mathcal{X}$ is a real topological vector space. After all, upon proving the real case to get a real-linear functional $u(x)$ we can then use the fact that there is a unique complex-valued linear functional $f(x) = u(x) - iu(ix)$ wtih $u(x)$ as it's real part. Also continuity is not an issue as $\|u\|_{\text{op}} = \|f\|_{\text{op}}$. (For a refresh see my math 240b notes...)

Proof of (a):
Fix $a_0 \in A$, $b_0 \in B$, and put $x_0 = b_0 - a_0$. Then put $C = A - B + x_0$. Now C is an open set containing 0. Also, B being convex is easily seen to imply $-B$ is convex. In turn, by basically identical reasoning to the *second lemma on page 236*, we know that C is convex.

As a side note, for this particular proof $A - B + x_0 = \{a - b + x_0 : a \in A, b \in B\}$ (i.e. $-$ is not a set difference operation).

Let p be the Minkowski functional of C. Since C isn't balanced, we won't have that p is a seminorm. But we can at least follow the reasoning in the proof of *(Zimmer) lemma 2.3.5 on pages 499-500* to get that p is a continuous well-defined sublinear functional. Also, $0 \leq p(x) \leq 1$ for all $x \in C$. That said we also have that $x_0 \notin A - B + x_0$. After all suppose:

$b_0 - a_0 = x_0 = a - b + x_0 = a - b + b_0 - a_0$ for some $a \in A$ and $b \in B$

That would imply that $a = b$, which contradicts that $A \cap B = \emptyset$. But now as C is convex and $x_0 \notin C$, we can can conclude that $p(x_0) \geq 1$.

Why?
Since C is convex, we have that if $0 \leq t_1 < t_2$ then $t_1 C \subseteq t_2 C$. When $t_1 = 0$ this is trivial since $0C = \{0\} \subseteq tC$ for all $t \in \mathbb{C}$ (which is true because C is a neighborhood of 0). Meanwhile, suppose $t_1 > 0$ and $x \in t_1 C$. Then as $t_2 C$ is convex and contains 0 and $\frac{t_2}{t_1}x$, we can conclude that $t_2 C$ contains $(1-s)0 + s\frac{t_2}{t_1}x = x$ where $s = \frac{t_1}{t_2}$.

Consequently, since $x_0 \notin 1C$, we know that $x_0 \notin tC$ for any $t < 1$. Hence, $p(x_0) \geq 1$.

Next, define a linear functional f on $\mathbb{R}x_0$ by setting $f(tx_0) = t$ for all $t \in \mathbb{R}$. By the Hahn-Banach theorem we know f extends to a linear functional λ on $\mathcal{X}$ with $\lambda \leq p$ for all $x \in \mathcal{X}$ and $\lambda(tx_0) = f(tx_0)$ for all $t \in \mathbb{R}$.

In particular, $\lambda(x) \leq 1$ for all $x \in C$. In turn $\lambda(x) \geq -1$ for all $x \in -C$ and $|\lambda(x)| \leq 1$ for all $x \in C \cap -C$. But $C \cap -C$ is a convex balanced neighborhood of 0. So, the Minkowski functional q associated with $C \cap -C$ satisfies that $|\lambda(x)| \leq q(x)$ and it follows that λ is continuous.

One more note is that since C is open, we actually have that $\lambda(x) < 1$ for all $x \in C$.

Proof:
If $m : \mathbb{R} \times \mathcal{X} \to \mathcal{X}$ is the scalar multiplication map, then because m is continuous and C is open we have that $m^{-1}(C)$ is open in $\mathbb{R} \times \mathcal{X}$. Also, for all $x \in C$ we have that $(1, x) \in m^{-1}(C)$. Therefore, we can find a small interval around $[\alpha, \beta]$ containing 1 such that $x \in t^{-1}C$ whenever $\alpha \leq t \leq \beta$.

Hence, if $a \in A$ and $b \in B$ then:

$$\lambda(a) - \lambda(b) + 1 = \lambda(a - b + x_0) \leq p_0(a - b + x_0) < 1$$

And, it follows that $\lambda(a) < \lambda(b)$ for all $a \in A$ and $b \in B$.

Finally, as λ is continuous and A and B are convex (and thus connected), we know that $\lambda(A)$ and $\lambda(B)$ are also connected in $\mathbb{R}$ and thus are intervals. Yet we also know that the right endpoint of $\lambda(A)$ is less than or equal to the left endpoint of $\lambda(B)$. So, set c equal to that endpoint.

Proof of (b):
By *lemma 3 on page 441*, we know there is an open set V containing 0 such that $(A + V) \cap (B + V) = \emptyset$. By the local convexity of $\mathcal{X}$ we can also assume V is convex. But now by part (a), we can find a functional $\lambda \in \mathcal{X}^*$ and $c \in \mathbb{R}$ such that:

$$\lambda(a + v) < c \leq \lambda(b) \text{ for all } a \in A,\, b \in B, \text{ and } v \in V.$$

But now we claim that if $C \subseteq \mathcal{X}$ is a set and $U \subseteq \mathcal{X}$ is an open neighborhood of 0, then for any $x \in C$ there exists $u_1, u_2 \in U$ such that $\lambda(x + u_1) < \lambda(x) < \lambda(x + u_2)$.

To prove this it suffices to show that there exists $u_1, u_2 \in U$ such that $\lambda(u_1) < 0 < \lambda(u_2)$. Fortunately, as λ is not the zero functional, we know there exists $y \in \mathcal{Y}$ such that $\lambda(y) > 0$. Then since U is absorbing, we can find $0 < s, t < 1$ such that $ty, s(-y) \in U$. And now $\lambda(-sy) < 0 < \lambda(ty)$.

It follows that $\lambda(A + V)$ is an open interval (α, β) with $\infty \leq \alpha < \beta \leq c$. At the same time, as A is compact and connected, we know that $\lambda(A)$ is a closed interval $[\gamma, \delta] \subseteq (\alpha, \beta)$. So, pick $\delta < c_1 < c_2 < c$. Then $\lambda(a) < c_1 < c_2 < \lambda(b)$ for all $a \in A$ and $b \in B$. ■

Theorem 3.7: Suppose B is a convex, balanced, closed set in a locally convex Hausdorff space $\mathcal{X}$ and $x_0 \in B^{\mathsf{C}}$. Then there exists $\lambda \in \mathcal{X}^*$ such that $|\lambda(x)| < 1$ for all $x \in B$ and $\lambda(x_0) > 1$.

Proof:
Apply part (b) of the last theorem (substituting in $A = \{x_0\}$) to get a linear function $\lambda_1 \in \mathcal{X}^*$ with $\lambda_1(x_0) = re^{i\theta}$ not in the closure of $\lambda_1(B)$. Since B is a balanced set, so is $\lambda_1(B)$.

(See *page 441* for why $\lambda_1(B)$ is balanced.)

In turn $\overline{\lambda_1(B)}$ is also balanced. So, we know there exists $s_1 \geq 0$ such that:

$$\overline{\lambda_1(B)} = \{z \in \mathbb{C} : |z| \leq s_1\} \text{ and } s_1 < r.$$

Finally choose $s \in (s, r)$ and set $\lambda = s^{-1}e^{i\theta}\lambda_1$. Then $\lambda \in \mathcal{X}^*$, $|\lambda(x)| \leq s^{-1}s_1 < 1$ for all $x \in B$, and $\lambda(x_0) = s^{-1}r > 1$. ■

Before doing the next theorem, here is an exercise from Folland Real Analysis.

Exercise 11.1: If G is a topological group with identity e and $E \subseteq G$, then:

$$\overline{E} = \bigcap\{EV : V \text{ is a neighborhood of } e\}.$$

Fix $A := \bigcap\{EV : V \text{ is a neighborhood of } e\}$. Then it's clear that $E \subseteq A$ since $E \subseteq EV$ for all sets V containing e. We also claim that A is closed.

Suppose $x \notin A$. Then we know there is a neighborhood V of e such that $x \notin EV$. Next, there is a smaller neighborhood U of e such that $UU \subseteq V$. And finally, we claim that xU^{-1} is a neighborhood of x disjoint from A. After all, suppose $y \in xU^{-1}$ is also in A. In particular that must mean that $y \in EU$. So, we can write

$$xu^{-1} = y = gu' \text{ where } u, u' \in U \text{ and } g \in E.$$

But that in turn would suggest that $x = gu'u \in E(UU) \subseteq EV$, a contradiction.

Hence, we've proven that every point in A^{C} is an interior point. Or in other words, A^{C} is an open set.

With that we've now shown that $\overline{E} \subseteq A$. To prove the other containment, it suffices to show that if $x \in A$ then for all neighborhoods V of x we have that $E \cap V \neq \emptyset$.

Fortunately, because $x^{-1}V$ is a neighborhood of e, we can find a symmetric neighborhood $U \subseteq x^{-1}V$ of e. But then because $x \in A$, we know that $x \in EU$. So, we can pick $g \in E$ and $u \in U$ such that $x = gu$. And since U is symmetric, we in turn know that $g = xu^{-1} \in xU$, thus proving that $E \cap xU \neq \emptyset$. Finally, since $U \subseteq x^{-1}V$, we know that $E \cap x(x^{-1}V) = E \cap V \neq \emptyset$. ■

Corollary: If G is a topological group with identity e and $U \subseteq G$ is a neighborhood of e, then there is an open set V satisfying that $e \in V \subseteq \overline{V} \subseteq U$.

Proof:

Let V be an open set containing e and satisfying that $VV \subseteq U$. Then by the prior exercise we know that:

$$\overline{V} = \bigcap\{VW : W \text{ is a neighborhood of } e\} \subseteq VV \subseteq U.$$

Because a topological vector space is a special case of a topological group, we'll be able to use this corollary in several later proofs.

Unfortunately I am out of time for now and need to pivot to doing complex analysis. So I will return to this on *page 540*.

1/11/2026

Math 220b Lecture Notes:

In math 220b, I'll denote uniform convergence on compact sets (also called local uniform convergence) by the notation $f_n \xrightarrow{\ell.u.} f$. Similarly, I'll denote uniform convergence by the notation $f_n \xrightarrow{u.} f$. (Technically the professor uses two arrows. However, I don't feel like figuring out how to type that since I feel like that notation is unnecessarily tall.)

The professor also uses the notation that an open ball centered at x with radius r is written as $\Delta(x, r)$. If we want to consider the closed ball then we write $\overline{\Delta}(x, r)$. Or, if we want to consider the boundary of that circle then we'll write $\partial\Delta(x, r)$.

One final notation quirk of this professor is that if $U \subseteq \mathbb{C}$ is open, then the set of holomorphic functions on U will be denoted $O(U)$.

Weierstraß Convergence Theorem: Suppose $\{f_n\}_{n\in\mathbb{N}}$ are a sequence of functions in $O(G)$ where G is a region, $f \in C(G)$, and $f_n \xrightarrow{\ell.u.} f$. Then f is holomorphic and $f_n' \xrightarrow{\ell.u.} f'$.

Proof:

To start off, note that if $f_n \xrightarrow{\ell.u.} f$ and γ is a C^1 path then $f_n \xrightarrow{u.} f$ on $\{\gamma\}$ since $\{\gamma\}$ is compact. Hence, $\int_\gamma f_n \to \int_\gamma f$ for any C^1 path γ with $\{\gamma\} \subseteq G$. But now it's an easy consequence of Morera's theorem (see my complex analysis notes from last Spring) that f is analytic. After all, if γ is a path going around the perimeter of a triangle contained in G, then $\int f_n = 0$ for all n and so $\int_\gamma f = \lim_{n\to\infty} \int_\gamma f_n = 0$.

With that we now know that f' exists on G. To show the other claim that $f_n' \xrightarrow{\ell.u.} f'$, it suffices to prove for any $a \in G$ that there is $r > 0$ small enough so that $\overline{\Delta}(a,r) \subseteq G$ and $f_n' \xrightarrow{u.} f'$ on $\overline{\Delta}(a,r)$. After all, we'll then be able to cover any compact set in G with finitely many of those disks.

Fortunately, we can pick $r > 0$ such that $\overline{\Delta}(a, 2r) \subseteq G$. Then we let $\gamma(t) = a + 2re^{it}$ for $t \in [0, 2\pi]$ and consider all $z \in \overline{\Delta}(a,r)$. By Cauchy's integral formula we have for all $z \in \overline{\Delta}(a,r)$ that $f(z) - f_n(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(w)-f_n(w)}{w-z}dw$. In particular, by applying *Leibniz's rule for contour integrals*, we have for all $z \in \overline{\Delta}(a,r)$ that $f'(z) - f_n'(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(w)-f_n(w)}{(w-z)^2}$.

But now if we let $M_n := \sup_{w\in\{\gamma\}} |f(w) - f_n(w)|$, we can conclude for all $z \in \overline{\Delta(a,r)}$ that: $|f'(z) - f_n'(z)| \leq \frac{1}{2\pi} \cdot \frac{M_n \cdot 2\pi(2r)}{r^2} = \frac{2M_n}{r}$. Since $M_n \to 0$ as $n \to \infty$, this proves that $f_n' \xrightarrow{u.} f'$ on $\overline{\Delta}(a,r)$. ■

Corollary: $O(G)$ is closed in $C(G)$ with respect to the topology of local uniform convergence.

As a side note, it is easy to see that if $\{f_n\}_{n\in\mathbb{N}} \subseteq C(G)$ converges locally uniformly to any $f \in \mathbb{C}^G$, then $f \in C(G)$. After all, as G is locally compact we can find for any $x \in G$ a neighborhood of x where $f_n \xrightarrow{u.} f$. So, f must be continuous at each individual point in G.

Hurwitz's Theorem: Let G be a region, $\{f_n\}_{n\in\mathbb{N}}$ a sequence of functions in $O(G)$ converging locally uniformly to f, and $\overline{\Delta}(a,R) \subseteq G$ satisfies that $f|_{\partial\Delta(a,R)}$ has no zeros. Then there exists $N \in \mathbb{N}$ such that for all $n \geq N$:

$$\#\text{zeros}(f_n|_{\overline{\Delta}(a,R)}) = \#\text{zeros}(f|_{\overline{\Delta}(a,R)}).$$

Proof:

Let $\varepsilon = \min\{|f(x)| : x \in \partial\Delta(a,R)\}$. By the assumption of the problem and the compactness $\partial\Delta(a,R)$, we know ε is well-defined and that $\varepsilon > 0$. Also as $f_n \xrightarrow{\ell.u.} f$, we know that $f_n \xrightarrow{u.} f$ on $\partial\Delta(a,R)$. Thus, there exists N such that $\forall n \geq N$, $|f_n(z) - f(z)| < \varepsilon$ for all $z \in \partial\Delta(a,R)$.

But now for all $n \geq N$, we have that $|f_n - f| < |f| \leq |f| + |-f_n|$ on $\partial\Delta(a,R)$. Thus, our desired result follows from *Rouché's theorem*. ■

Corollary: If G is a region, and $\{f_n\}_{n\in\mathbb{N}} \subseteq O(G)$ is a sequence of zero-free functions converging locally uniformly to f, then either f is zero-free or $f \equiv 0$.

Proof:

Assume $f \not\equiv 0$ but that f has a zero. Since Z_f has no limit points in G, we can find a closed disk $\overline{\Delta}(a, R)$ centered at a zero such that $f(z) \neq 0$ for all $z \in \partial\Delta(a, R)$. Now we can use Hurwitz's theorem to get a contradiction. ■

Corollary 2: If G is a region and $\{f_n\}_{n\in\mathbb{N}} \subseteq O(G)$ is a sequence of injective functions converging locally uniformly to f, then either f is injective or f is contant.

Proof:

Assume f is not injective. Then there exists $a \neq b$ such that $f(a) = f(b)$. Now let $\widetilde{f}_n(z) \coloneqq f_n(z) - f_n(a)$, $\widetilde{f}(z) \coloneqq f(z) - f(a)$, and $\widetilde{G} \coloneqq G - \{a\}$.

All the $\widetilde{f}_n$ are zero free in the region $\widetilde{G}$ since all f_n are injective. Also, $\widetilde{f}_n \xrightarrow{\ell.u.} \widetilde{f}$. So, either $\widetilde{f}$ is zero free on $\widetilde{G}$ or $\widetilde{f} \equiv 0$. Since $\widetilde{f}(b) = 0$ and $b \in \widetilde{G}$, we know it's the latter case that is true. So, $\widetilde{f}(z) = f(z) - f(a) = 0$ for all $z \in G - \{a\}$. This proves that $f \equiv f(a)$. ■

We say $\sum f_n$ converges absolutely locally uniformly iff $\sum |f_n|$ converges locally uniformly. Meanwhile, we say $\sum f_n$ converges normally if for all compact sets K in the domain of the f_n, $\sum \sup_K |f_n| < \infty$. By the Weierstrass M-test, normal convergence implies absolute local uniform convergence.

Math 220B Homework Assignment 1:

For this homework assignment, $\Delta \coloneqq \{z \in \mathbb{C} : |z| < 1\}$.

1. Qualifying Exam, Spring 2023: Let $P_1, \ldots, P_m$ be points on the unit circle. Show that there is a point Q on the unit circle such that $\prod_{i=1}^m |P_i - Q| \geq 1$.

Let $f(z) = \prod_{i=1}^m (P_i - z)$. Then $f \in O(\Delta) \cap C(\overline{\Delta})$. Also $\overline{\Delta}$ is compact. So, we can conclude by the maximum modulus principle that $\max_{z\in\overline{\Delta}}(|f(z)|)$ exists and equals $\max_{z\in\partial\Delta}(|f(z)|)$. But also note that $0 \in \Delta$ satisfies that $|f(0)| = 1$ since $|P_i| = 1$ for all i. Therefore, there must exist a $Q \in \partial\Delta$ with $1 \leq |f(Q)| = \prod_{i=1}^m |P_i - Q|$.

2. Qualifying Exam, Spring 2021: Let $f : \Delta(0,2) \to \mathbb{C}$ be a holomorphic function such that $|f(z)| < 1$ for all $z \in \Delta(0,2)$. Assume $f(1) = f(-1) = f(i) = f(-i) = 0$. Then show that $|f(0)| \leq 1/15$.

Let $g(z)$ be the holomorphic function on $\Delta(0,2)$ satisying that:

$$(z^4 - 1)g(z) = (z-1)(z+1)(z-i)(z+i)g(z) = f(z).$$

Then by the maximum modulus principle, for any $r < 2$ we have that:

$$\frac{|f(0)|}{|0-1|} \leq \max_{z\in\Delta(0,r)} |g(z)| = \max_{|z|=r} |g(z)| = \max_{|z|=r} \frac{|f(z)|}{|z^4-1|} \leq \max_{|z|=r} \frac{1}{|z^4-1|} = (\min_{|z|=r} |z^4 - 1|)^{-1}$$

But note that when z is confined to a circle, then $|z - 1|$ is minimized when the vector z is pointing opposite to -1. Therefore $(\min_{|z|=r} |z^4 - 1|)^{-1} = r^4 - 1$. So, we've shown that $|f(0)| \leq (r^4 - 1)^{-1}$. Finally, by taking $r \to 2$ we get that $|f(0)| \leq 1/15$.

3. Qualifying Exam, Spring 2024: Let $f : \mathbb{C} \to \mathbb{C}$ be an entire nowhere zero function which is not a constant. Define $U := \{z : |f(z)| < 1\}$. Then show that the connected components of U are unbounded.

Suppose $G \subseteq U$ is a bounded connected component of U. Then we know that $\overline{G}$ is compact. Hence there exists $a \in \overline{G}$ such that $|f(a)| = \min_{z \in \overline{G}} |f(z)|$. But now note that $\overline{G} - G$ must be disjoint from U since otherwise would contradict that G is a connected component of U. It follows that $a \in G$ since $|f(z)| < 1$ when $z \in G$ and $|f(z)| \geq 1$ when $z \in \overline{G} - G$.

Next note that as $U \subseteq \mathbb{C}$ is open and $\mathbb{C}$ is locally connected, all the connected components of U (including G) are open. For another way to see that G is open, note that for any $z \in G$ we can pick (using the openness of U) an $r > 0$ with $\Delta(z, r) \subseteq U$. Then since $\Delta(z, r) \subseteq U$ is connected and contains a point in G, we also have that $\Delta(z, r) \subseteq G$.

But as a consequence, we've now shown that $|f|$ has a local minimum at $z = a$. Since, f is zero-free, we know by applying the maximum modulus principle to f that $f \equiv f(a)$. Yet this contradicts our other initial assumption that f isn't constant.

4. Qualifying Exam, Fall 2017: Let $G \subseteq \mathbb{C}$ be a region containing $\overline{\Delta}$. Show for all holomorphic functions $f : G \to \mathbb{C}$ that $\max_{|z|=1} |f(z) - \frac{e^z}{z}| \geq 1$.

Consider the function $g(z) = zf(z) - e^z$. Clealy $g \in O(\Delta) \cap C(\overline{\Delta})$. Thus by the maximum modulus principle plus the fact that $\overline{\Delta}$ is compact, we know that:

$$1 = |g(0)| \leq \max_{z \in \overline{\Delta}} |g(z)| = \max_{|z|=1} |g(z)| = \max_{|z|=1} |zf(z) - e^z|.$$

Then because $|zf(z) - e^z| = |\frac{zf(z) - e^z}{z}| = |f(z) - \frac{e^z}{z}|$ when $|z| = 1$, we can conclude that $\max\limits_{|z|=1} |f(z) - \frac{e^z}{z}| \geq 1$.

5. Let f be a continuous function in the closed disk $\overline{\Delta}$ and holomorphic in Δ. Assume $|f(z)| = 1$ whenever $|z| = 1$.

(i) If f is not constant, show that f has a zero in Δ.

Suppose f is zero-free on Δ. Then since $\overline{\Delta}$ is compact and f is zero-free on $\partial\Delta$, we can apply the maximum modulus principle to $1/f$ to get that:

$$\left(\min_{|z| \leq 1} |f(z)|\right)^{-1} = \max_{|z| \leq 1} |f(z)|^{-1} = \max_{|z|=1} |f(z)|^{-1} = 1.$$

Yet simultaneously, we know from applying the maximum modulus principle to f that:

$$\max_{|z| \leq 1} |f(z)| = \max_{|z|=1} |f(z)| = 1.$$

This lets us conclude that $|f(z)| = 1$ for all $z \in \overline{\Delta}$. But now from an exercise we did in math 220a, we are able to conclude that f is constant on Δ (see *(Conway) Exercise III.3.17*). For a faster proof of this claim (that I would actually want to use on a qual), note that if f is nonconstant then $f(\Delta)$ must be a nonempty open set in $\mathbb{C}$ by the open map theorem. Yet, we know $f(\Delta) \subseteq \partial\Delta$ and the latter set has an empty interior. Thus, to avoid a contradiction we must have that f is constant on Δ.

Finally, because continuous functions are uniquely determined by their values on a dense subset of their domain and Δ is dense in $\overline{\Delta}$, we can conclude that f is constant on all of $\overline{\Delta}$.

(ii) If f has a zero at 0 of order 1 and no other zeros, then $f(z) = \alpha z$ for some $\alpha \in \mathbb{C}$.

Let $g(z)$ be the function in $O(\Delta) \cap C(\overline{\Delta})$ satisfying that $zg(z) = f(z)$. Then g is zero-free on Δ and satisfies that $|g(z)| = 1$ whenever $|z| = 1$. By part (a), we can conclude that $g \equiv \alpha$ for some $\alpha \in \mathbb{C}$. Hence, we've proven that $f(z) = zg(z) = \alpha z$.

6. Qualifying Exam, Fall 2020: Let $f : \mathbb{C} \to \mathbb{C}$ be an entire function. Show that the series $\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n$ converges uniformly on compact subsets of $\mathbb{C}$.

To start off, if $K \subseteq \mathbb{C}$ is compact (and thus bounded), then we can find $R > 0$ such that $K \subseteq \Delta(0, R)$. Next, since R is less than the radius of convergence of the power series $\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n$, we know that $\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n$ converges uniformly on $\overline{\Delta(0, R)}$. And as $K \subseteq \overline{\Delta(0, R)}$, we are done.

7. Assume that f is an entire function such that the sequence of derivatives $f, f', f'', f^{(3)}, \ldots$ converges locally uniformly to a function g with $g(0) = 1$. Then show there exists N such that $f^{(n)}(z) \neq 0$ for all $n \geq N$ and $|z| < 1$.

We know g is an entire function since $f^{(n)}$ is holomorphic for all n and $f^{(n)} \to g$ locally uniformly. In particular, note by the Weierstraß Convergence Theorem that $f^{(n+k)} = (f^{(n)})^{(k)} \to g^{(k)}$ as $n \to \infty$ for all k. But $\lim_{n\to\infty} f^{(n+k)} = \lim_{n\to\infty} f^{(n)}$ for all k. So, we actually have that $g^{(k)}(z) = g(z)$ for all $k \in \mathbb{Z}_{\geq 0}$ and $z \in \mathbb{C}$.

Taking the power series expansion of g centered at 0, we get that:

$$g(z) = \sum_{n=0}^{\infty} \frac{g^{(n)}(0)}{n!} z^n = \sum_{n=0}^{\infty} \frac{g(0)}{n!} z^n = \sum_{n=0}^{\infty} \frac{1}{n!} z^n = e^z.$$

This shows that g is a zero-free function. And now the stated claim is a simple corollary of Hurwitz's theorem.

1/14/2026

Math 200B Homework Set 1:

Set 1 Problem 1: A Bezout domain is an integral domain D in which for all $a, b \in D$ there exists $c \in D$ such that $\langle a, b \rangle = \langle c \rangle$.

(a) Prove that an integral domain D is a Bezout domain if and only if for all $a, b \in D - \{0\}$ there exists $d \in D$ such that $d \mid a$, $d \mid b$, and $d \in \langle a, b \rangle$.

$(\Longrightarrow)$
Suppose $a, b \in D - \{0\}$ and let $d \in D$ be such that $\langle a, b \rangle = \langle d \rangle$. Then d satisfies the three properties we want. After all, since $a, b \in \langle d \rangle$, we know that there exists $x_1, x_2 \in D$ such that $a = dx_1$ and $b = dx_2$. Hence, $d \mid a$ and $d \mid b$. Furthermore, $d \in \langle a, b \rangle$.

$(\Longleftarrow)$
If a is zero then $\langle a, b \rangle = \langle b \rangle$. Similarly, if b is zero then $\langle a, b \rangle = \langle a \rangle$. So, the only

nontrivial case we need to worry about when proving that D is a Bezout domain is when both a and b are nonzero.

Let $d \in D$ satisfy that $d \mid a$, $d \mid b$, and $d \in \langle a, b\rangle$. Then it's clear that $\langle d\rangle \subseteq \langle a, b\rangle$ since $d \in \langle a, b\rangle$. Also, as both $a \in \langle d\rangle$ and $b \in \langle d\rangle$, we have that $\langle a, b\rangle \subseteq \langle d\rangle$.

(Notice that if d satisfies the above three properties and d' is another common divisor of a and b, then $d' \mid d$. So, we say d is a greatest common divisor of a and b.)

Why?
We know $\langle d\rangle = \langle a, b\rangle$ due to those three properties. At the same time, we know that $\langle a, b\rangle \subseteq \langle d'\rangle$ since $d' \mid a$ and $d' \mid b$. It follows that $\langle d\rangle \subseteq \langle d'\rangle$, and that implies that $d' \mid d$.

(b) Prove that every finitely generated ideal of a Bezout domain is principal.

By the definition of a Bezout domain, we know that any ideal generated by two elements of D is a principal ideal. Meanwhile, suppose $n > 2$ and consider an ideal $\langle a_1, \ldots, a_{n-1}, a_n\rangle$.

If $a_n = 0$ or $a_{n-1} = 0$, then we can trivially say that the ideal $\langle a_1, \ldots, a_{n-1}, a_n\rangle$ is generated by $n-1$ elements (just remove 0 from the generating set). As for if $a_n \neq 0 \neq a_{n-1}$, then we can pick (by part (a)) some $d \in D$ such that $d \mid a_n$, $d \mid a_{n-1}$, and $d \in \langle a, b\rangle$. Now I claim that $\langle a_1, \ldots, a_{n-1}, a_n\rangle = \langle a_1, \ldots, a_{n-2}, d\rangle$.

Since $d \in \langle a_{n-1}, a_n\rangle \subseteq \langle a_1, \ldots, a_{n-1}, a_n\rangle$, we have that:

$$\langle a_1, \ldots, a_{n-1}, a_n\rangle \supseteq \langle a_1, \ldots, a_{n-2}, d\rangle$$

On the other hand, as $a_n, a_{n-1} \in \langle d\rangle \subseteq \langle a_1, \ldots, a_{n-2}, d\rangle$ we have that:

$$\langle a_1, \ldots, a_{n-1}, a_n\rangle \subseteq \langle a_1, \ldots, a_{n-2}, d\rangle$$

Now by induction we can conclude our ideal generated by $n-1$ elements is a principal ideal.

(c) Prove that D is a P.I.D. if and only if it is both a U.F.D. and a Bezout domain.

$(\Longrightarrow)$
We already showed in class that all P.I.D.s are U.F.D.s. Also, all P.I.D.s are trivially Bezout domains.

$(\Longleftarrow)$
Let $\mathfrak{a}$ be any nonzero ideal in D. Then pick $a \in \mathfrak{a}$ with the fewest factors compared to all $b \in \mathfrak{a}$ in its decomposition as a product of irreducibles. It's trivial that $\langle a\rangle \subseteq \mathfrak{a}$. To show the other inclusion, suppose $b \in \mathfrak{a}$. Also without loss of generality we can assume $b \neq 0$ since $0 \in \langle a\rangle$.

Since D is a Bezout domain, we know there exists $c \in D$ such that $\langle c\rangle = \langle a, b\rangle \subseteq \mathfrak{a}$ and in particular that $c \mid a$. But now by how we chose a, we know that c has at least as many irreducible factors as a does. Since we also have that $\nu_p(c) \leq \nu_p(a)$ for any irreducible p, (since $c \mid a$), we can actually conclude that c and a are companions. Hence, $\langle a, b\rangle = \langle c\rangle = \langle a\rangle$. In particular, this let's us conclude that $b \in \langle a\rangle$ for any $b \in \mathfrak{a}$. So, $\mathfrak{a} \subseteq \langle a\rangle$.

Set 1 Problem 2: Let A be a subring of $\mathbb{Q}[x,y]$ which is generated by $\mathbb{Q}, x, xy, xy^2, \ldots$

In other words, $A := \mathbb{Q}[x, xy, xy^2, \ldots]$. What this means is that if we consider the evaluation homomorphism $e : \mathbb{Q}[x_0, x_1, x_2, \ldots] \to \mathbb{Q}[x,y]$ mapping x_k to xy^k, then A is the range of e. A slightly more practical way of viewing A that works in this specific case is that A consists of all polynomials $\sum\limits_{n_1,n_2 \in \mathbb{Z}_{\geq 0}} c_{n_1,n_2} x^{n_1} y^{n_2}$ where $c_{n_1,n_2} = 0$ if $n_1 = 0$ and $n_2 \neq 0$.

Prove that A is not Noetherian.

Let $\mathfrak{a}_n := \langle \{xy^k : k \in \{0, 1, \ldots, n\}\} \rangle$ for all $n \in \mathbb{Z}_{\geq 0}$. Then it's clear that $\mathfrak{a}_0 \subseteq \mathfrak{a}_1 \subseteq \cdots$. That said, I also claim that $xy^{n+1} \notin \langle x, xy, \ldots, xy^n \rangle$.

Suppose $p_1(x,y)x + \ldots + p_n(x,y)xy^n = xy^{n+1}$ for some arbitrary polynomials $p_1, \ldots, p_n$ in $\mathbb{Q}[x,y]$. Focusing on just the xy^{n+1}, terms, we thus have that there are numbers $c^{(1)}, \ldots, c^{(n)} \in \mathbb{Q}$ such that:

$$(c^{(1)}y^{n+1})x + (c^{(2)}y^n)xy + \cdots + (c^{(n)}y^1)xy^n = 1xy^{n+1}.$$

But now if each $p_i \in A$, we must have that each $c^{(i)} = 0$. Yet simultaneously, $0 + \cdots + 0 \neq 1$. Hence, $xy^{n+1} \notin Ax + Axy + \ldots + Axy^n = \mathfrak{a}_n$.

It follows that $\mathfrak{a}_n$ is a proper subset of $\mathfrak{a}_{n+1}$ for all n. The fact that such a sequence $\{\mathfrak{a}_n\}_{n \in \mathbb{N}}$ of ideals in A without a maximal element exists proves that A is not Noetherian.

Set 1 Problem 3: Let D be a U.F.D. and $F = Q(D)$ be the field of fractions of D.

(a) (Rational root criterion) Suppose $p(x) = a_n x^n + \ldots + a_1 x + a_0 \in D[x]$, $\frac{r}{s} \in F$ is a zero of $p(x)$, and r and s have no common irreducible factors. Then $s \mid a_n$ and $r \mid a_0$.

Multiplying both sides of the equation $a_n(\frac{r}{s})^n + \ldots + a_1(\frac{r}{s}) + a_0 = 0$ by s^n and rearranging, we get that:

- $-sa_{n-1}r^{n-1} - \cdots - s^{n-1}a_1 r - s^n a_0 = a_n r^n$;
- $-a_n r^n - sa_{n-1}r^{n-1} - \cdots - s^{n-1}a_1 r = s^n a_0$.

In particular, this shows that $s \mid a_n r^n$ and $r \mid s^n a_0$.

Yet note that since D is a U.F.D., $s \mid a_n r^n$, and s shares no irreducible factors with r, we know for any irreducible $p \in D$ with $\nu_p(s) \neq 0$ that:

$$\nu_p(s) \leq \nu_p(a_n r^n) = \nu_p(a_n) + n\nu_p(r^n) = \nu_p(a_n)$$

It follows that $\nu_p(s) \leq \nu_p(a_n)$ for all irreducible elements p, and that implies that $s \mid a_n$. By analogous reasoning, we can also show $r \mid a_0$.

(b) (Integrally closed) Prove that a fraction $\frac{r}{s}$ is a zero of some monic polynomial in $D[x]$ if and only if $\frac{r}{s} \in D$.

$(\Longrightarrow)$
By part (a), we know that $\frac{r}{s}$ being a zero of a polynomial in $D[x]$ implies that s divides the leading term of that polynomial. Yet if the leading term of that polynomial is 1, this implies that s is a unit. So $\frac{r}{s} = rs^{-1}$ in D.

$(\Longleftarrow)$
If $d \in D$, then just consider the monic polynomial $x - d$.

(c) Prove that $\mathbb{Z}[2\sqrt{2}]$ is not a U.F.D.

Consider the polynomial $x^2 - 2 \in (\mathbb{Z}[2\sqrt{2}])[x]$ which is monic and has $\sqrt{2} = \frac{2\sqrt{2}}{2} \in Q(\mathbb{Z}[2\sqrt{2}])$ as a zero. Thus if $\mathbb{Z}[2\sqrt{2}]$ were a U.F.D., we'd know by part (a) that $\sqrt{2} \in \mathbb{Z}[2\sqrt{2}]$.

Yet I claim that $\sqrt{2} \notin \mathbb{Z}[2\sqrt{2}]$. After all, suppose for the sake of contradiction that there exists $a_0, \ldots, a_{2n+1} \in \mathbb{Z}$ (with a_{2n+1} allowed to be zero) such that $\sqrt{2} = \sum_{k=0}^{2n+1} a_k(2\sqrt{2})^k$. After regrouping terms, we can in particular say that:

$$\sqrt{2} = (a_0 + \textstyle\sum_{k=1}^{n} 2^{3k}a_{2k}) + (\sum_{k=0}^{n} 2^{3k}a_{2k+1})2\sqrt{2}.$$

Or in other words, there exists integers $n, m \in \mathbb{Z}$ such that $2 \mid m$ in $\mathbb{Z}$ and $\sqrt{2} = n + m\sqrt{2}$. But note that if $m \neq 1$, then that would imply that $\sqrt{2} = \frac{n}{1-m} \in \mathbb{Q}$ (which we know is not true). Hence we have a contradiction as m must equal 1 and yet $2 \nmid 1$.

As a result, we know that $\sqrt{2} \notin \mathbb{Z}[2\sqrt{2}]$. The only way this is possible is if $\mathbb{Z}[2\sqrt{2}]$ isn't a U.F.D.

Set 1 Problem 4: Let $A := \mathbb{Z} + x\mathbb{Q}[x]$. In other words:

$$A = \{a_0 + a_1x + \ldots + a_nx^n : a_0 \in \mathbb{Z}, n \in \mathbb{N}, \text{ and } a_1, \ldots, a_n \in \mathbb{Q}\}$$

(a) Prove that $f(x) \in A$ is irreducible iff either $f(x) = \pm p$ where p is a prime integer or $f(x) \in \mathbb{Q}[x]$ is irreducible and $f(0) = \pm 1$.

If we view A as a subring of $\mathbb{Q}[x]$, then $u \in A$ is a unit of A iff u^{-1} exists in $\mathbb{Q}[x]$ and is contained in A. In particular, this shows that the only units of A are ± 1.

Now we break into cases to show when a polynomial in $A - (\{0\} \cup A^\times)$ is irreducible.

- Suppose $\deg(f) = 0$. Thus $f(x) = a$ where $a \in \mathbb{Z} - \{-1, 0, 1\}$.
  Next, write $f(x) = g_1(x)g_2(x)$ where $g_1(x), g_2(x) \in A$. Then as $0 = \deg(f) = \deg(g_1) + \deg(g_2)$, we can actually conclude $\deg(g_i) = 0$ and $g_i(x) = b_i \in \mathbb{Z}$ for both i. So, $f(x) = a$ is irreducible in A if and only if $a = b_1b_2 \Longrightarrow b_1 = \pm 1$ or $b_2 = \pm 1$. But that means that a degree 0 polynomial in A is irreducible iff it equals $\pm p$ where p is a prime.

- Suppose $f(x) = a_1x + \ldots + a_nx^n$ where $a_n \neq 0$. As $a_1 \in \mathbb{Q}$, we know there exists some integer $m \in \mathbb{Z} - \{0\}$ such that $a_1m \in \mathbb{Z}$. In turn we know that f is *not* irreducible in A since:
  $$f(x) = (\tfrac{1}{2m}x) \cdot (2(a_1m) + 2(a_2m)x + \ldots + 2(a_nm)x^{n-1})$$
  where neither of the right-side polynomials two polynomials are ± 1.

- Suppose $f(x) = a_0 + a_1x + \ldots + a_nx^n$ where $n \geq 1$, $a_n \neq 0$ and $|a_0| \geq 2$. Then $f(x)$ is *not* irreducible in A since:
$$f(x) = a_0 \cdot (1 + \tfrac{a_1}{a_0}x + \ldots + \tfrac{a_n}{a_0}x^n)$$
where neither of the right-side polynomials are ± 1.

- Suppose $f(x) = a_0 + a_1x + \ldots + a_nx^n$ where $n \geq 1$, $a_n \neq 0$, and $a_0 = \pm 1$. If f is not irreducible in $\mathbb{Q}[x]$, then we can find polynomials $g_1, g_2 \in \mathbb{Q}[x]$ with degree at least 1 such that $f(x) = g_1(x)g_2(x)$. Then by Gauss's lemma v3, we can without loss of generality assume $g_1, g_2 \in \mathbb{Z}[x] \subseteq A$. So, f is *not* irreducible in A either since neither g_1 nor g_2 equals ± 1.

- Finally, suppose $f(x) = a_0 + a_1x + \ldots + a_nx^n$ where $n \geq 1$, $a_n \neq 0$, $a_0 = \pm 1$, and f is irreducible in $\mathbb{Q}[x]$. Then write $f(x) = g_1(x)g_2(x)$ where $g_1, g_2 \in A$. Since f is irreducible in $\mathbb{Q}[x]$, we know that $\deg(g_i) = 0$ for one of the i. So without loss of generality we can assume $g_1(x) = b_0$ where $b_0 \in \mathbb{Z}$.

  Next, as $\pm 1 = b_0g_2(0)$ where both b_0 and $g_2(0)$ are integers, we must have that $b_0, g_2(0) = \pm 1$. Hence, we've proven that $g_1 \in A^\times$ and thus f is irreducible in A.

(b) Prove that x cannot be written as a product of irreducibles in A.

Suppose for the sake of contradiction that $x = p_1 \cdots p_n$ where each p_i is irreducible in A. Then as $1 = \deg(x) = \sum_{i=1}^n \deg(p_i)$, we must have that exactly one p_i is a degree one polynomial and the rest are constant polynomials. Hence, by part (a) we can equivalently say that there are prime integers $q_1 \cdots q_{n-1}$, as well as some $r \in \mathbb{Q}$ such that $x = \pm(1 + rx)\prod_{i=1}^{n-1} q_i$.

But as $\mathbb{Z}$ is an integral domain, there is no way for $\pm 1 \cdot \prod_{i=1}^{n-1} q_i$ to equal zero. Hence, this equation is impossible to solve.

(c) Prove that A is neither a U.F.D. nor Noetherian.

We know as a trivial consequence of part (b) that A is not a U.F.D. Meanwhile, recall a lemma on *page 488* saying that if A is a Noetherian integral domain then every nonzero-nonunit element of A can be factored into a product of irreducible elements. Yet $x \notin A^\times \cup \{0\}$ and x can't be factored into a product of irreducible elements by part (b). It follows that A must either not be a domain or not be Noetherian. Since A is a subring of $\mathbb{Q}[x]$ which is itself a domain, we know that A is also a domain. Thus, it's the latter claim that A is Noetherian that must fail.

Set 1 Problem 5: Suppose A is a unital commutative ring.

(a) Let $\Sigma := \{\mathfrak{a} \lhd A : \mathfrak{a} \text{ is not finitely generated}\}$. Suppose Σ is not empty. Then Σ has a maximal element.

Let $\mathcal{C} \subseteq \Sigma$ be a chain (ordered by inclusion). Then set $\mathfrak{b} := \bigcup_{\mathfrak{a} \in \mathcal{C}} \mathfrak{a}$. By a lemma from class (see *page 455*), we know that $\mathfrak{b} \lhd A$. Also, I claim that $\mathfrak{b}$ is not finitely generated. After all, suppose to the contrary that $\mathfrak{b} = \langle b_1, \ldots b_n \rangle$. By the definition of $\mathfrak{b}$, for each b_i we can find $\mathfrak{a}_i \in \mathcal{C}$ such that $b_i \in \mathfrak{a}_i$. Then using the fact that $\mathcal{C}$ is a chain, we can take the max element of $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ to get an ideal $\mathfrak{a} \in \mathcal{C}$ containing $b_1, \ldots, b_n$. Yet now $\langle b_1, \ldots, b_n \rangle \subseteq \mathfrak{a} \subseteq \mathfrak{b} = \langle b_1, \ldots, b_n \rangle$ implies that $\mathfrak{a} = \langle b_1, \ldots, b_n \rangle$ is finitely generated. This is a contradiction.

We now conclude by Zorn's lemma that Σ has a maximal element.

(b) Suppose $\mathfrak{p}$ is a maximal element of Σ. Prove that $\mathfrak{p}$ is a prime ideal.

Suppose for the sake of contradiction that $\mathfrak{p}$ is a maximal element of Σ but that $\mathfrak{p}$ is not a prime ideal. Since $A = \langle 1 \rangle$ is finitely generated and $\mathfrak{p}$ is not finitely generated, we know that $\mathfrak{p}$ is a proper ideal of A. It follows by a lemma from class (see *page 452*) that there exists $a, b \notin \mathfrak{p}$ with $ab \in \mathfrak{p}$.

Next note that $\mathfrak{p} + \langle a \rangle$ is an ideal of A properly containing $\mathfrak{p}$. It follows that $\mathfrak{p} + \langle a \rangle \notin \Sigma$ since otherwise would contradict the maximality of $\mathfrak{p}$. But in turn we know there exists $p_i \in \mathfrak{p}$ and $r_i \in A$. such that:

$$\mathfrak{p} + \langle a \rangle = \langle p_1 + r_1 a, \ldots, p_n + r_n a \rangle$$

Meanwhile, let $(\mathfrak{p} : \langle a \rangle) := \{x \in A : xa \in \mathfrak{p}\}$. This is an ideal.

If $x_1, x_2 \in A$ satisfy that $ax_1 a, x_2 a \in \mathfrak{p}$, then $(x_1 + x_2)a = x_1 a + x_2 a \in \mathfrak{p}$. Similarly, if $x \in A$ satisfies that $xa \in \mathfrak{p}$ and r is any other element in A, then $(rx)a = r(xa) \in \mathfrak{p}$.

Furthermore $(\mathfrak{p} : \langle a \rangle)$ properly contains $\mathfrak{p}$. After all, we know $xa \in \mathfrak{p}$ for all $x \in \mathfrak{p}$ (by the definition of an ideal). At the same time, we know $b \notin \mathfrak{p}$ but that $ba \in \mathfrak{p}$. Hence, $b \in (\mathfrak{p} : \langle a \rangle)$. By similar reasoning to before, we can thus conclude that $(\mathfrak{p} : \langle a \rangle)$ is finitely generated, meaning there exists $s_i \in A$ such that:

$$(\mathfrak{p} : \langle a \rangle) = \langle s_1, \ldots, s_m \rangle$$

Finally, the provided hint claims that $\mathfrak{p} = \langle p_1, \ldots, p_n, s_1 a, \ldots, s_m a \rangle$.

To see this first note that $s_i a \in \mathfrak{p}$ for all $i \in \{1, \ldots, m\}$ by the definition of $(\mathfrak{p} : \langle a \rangle)$. And as each $p_i \in \mathfrak{p}$, we can conclude that:

$$\mathfrak{p} \supseteq \langle p_1, \ldots, p_n, s_1 a, \ldots, s_m a \rangle.$$

To show the other inclusion, suppose $x \in \mathfrak{p}$. Then as $x \in \mathfrak{p} + \langle a \rangle$, we can conclude there exists $c_1, \ldots, c_n \in A$ such that:

$$x = c_1(p_1 + r_1 a) + \ldots + c_n(p_n + r_n a)$$

But then $(\sum_{i=1}^n c_i r_i)a = x - \sum_{i=1}^n c_i p_i$ where the latter element is in $\mathfrak{p}$. It follows that $\sum_{i=1}^n c_i r_i \in (\mathfrak{p} : \langle a \rangle)$ and hence, we can find $d_1, \ldots, d_m \in A$ such that $\sum_{i=1}^n c_i r_i = d_1 s_1 + \ldots + d_m s_m$.

Finally, we have shown that:

$$x = c_1 p_1 + \ldots + c_n p_n + d_1 s_1 a + \ldots + d_m s_m a \in \langle p_1, \ldots, p_n, s_1 a, \ldots, s_m a \rangle.$$

(c) (Cohen) Suppose all the prime ideals of A are finitely generated. Prove that A is Noetherian.

If A isn't Noetherian, then we'd know there is an ideal in A not generated by any finite set. In turn, Σ wouldn't be empty and by parts (a) and (b) we'd be able to find a prime ideal $\mathfrak{p}$ not generated by any finite set. ■

Set 1 Problem 6: Suppose $f(x) \in (\mathbb{Z}/n\mathbb{Z})[x]$ is a monic polynomial of degree d. Prove that $|(\mathbb{Z}/n\mathbb{Z})[x]/\langle f(x)\rangle| = n^d$.

Since $f(x)$ is monic, we can apply the long division theorem to get unique polynomials $r(x), q(x)$ for any $g(x) \in (\mathbb{Z}/n\mathbb{Z})[x]$ satisfying that $g(x) = q(x)f(x) + r(x)$ and $\deg(r) \leq d - 1$.

As a consequence of the existance $q(x)$ and $r(x)$, we know that:

$$g(x) + \langle f(x)\rangle = r(x) + \langle f(x)\rangle.$$

Meanwhile, as a consequence of the uniqueness of $q(x)$ and $r(x)$, we know that if $r'(x)$ is any other polynomial in $(\mathbb{Z}/n\mathbb{Z})[x]$ of degree at most $d-1$, then there can't exist a polynomial $q'(x)$ with $g(x) = q'(x)f(x) + r'(x)$. Hence:

$$g(x) + \langle f(x)\rangle \neq r'(x) + \langle f(x)\rangle.$$

It follows that every equivalence class of $\frac{(\mathbb{Z}/n\mathbb{Z})[x]}{\langle f(x)\rangle}$ contains exactly one polynomial of degree at most $d-1$. So:

$$\left|\frac{(\mathbb{Z}/n\mathbb{Z})[x]}{\langle f(x)\rangle}\right| = \#\text{polynomials of degree at most } d-1 = n^d.$$

Set 1 Problem 7: Suppose $p \in \mathbb{Z}$ is prime. Prove the following statements are equivalent:

(a) p is not irreducible in $\mathbb{Z}[i] = \{m + in : m, n \in \mathbb{Z}\}$;

(b) There exists $a, b \in \mathbb{Z}$ such that $p = a^2 + b^2$;

(c) $x^2 \equiv -1 \pmod p$ has a solution.

$(a \implies b)$
Firstly, note that $|w|^2 \in \mathbb{Z}_{\geq 0}$ for all $w \in \mathbb{Z}[i]$. After all, if $w = a + bi \in \mathbb{Z}[i]$ then $|w|^2 = a^2 + b^2$.

Now suppose $p = z_1 z_2$ where $z_1, z_2 \in \mathbb{Z}[i]$ are not units. Then we first claim that $|z_j| \neq 1$ for either j.

Suppose $w \in \mathbb{Z}[i]$ is a unit. Then, there exists $w^{-1} \in \mathbb{Z}[i]$ satisfying that $1 = w \cdot w^{-1}$ and in turn we have that $1 = |1|^2 = |w|^2|w^{-1}|^2$. But now since $|w|^2$ and $|w^{-1}|^2$ are nonnegative integers, the only way this is possible is if $|w|^2 = 1 = |w^{-1}|^2$. Hence, any unit w in $\mathbb{Z}[i]$ must satisfy that $|w|^2 = 1$.

But also note that $p^2 = |p|^2 = |z_1 z_2|^2 = |z_1|^2|z_2|^2$. Therefore, we must have that $|z_j|^2 = p$ for both j. Then upon writing $z_1 = a + bi$ where $a, b \in \mathbb{Z}$, we have that $p = a^2 + b^2$.

$(b \implies c)$
Suppose $p = a^2 + b^2$. Then we know that $0 \equiv a^2 + b^2 \pmod p$. Also, I claim that $b \not\equiv 0 \pmod p$.

If not then we must also have that $a^2 \equiv 0 \pmod p$, and the only way that is possible is if $a \equiv 0 \pmod p$. It follows after writing $a = mp$ and $b = np$ where $m, n \in \mathbb{Z}$ that:

$$p = a^2 + b^2 = m^2p^2 + n^2p^2 = (m^2 + n^2)p^2$$

This is a contradiction as the above equation is impossible (it violates the uniqueness of the factorization of p into a product of primes).

It follows that b is invertible in $\mathbb{Z}/p\mathbb{Z}$. After representing that inverse as $c + p\mathbb{Z}$ and multiplying it to both sides of the equation $0 \equiv a^2 + b^2 \pmod p$, we get that:
$$0 \equiv (ac)^2 + 1 \pmod p$$

In other words, $-1 \equiv (ac)^2 \pmod p$

$(c \Longrightarrow a)$
We know there exists $x_0 \in \mathbb{Z}$ such that $p \mid x_0^2 + 1$. Hence $p \mid (x_0 + i)(x_0 - i)$. That said, note that $p \not\mid (x_0 + i)$ and $p \not\mid (x_0 - i)$. After all, $\mathrm{Im}(p(a + bi)) \equiv 0 \pmod p$ for all $a + bi \in \mathbb{Z}[i]$ and $\mathrm{Im}(x_0 \pm i) \not\equiv 0 \pmod p$. It follows that p is not a prime element of $\mathbb{Z}[i]$.

Finally, it was covered in a lecture (and in my math 100b notes) that $\mathbb{Z}[i]$ is an E.D. when equipped with $|\cdot|^2$. It thus follows that $\mathbb{Z}[i]$ is a P.I.D. and in turn a U.F.D.. So, p being not prime in $\mathbb{Z}[i]$ implies that p is not irreducible either.

Set 1 Problem 8: Suppose $p \in \mathbb{Z}$ is prime. Prove the following statements are equivalent:

(a) p is not irreducible in $\mathbb{Z}[\omega] = \{m + n\omega : m, n \in \mathbb{Z}\}$ where $\omega := \frac{-1+i\sqrt{3}}{2}$;

(b) There exists $a, b \in \mathbb{Z}$ such that $p = a^2 - ab + b^2$;

(c) $x^2 - x + 1 \equiv 0 \pmod p$ has a solution.

$(a \Longrightarrow b)$
Like in problem 7, note that $|w|^2 \in \mathbb{Z}_{\geq 0}$ for all $w \in \mathbb{Z}[\omega]$. After all if $w = a + b\omega \in \mathbb{Z}[i]$ then we know (by problem 4 on the 9th problem set of math 200a on *pages 463-464*) that:
$$|w|^2 = a^2 - ab + b^2 \geq a^2 - 2ab + b^2 = (a - b)^2.$$

It follows like in problem 7 that $w \in \mathbb{Z}[\omega]$ is a unit only if $|w|^2 = 1$. So if $p = z_1 z_2$ where $z_1, z_2 \in \mathbb{Z}[\omega]$ aren't units, then the fact that $p^2 = |p|^2 = |z_1 z_2|^2 = |z_1|^2 |z_2|^2$ mean means that $|z_j|^2 = p$ for both j. Then upon writing $z_1 = a + b\omega$ where $a, b \in \mathbb{Z}$, we will have that $p = a^2 - ab + b^2$.

$(b \Longrightarrow c)$
Suppose $p = a^2 - ab + b^2$. Then we know that $0 \equiv a^2 - ab + b^2 \pmod p$. Also I claim that $b \not\equiv 0 \pmod p$.

If not then we have that $a \equiv 0 \pmod p$ for identical reasons as in problem 7. But then upon writing $a = mp$ and $b = np$ where $m, n \in \mathbb{Z}$, we have that:
$$p = a^2 - ab + b^2 = m^2p^2 - mnp^2 + n^2p^2 = (m^2 - mn + n^2)p^2$$

Like in problem 7 this contradicts the uniqueness of the factorization of p into a product of primes.

Thus we know b has some inverse element $c + p\mathbb{Z}$ in $\mathbb{Z}/p\mathbb{Z}$. After multiplying it to both sides of the equation $0 \equiv a^2 - ab + b^2 \pmod p$, we get that:
$$0 \equiv (ac)^2 - (ac) + 1 \pmod p$$

In other words, $0 \equiv (ac)^2 - (ac) + 1 \pmod p$

$(c \Longrightarrow a)$
Suppose $x_0 \in \mathbb{Z}$ satisfies that $p \mid x_0^2 - x_0 + 1$. Then note $x^2 - x + 1 = (x + \omega)(x - 1 - \omega)$

To see how I know that, note that:
$$x^6 - 1 = (x^3 + 1)(x^3 - 1) = (x + 1)(x^2 - x + 1)(x - 1)(x^2 + x + 1)$$

Then after comparing the zeros of both sides of the above equation, we can conclude that the zeros of $(x^2 - x + 1)$ must be $-\omega$ and $-\omega^2$. And by the exercise on *pages 463-464*, I know that $\omega^2 = -1 - \omega$.

More boringly, you could also use quadratic formula to find the roots of $x^2 - x + 1$.

It follows that $p \mid (x_0 + \omega)(x_0 - 1 - \omega)$. Yet for identical reasons as in problem 7, we know $p \nmid (x_0 + \omega)$ and $p \nmid (x_0 - 1 - \omega)$. So, p is not prime in $\mathbb{Z}[\omega]$.

Now in problem 4 on the 9th problem set of math 200a on *pages 463-464*, we showed $\mathbb{Z}[\omega]$ is a P.I.D. Hence, p not being prime in $\mathbb{Z}[\omega]$ implies that p isn't irreducible either. ■

1/17/2026

To start off today, I want to do some more stuff with topological vector spaces. That said, I'll be taking a bit of a detour in Rudin functional analysis from what I was working on back on *page 528*.

Recall the Baire Category theorem which says that a complete metric space is not a countable union of nowhere dense sets. For reasons, in math 240b I didn't do the homework assignment exploring the consequences of this theorem. So I'd like to explore this theorem now and with perhaps slightly more generality than was possible in math 240b.

Baire Category Theorem For LCH Spaces: If X is an LCH space, then:

(a) the intersection of any countable collection of open dense sets in X is dense in X;

Proof:
Let $\{U_n\}_{n\in\mathbb{N}}$ be a sequence of open dense sets in X. Then it suffices to show for any nonempty open set $W \subseteq X$ that $W \cap (\bigcap_{n\in\mathbb{N}} U_n) \neq \emptyset$.

Now note that $U_1 \cap W$ is nonempty (since U_1 is dense in X) and open. It follows that for some $x_1 \in U_1 \cap W$ we can find a precompact open set V_1 such that:

$$x_1 \in V_1 \subseteq \overline{V_1} \subseteq U_1 \cap W.$$

Next, for all $n > 1$ we can inductively continue picking nonempty precompact open sets V_n such that $\overline{V_n} \subseteq U_n \cap V_{n-1}$ as follows:

$U_n \cap V_{n-1}$ is open and nonempty (the latter being true because U_n is dense in X and V_{n-1} is a nonempty open set). So let $x_n \in U_n \cap V_{n-1}$ and then pick a precompact open set V_n such that $x_n \in V_n \subseteq \overline{V_n} \subseteq U_n \cap V_{n-1}$.

Now $\{\overline{V_n}\}_{n\in\mathbb{N}}$ is a decreasing sequence of nonempty closed sets in $\overline{V_1}$ that have the finite intersection property. By the compactness of $\overline{V_1}$, we can conclude that:

$$V := \bigcap_{n\in\mathbb{N}} \overline{V_n} \neq \emptyset.$$

At the same time:

$$V \subseteq \overline{V_n} \subseteq U_n \cap V_{n-1} \subseteq U_n \text{ for all } n > 1 \text{ and also } V \subseteq \overline{V_1} \subseteq U_1 \cap W$$

It follows that $W \cap (\bigcap_{n\in\mathbb{N}} U_n) \neq \emptyset$ since it contains the nonempty set V.

(b) X is not a countable union of nowhere dense sets.

Proof:
If $\{E_n\}_{n\in\mathbb{N}}$ is any sequence of nowhere dense sets then we know that $\{(\overline{E_n})^\mathsf{C}\}_{n\in\mathbb{N}}$ is a sequence of open dense sets in X. It follows that $\bigcup_{n\in\mathbb{N}} E_n \neq X$ since $\bigcup_{n\in\mathbb{N}} E_n \subseteq \bigcup_{n\in\mathbb{N}} \overline{E_n}$ and:

$$(\bigcup_{n\in\mathbb{N}} \overline{E_n})^\mathsf{C} = \bigcap_{n\in\mathbb{N}} (\overline{E_n})^\mathsf{C} \neq \emptyset \text{ by claim (a). } \blacksquare$$

I wrote down this proof mostly because it's mostly similar to the proof for complete metric spaces and I wanted some record of the Baire category theory and how to prove it in my latex notes.

Next, following grandpa Rudin I'm going to cover the uniform boundedness principal in a bit more generality than in math 240b.

Recall that if X is a topological space, then $\mathcal{F} \subseteq C(X)$ is called equicontinuous if for all $x \in X$ and $\varepsilon > 0$, there exists a neighborhood $V \subseteq X$ of x with $|f(y) - f(x)| < \varepsilon$ for all $f \in \mathcal{F}$ and $y \in V$.

Given a topological vector space $\mathcal{Y}$, we can extend the spirit of the prior definition as follows.

We say $\mathcal{F} \subseteq C(X, \mathcal{Y})$ is equicontinuous if for any $x \in X$ and neighborhood $W \subseteq \mathcal{Y}$ of 0, there exists a neighborhood $V \subseteq X$ of x with $(f(y) - f(x)) \in W$ for all $y \in V$ and $f \in \mathcal{F}$.

Note that this maintains the key property that $f(y) \to f(x)$ uniformly over all $f \in \mathcal{F}$ as $y \to x$.

Additionally, note in particular that if $\mathcal{X}, \mathcal{Y}$ are topological vector spaces, then $\mathcal{F} \subseteq B(\mathcal{X}, \mathcal{Y})$ is equicontinuous iff for every neighborhood $W \subseteq \mathcal{Y}$ of 0 there exists a neighborhood $V \subseteq \mathcal{X}$ of 0 such that $A(V) \subseteq W$ for all $A \in \mathcal{F}$.

Theorem 2.4: Suppose $\mathcal{X}$ and $\mathcal{Y}$ are topological vector spaces, $\mathcal{F} \subseteq B(\mathcal{X}, \mathcal{Y})$ is an equicontinuous collection of linear maps, and E is a (Von Neumann) bounded subset of $\mathcal{X}$. Then $\mathcal{Y}$ has a bounded subset F such that $A(E) \subseteq F$ for all $A \in \mathcal{F}$.

Proof:
Let $F := \bigcup_{A\in\mathcal{F}} A(E)$. Then to prove that F is bounded, suppose W is any neighborhood of 0 in $\mathcal{Y}$. Since $\mathcal{F}$ is equicontinuous, there exists a neighborhood V of 0 in $\mathcal{X}$ such that $A(V) \subseteq W$ for all $A \in \mathcal{F}$. Then since E is bounded in $\mathcal{X}$, we know there exists $t > 0$ such that $E \subseteq tV$. Finally, note that $A(E) \subseteq A(tV) = tA(v) \subseteq tW$ for all $A \in \mathcal{F}$. Thus, $F \subseteq tW$. $\blacksquare$

Theorem 2.5 (Banach-Steinhaus): Suppose $\mathcal{X}$ and $\mathcal{Y}$ are topological vector spaces, $\mathcal{F} \subseteq B(\mathcal{X}, \mathcal{Y})$, and $E \subseteq \mathcal{X}$ is the set of all $x \in \mathcal{X}$ such that $\mathcal{F}(x) := \{Ax : A \in \mathcal{F}\}$ is bounded in $\mathcal{Y}$. If E is not meager, then $E = \mathcal{X}$ and $\mathcal{F}$ is equicontinuous.

(As a remind, being meager just means a set is a countable union of disjoint nonempty sets.)

Proof:
Let W be any neighborhood of 0 in $\mathcal{Y}$ and let U be a balanced neighborhood of 0 in $\mathcal{Y}$ such that $\overline{U} + \overline{U} \subseteq W$. We can do this by (Folland) exercise 11.1 (see *pages 527-528*) as well as other standard arguments. Next put $S := \bigcap_{A \in \mathcal{F}} A^{-1}(\overline{U})$.

If $x \in E$, then $\mathcal{F}(x) \subseteq nU$ for some $n \in \mathbb{N}$. In turn $x \in \bigcap_{A \in \mathcal{F}} A^{-1}(nU) = nS$. So, we can say that $E \subseteq \bigcup_{n=1}^{\infty} nS$. Now as E is nonmeager, we know that at least one nS is also nonmeager. Then by applying the homeomorphism $x \mapsto n^{-1}x$, we get that S is nonmeager. But S is closed because each $A \in \mathcal{F}$ is continuous. Thus, S must have a nonempty interior (as otherwise S would be nowhere dense and thus trivially meager).

Let $x \in S^{\circ}$. Then $x + (-S)$ contains a neighborhood V of 0 in $\mathcal{X}$. Also:

$$A(V) \subseteq Ax + (-A(S)) \subseteq \overline{U} + \overline{U} \subseteq W \text{ for every } A \in \mathcal{F}.$$

This proves that $\mathcal{F}$ is equicontinuous. Finally, the fact that $E = \mathcal{X}$ is a simple corollary of theorem 2.4 applied to the set $\{x\}$ for all $x \in \mathcal{X}$. ■

Since I was pondering it, I thought I might as well mention it. Meagerness is a particularly nice way of characterizing whether a set is big or small because it is easy to see that if $A \subseteq B$ and B is meager, then so is A. (Also consequently, if $A \subseteq B$ and A is nonmeager, then so is B).

On a different note, if $\mathcal{X}$ is a Banach space and $\mathcal{Y}$ is a normed space, then $\mathcal{F} \subseteq B(\mathcal{X}, \mathcal{Y})$ is equicontinuous if and only if $\sup_{T \in \mathcal{F}} \|T\|_{\text{op}} < \infty$.

$(\Longrightarrow)$
Consider the set $E = \{x \in \mathcal{X} : \|x\| < 1\}$. By theorem 2.4, there exists a bounded subset $F \subseteq \mathcal{Y}$ with $T(E) \subseteq F$ for all $T \in \mathcal{F}$. In particular, after letting $C = \sup_{y \in F} \|y\| < \infty$, we have that $\|T\|_{\text{op}} \leq Cy$ for all $T \in \mathcal{F}$.

$(\Longleftarrow)$
Suppose W is any open neighborhood of 0 in $\mathcal{Y}$. Then there exists $M > 0$ such that $W' = \{y \in \mathcal{Y} : \|y\| < M_1\} \subseteq W$. So, let $V = \{x \in \mathcal{X} : \|x\| < \frac{M}{C}\}$.
Then $Tx \in W' \subseteq W$ for all $x \in V$ and $T \in \mathcal{F}$.

Thus, theorem 2.5 above is purely a generalization of the uniform boundedness principle in my math 240b notes.

Where the Baire Category theorem becomes relevant is that any Fréchet space (and thus also any Banach or Hilbert space) is a complete metric space and thus nonmeager. So, we can come up with some really powerful theorems about linear maps to and from Fréchet spaces.

Lemma: If $\mathcal{X}$ is a topological vector space whose topology is generated by a family $\{p_\alpha\}_{\alpha\in A}$ of seminorms, then $E \subseteq \mathcal{X}$ is Von Neumann bounded if and only if $\sup_{x\in E} p_\alpha(x) < \infty$ for all $\alpha \in A$.

($\Longrightarrow$)
Let $U_\alpha = \{x' \in \mathcal{X} : p_\alpha(x') < 1\}$. Since E is bounded, we know that $E \subseteq tU_\alpha$ for some $t > 0$ In turn, $p_\alpha(x) < t$ for all $x \in E$.

($\Longleftarrow$)
Suppose U is any open neighborhood of 0 in $\mathcal{X}$. Then there exists $\alpha_1, \ldots, \alpha_n \in A$ and $\varepsilon_1, \ldots, \varepsilon_n > 0$ such that $\bigcap_{k=1}^n \{x' \in \mathcal{X} : p_{\alpha_k}(x') < \varepsilon_k\} \subseteq U$. In particular, by taking $\varepsilon = \min(\varepsilon_1, \ldots, \varepsilon_n)$ we get that:

$$\bigcap_{k=1}^n \{x' \in \mathcal{X} : p_{\alpha_k}(x') < \varepsilon\} \subseteq U.$$

Now fix $C > 0$ such that $p_{\alpha_k}(x) \leq C$ for all $x \in E$ and $1 \leq k \leq n$. Then:

$$E \subseteq \tfrac{C}{\varepsilon} \bigcap_{k=1}^n \{x' \in \mathcal{X} : p_{\alpha_k}(x') < \varepsilon\} \subseteq \tfrac{C}{\varepsilon} U. \blacksquare$$

Before going home and having to work on other math, I want to do an exercises from Folland Real Analysis.

Exercise 5.39: Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be Banach spaces and let $B : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a separately continuous bilinear map. That is: $B(x, \cdot) \in B(\mathcal{Y}, \mathcal{Z})$ for all $x \in \mathcal{X}$ and $B(\cdot, y) \in B(\mathcal{X}, \mathcal{Z})$ for all $y \in \mathcal{Y}$. Then B is jointly continuous.

Consider any fixed $x \in \mathcal{X}$. Then for all $y \in \mathcal{Y}_1$ we have that:

$$\|B(x,y)\| \leq \|B(x,\cdot)\|_{\mathrm{op}} \|y\| \leq \|B(x,\cdot)\|_{\mathrm{op}}.$$

Thus $\sup_{y\in\mathcal{Y}_1} \|B(x,y)\| < \infty$ for all $x \in \mathcal{X}$. And by the uniform boundedness principle we can conclude that $\sup_{y\in\mathcal{Y}_1} \|B(\cdot,y)\|_{\mathrm{op}} < \infty$. So, we can pick $C \geq 0$ such that $\|B(\cdot,y)\|_{\mathrm{op}} \leq C$ when $\|y\| \leq 1$.

Now suppose $(x,y) \in \mathcal{X} \times \mathcal{Y}$. If $y \neq 0$ then:

$$\|B(x,y)\| = \|B(x, \tfrac{y}{\|y\|})\| \|y\| \leq C\|x\|\|y\|$$

Meanwhile, if $y = 0$ then trivially $\|B(x,y)\| = 0 \leq C\|x\|\|y\|$. So, we have found some constant $C \geq 0$ such that:

$$\|B(x,y)\| \leq C\|x\|\|y\| \text{ for all } x \in \mathcal{X} \text{ and } y \in \mathcal{Y}.$$

Finally, I claim this means B is jointly continuous. After all, suppose $x_n \to x$ in $\mathcal{X}$ and $y_n \to y$ in $\mathcal{Y}$. Then:

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

Since $\|x_n\| \to \|x\|$ as $n \to \infty$, we know that $\|x_n\|$ is bounded as $n \to \infty$. In turn, it is easy to see that $C\|x_n\|\|y_n - y\| + C\|x_n - x\|\|y\| \to 0$ as $n \to \infty$.

So, we've proven that $B(x_n, y_n) \to B(x,y)$ as $n \to \infty$. $\blacksquare$

I'll work on more functional analysis soon on *page* ___.

Math 220b Lecture Notes:

Suppose $\{p_k\}_{k\in\mathbb{N}} \subseteq \mathbb{C}$. Then how should we define an infinite product $\prod_{k=1}^{\infty} p_k$?

The naive thing to do is to just define $\prod_{k=1}^{\infty} p_k = \lim_{n\to\infty} \prod_{k=1}^{n} p_k$ if the latter limit exists. This is precisely what I assumed the definition to be on *pages 51-52* of my journal. Also, when working with formal powers this is also how I defined infinite products. However, for the purposes of this class this definition is not desirable. This is because:

- We want $\mathbb{C}$ to remain an integral domain with respect to our infinite product. In other words, we'd like to consider $\prod_{k=1}^{\infty} \frac{1}{k}$ to be a divergent product.
- We want a definition of an infinite product where noncovergent products don't converge to 0 if we append 0 to the sequence we're producting over.

I also had a weirder definition or products on *page 220*. I came up with this on my own without outside help specifically to study probability measures. As a result, the definition on that page is utterly unsuited to working with general complex numbers.

Instead we define:
$\prod_{k=1}^{\infty} p_k := P$ iff $\exists M > 0$ such that $\lim\limits_{n\to\infty} \prod_{k=M}^{n} p_k = \widehat{P} \neq 0$ and $P = p_1 \cdots p_{M-1}\widehat{P}$.

Observations:

1. Suppose $M < M'$, $\lim_{n\to\infty} \prod_{k=M}^{n} p_k = \widehat{P}$, and $\lim_{n\to\infty} \prod_{k=M'}^{n} p_k = \widehat{Q}$ Then it is easy to see that $\widehat{P} = p_M p_{M+1} \cdots p_{M'-1}\widehat{Q}$. Therefore, $\prod_{k=1}^{\infty} p_k$ does not depend on M.
2. In a convergent product, only finitely many p_k are zero. After all, otherwise
$$\lim_{n\to\infty} \prod_{k=M}^{n} p_k = 0 \text{ for all } M \in \mathbb{N}.$$
3. $\prod_{k=1}^{\infty} p_k = 0$ iff there exists some $k' \in \mathbb{N}$ with $p_{k'} = 0$.
4. Suppose $\prod_{k=1}^{\infty} p_k = p_1 \cdots p_{M-1}\widehat{P}$ where $\widehat{P} = \lim_{n\to\infty} \prod_{k=M}^{n} p_k \neq 0$. Then for $m > M$, we have that:
$$p_m = \frac{\prod_{k=M}^{m} p_k}{\prod_{k=M}^{m-1} p_k} \to \frac{\widehat{P}}{\widehat{P}} = 1 \text{ as } m \to \infty.$$
Because of this, we typically write $p_k = 1 + a_k$ for each k. Then we know that if the product converges, $a_k \to 0$.

The following lemmas will be reminiscent of my proof of (Folland exercise 1.32) on *pages 51-52*. That said, the trick of using logarithms will not work as directly in the complex analysis case as in the real analysis case.

We'll use the following conventions in this class:

- We denote the principal branch of the logarithm as $\mathrm{Log} : \mathbb{C} - \mathbb{R}_{\leq 0} \to \mathbb{C}$ where $\mathrm{Log}(re^{i\theta}) = \log(r) + i\theta$ and $\theta \in (-\pi, \pi)$.
- If α is fixed, we define $\mathrm{Log}_\alpha : \mathbb{C} - e^{i\alpha}\mathbb{R}_{\geq 0} \to \mathbb{C}$ by $\mathrm{Log}_\alpha(re^i\theta) = \log(r) + i\theta$ where $\theta \in (\alpha, \alpha + 2\pi)$.

Lemma: $\prod_{k=1}^{\infty}(1 + a_k)$ converges iff $\exists M \in \mathbb{N}$ such that $\sum_{k=M}^{\infty} \mathrm{Log}(1 + a_k)$ converges.

$(\Longleftarrow)$
Let $S_n = \sum_{k=M}^{n} \mathrm{Log}(1 + a_k)$ and suppose $S_n \to S$. Then set $P_n = \prod_{k=M}^{n}(1 + a_k)$. Importantly, $P_n = e^{S_n}$. Thus by the coninuity of the exponential function, we know that $P_n \to e^S$ as $n \to \infty$. But $e^z \neq 0$ for any z. So, $\widehat{P} = e^S \neq 0$.

This proves that $\prod_{k=1}^{\infty}(1 + a_k)$ converges.

$(\Longrightarrow)$
Let $M \in \mathbb{N}$ satisfy that $P_n = \prod_{k=M}^{n}(1 + a_k) \to \widehat{P} \neq 0$ as $n \to \infty$ and $|a_k| < 1$ for all $k \geq M$. Then choose $\alpha \in \mathbb{R}$ such that $\widehat{P} \notin e^{i\alpha}\mathbb{R}_{\geq 0}$. Since $P_n \to \widehat{P}$, we know that for n sufficiently large, $\mathrm{Log}_\alpha(P_n)$ will be defined.

Like in the other proof direction, let $S_n = \sum_{k=M}^{n} \mathrm{Log}(1 + a_k)$. In turn, for each n we have that $e^{S_n} = P_n = e^{\mathrm{Log}_\alpha(P_n)}$. So, there exists some sequence $\{\ell_n\}_{n \in \mathbb{N}} \subseteq \mathbb{Z}$ such that $S_n = \mathrm{Log}_\alpha(P_n) + 2\pi i \ell_n$ for all n.

But $\mathrm{Log}(1 + a_n) = S_n - S_{n-1} = \mathrm{Log}_\alpha(P_n) - \mathrm{Log}_\alpha(P_{n-1}) + 2\pi i(\ell_n - \ell_{n-1})$. By taking $n \to \infty$ we get that $\mathrm{Log}(1 + a_n) \to \mathrm{Log}(1) = 0$. Similarly, $\mathrm{Log}_\alpha(P_n) - \mathrm{Log}_\alpha(P_{n-1}) \to 0$ as $n \to \infty$. Hence, we may conclude that:

$$\ell_n - \ell_{n-1} \to 0 \text{ as } n \to \infty.$$

That combined with the fact that ℓ_n is integer valued means there must exist some $N \in \mathbb{N}$ such that ℓ_n is some constant ℓ for all $n \geq N$. But finally, we now have that:

$$S_n = \mathrm{Log}_\alpha(P_n) + 2\pi i \ell_n \to \mathrm{Log}_\alpha(\widehat{P}) + 2\pi i \ell \text{ as } n \to \infty. \blacksquare$$

Based on the above lemma, we shall now define $\prod_{k=1}^{\infty}(1 + a_k)$ as converging absolutely if and only if there exists $M \in \mathbb{N}$ such that $\sum_{k=M}^{\infty} \mathrm{Log}(1 + a_k)$ converges absolutely.

As for why we don't define absolute convergence to be if $\prod_{k=1}^{\infty} |1 + a_k|$ converges, note that $\prod_{k=1}^{\infty} |1 + a_k|$ can converge even if $\prod_{k=1}^{\infty}(1 + a_k)$. Preferably though, we'd want absolute convergence to be a stronger property though than normal convergence.

Lemma: The following are equivalent:

(a) $\prod_{k=1}^{\infty}(1 + a_k)$ converges absolutely;

(b) $\sum_{k=1}^{\infty} a_k$ converges absolutely;

(c) $\prod_{k=1}^{\infty}(1 + |a_k|)$ converges.

Before proving the above lemma, it will help to define a constant. Also, I will be continuing to use this constant in other proofs so I will be marking this result with an asterisk.

Note that the power series expansion for $\mathrm{Log}(1+z)$ about 0 is $z - \frac{1}{2}z^2 + \cdots$. In turn, $\frac{\mathrm{Log}(1+z)}{z} = 1 - \frac{1}{2}z + \cdots$. And by taking the limit of the latter expression as $z \to 0$ (and using the fact that power series are continuous), we get that: $\lim_{z\to\infty} \frac{\mathrm{Log}(1+z)}{z} = 1$.

It follows that there exists $0 < \delta < 1$ such that $\frac{1}{2} \leq \left|\frac{\mathrm{Log}(1+z)}{z}\right| \leq \frac{3}{2}$ when $|z| < \delta$. In turn we also have that:

$$\tfrac{1}{2}|z| \leq |\,\mathrm{Log}(1+z)| \leq \tfrac{3}{2}|z| \text{ when } |z| < \delta. \; (*)$$

Proof:

$(a \implies b)$

By definition, (a) holds if and only if there exists some M such that $\sum_{k=M}^{\infty} \mathrm{Log}(1+a_k)$ converges absolutely. Yet in that case we know that $a_k \to 0$. So, there exists $N \geq M$ such that $|a_k| < \delta$ (see $(*)$ when $k \geq N$. By comparison test, we can thus conclude that $\sum_{k=N}^{\infty} a_k$ converges absolutely. And because $\sum_{k=1}^{N-1} |a_k|$ is finite, we can also conclude that $\sum_{k=1}^{\infty} a_k$ converges absolutely.

$(b \implies a)$

If $\sum_{k=1}^{\infty} a_k$ converges absolutely, then we know that $a_k \to 0$. So there must exist $M > 0$ such that $|a_k| < \delta$ (see $(*)$) for all $k \geq M$. In turn, by the comparison test we can conclude that $\sum_{k=M}^{\infty} \mathrm{Log}(1+a_k)$ converges absolutely.

$(c \iff b)$

By our previous lemma, we know that (c) holds if and only if $\sum_{k=1}^{\infty} \mathrm{Log}(1+|a_k|)$ converges (and we don't need to worry about an M here since $1+|a_k|$ is always in the domain of the principal branch). Then by the comparison test using $(*)$, we can conclude that $\sum_{k=1}^{\infty} \mathrm{Log}(1+|a_k|)$ converges if and only if (b) holds. ■

A useful fact is as follows: If $\prod_{k=1}^{\infty}(1+a_k)$ absolutely converges to P, then $\prod_{k=1}^{\infty}(1+a_{\sigma(k)})$ also absolutely converges to P for any permutations $\sigma \in S_{\mathbb{N}}$.

Why?

By part (b) of our prior lemma plus the fact absolute convergence of sums isn't affected by rearranging the sequence we're summing over, we have that $\prod_{k=1}^{\infty}(1+a_k)$ absolutely converges if and only if $\prod_{k=1}^{\infty}(1+a_{\sigma(k)})$ does.

Next, pick $M \in \mathbb{N}$ large enough so that $|a_k|, |a_{\sigma(k)}| < 1$ for all $k \geq M$. Then let $r_1, \ldots, r_m$ be all the integers in $\{1, \ldots, M-1\}$ satisfing that $\sigma(r_j) \geq M$ for each j and let $s_1, \ldots, s_\ell$ be all the integers $\geq M$ satisfying that $\sigma(s_j) < M$ for each j. Since $\sum_{k=M}^{\infty} \mathrm{Log}(1+a_k)$ absolutely converges, we have that:

$\sum_{k=M}^{\infty} \mathrm{Log}(1+a_{\sigma(k)}) = \sum_{k=M}^{\infty} \mathrm{Log}(1+a_k) + \sum_{j=1}^{m} \mathrm{Log}(1+a_{r_j}) - \sum_{j=1}^{\ell} \mathrm{Log}(1+a_{s_j})$

Finally:

$$\prod_{k=1}^{\infty}(1+a_\sigma(k)) = (\prod_{k=1}^{M-1}(1+a_{\sigma(k)}))\exp(\sum_{k=M}^{\infty}\operatorname{Log}(1+a_{\sigma(k)}))$$

$$= (\prod_{k=1}^{M-1}(1+a_{\sigma(k)}))\exp(\sum_{k=M}^{\infty}\operatorname{Log}(1+a_k)+\sum_{j=1}^{m}\operatorname{Log}(1+a_{r_j})-\sum_{j=1}^{\ell}\operatorname{Log}(1+a_{s_j}))$$

$$= (\prod_{k=1}^{M-1}(1+a_k))\exp(\sum_{k=M}^{\infty}\operatorname{Log}(1+a_k)) = \prod_{k=1}^{\infty}(1+a_k). \blacksquare$$

Now we are ready to talk about infinite products of functions. Suppose $U \subseteq \mathbb{C}$ is open and $\{f_k\}_{k\in\mathbb{N}} \subseteq O(U)$ satisfies that $\sum_{k=1}^{\infty} f_k$ converges absolutely locally uniformly. Then it is well-defined to set:

$$F(z) := \prod_{k=1}^{\infty}(1+f_k(z)) \text{ for all } z \in U.$$

Before getting to the next theorem, note in this class that if $g \in O(U)$ and $g(a) = 0$, then we denote $\operatorname{ord}(g,a)$ to be the order of the zero of g at a.

Theorem:

1. The product converges absolutely and locally uniformly to a holomorphic function.
2. $F(a) = 0$ if and only if there exists k such that $1 + f_k(a) = 0$. Moreover;
$$\operatorname{ord}(F,a) = \sum_{k=1}^{\infty}\operatorname{ord}(1+f_k,a).$$

Proof:
Consider any $a \in u$ and fix $r > 0$ such that $\overline{\Delta}(a,2r) \subseteq U$. Then $\sum|f_k|$ converges uniformly on $\overline{\Delta}(a,2r)$, which in turn tells us that $|f_k|$ converges uniformly to 0 on $\overline{\Delta}(a,2r)$. But now recall the constant δ from $(*)$. We know there must exist N such that $|f_k(z)| < \delta$ for all $k \geq N$ and $z \in \overline{\Delta}(a,2r)$. Thus by comparison test, we can conclude that $\sum_{k=N}^{\infty}\operatorname{Log}(1+f_k)$ converges absolutely uniformly on $\overline{\Delta}(a,2r)$ to a function h.

By definition this shows that F converges absolutely on $\overline{\Delta}(a,2r)$. Also by the Weierstraß convergence theorem, we know that h is holomorphic on $\Delta(a,2r)$. In turn:

$$F := (\prod_{k=1}^{N-1}(1+f_k))\exp(\sum_{k=N}^{\infty}\operatorname{Log}(1+f_k)) = (\prod_{k=1}^{N-1}(1+f_k))e^h$$

is also holomorphic on $\Delta(a,2r)$.

Finally, we rely on two lemmas I shall prove below.

Lemma 1: If K is a compact set, $h_n, h \in C(K)$ satisfy that $h_n \xrightarrow{u.} h$, and g is continuous on $\mathbb{C}$, then $g \circ h_n \xrightarrow{u.} g \circ h$.

Proof:
Since $h(K)$ is compact, we know there exists $R > 0$ such that $h(K) \subseteq \Delta(0,R)$. Then as $h_n \xrightarrow{u.} h$, we can find some N such that $h_n(K) \subseteq \Delta(0,R+1)$ for all $n \geq N$. Also importantly, we have that g is uniformly continuous on $\overline{\Delta}(0,R+1)$.

Consider $\varepsilon > 0$. Then let $\delta > 0$ be such that $|g(x) - g(y)| < \varepsilon$ whenever $x, y \in \overline{\Delta}(0, R+1)$ with $|x-y| < \delta$. Thirdly, pick $M > N$ such that $|h_n(x) - h(x)| < \delta$ for all $n \geq M$ and $x \in K$. Then $|(g \circ h_n)(x) - (g \circ h)(x)| < \varepsilon$ for all $n \geq M$ and $x \in K$. So, $g \circ h_n \xrightarrow{u.} g \circ h$. ■

An obvious corollary is that if $h_n \xrightarrow{u.} h$ on a compact set, then also $\exp(h_n) \xrightarrow{u.} \exp(h)$ on that compact set.

Lemma 2: If X is a topological space, $f_n, g_n, f, g \in C(X)$ satisfy that $f_n \xrightarrow{u.} f$ and $g_n \xrightarrow{u.} g$, $\sup_{n \in \mathbb{N}} \|f_n\|_u = \alpha < \infty$, and $\|g\|_u < \infty$, then $f_n g_n \xrightarrow{u.} fg$.

Proof:

$$\begin{aligned} \|f_n g_n - fg\|_u &\leq \|f_n g_n - f_n g\|_u + \|f_n g - fg\|_u \\ &\leq \|f_n\|_u \|g_n - g\|_u + \|f_n - f\|_u \|g\|_u \\ &\leq \alpha \|g_n - g\|_u + \|f_n - f\|_u \|g\|_u \to 0 \text{ as } n \to \infty. \blacksquare \end{aligned}$$

An obvious corollary is that if K is compact, g is continuous on K, and $h_n \xrightarrow{u.} h$ on K, then $gh_n \xrightarrow{u.} gh$ on K as well.

Also as a side note: Why did I just process that $\|\cdot\|_u$ turns $C(K)$ into a Banach algebra?!

Based on the last two lemmas, we can easily conclude that:

$$\prod_{k=1}^{m}(1+f_k) = \left(\prod_{k=1}^{N-1}(1+f_k)\right)\exp\left(\sum_{k=N}^{m}\operatorname{Log}(1+f_k)\right)$$

uniformly converges to $(\prod_{k=1}^{N-1}(1+f_k))e^h = F$ on $\overline{\Delta}(a, r)$.

To finish off, we can just cover any compact set $K \subseteq U$ with finitely many of those balls $\overline{\Delta}(a, r)$ to get that $\prod_{k=1}^{m}(1+f_k) \xrightarrow{\ell.u.} F$ as $m \to \infty$. This proves part 1.

To show part 2, continue letting a, N, and h be from before. Then note that $F(a) = 0$ if and only if $e^{h(a)} \prod_{k=1}^{N-1}(1+f_k(a)) = 0$. But $e^{h(a)}$ can't be zero. So, we equivalently have that there exists $k \in \{1, \ldots, N-1\}$ such that $1 + f_k(a) = 0$.

Furthermore, $\operatorname{ord}(F, a) = \operatorname{ord}(e^h \prod_{k=1}^{N-1}(1+f_k), a) = 0 + \sum_{k=1}^{N-1} \operatorname{ord}(1+f_k, a)$. As for $k \geq N$, since $|f_k(z)| < \delta < 1$ for all $z \in \overline{\Delta}(a, 2r)$ we know that $(1+f_k)$ has no zeros on $\overline{\Delta}(a, 2r)$. ■

Now that we've proven that the derivative of $\prod_{k=1}^{\infty}(1+f_k(z))$ exists when $\sum_{k=1}^{\infty} f_k$ converges absolutely locally uniformly, it'd be nice to have a way of calculating it. So, we bring in the logarithmic derivative.

Note that if $g_1, \ldots, g_n$ are holomorphic functions and $h = \prod_{k=1}^{n} g_k$, then:

$$\frac{h'}{h} = \sum_{k=1}^{n} \frac{g_k'}{g_k}.$$

We shall see that this generalizes to the case of infinite products.

Theorem: Let $f_k : U \to \mathbb{C}$ be holomorphic and suppose $\sum_{k=1}^{\infty} f_k$ converges absolutely locally uniformly. Then after setting $g_k = 1 + f_k$ for all k, we have that $F(x) := \prod_{k=1}^{\infty} g_k$ converges absolutely and locally uniformly. We claim that:

$\frac{F'}{F} = \sum_{k=1}^{\infty} \frac{g_k'}{g_k}$ with the latter converging locally uniformly on $U - F^{-1}(\{0\})$.

Proof:
Let Δ be an open disk such that $\overline{\Delta} \subseteq U - F^{-1}(\{0\})$. Then like in the proof of the previous theorem there exists $N > 0$ such that $h = \sum_{k=N}^{\infty} \operatorname{Log}(1 + f_k)$ is well-defined and converges uniformly in $\overline{\Delta}$. In turn:

$$e^h = \prod_{k=N}^{\infty}(1 + f_k) = \prod_{k=N}^{\infty} g_k \text{ on } \overline{\Delta}.$$

But now as $F = e^h \prod_{k=1}^{N-1} g_k$, we have by our knowledge of logarithmic derivatives of finite products that:

$$\frac{F'}{F} = \frac{e^h h'}{e^h} + \sum_{k=1}^{N-1} \frac{g_k'}{g_k} = h' + \sum_{k=1}^{N-1} \frac{g_k'}{g_k} \text{ on } \Delta$$

Also, by the Weierstrass convergence theorem, we have that:

$$h' = \lim_{m\to\infty} \left(\sum_{k=N}^{m} \operatorname{Log}(1 + f_k)\right)' = \lim_{m\to\infty} \sum_{k=N}^{m} \frac{f_k'}{1+f_k} = \sum_{k=N}^{\infty} \frac{g_k'}{g_k}$$

where the latter converges locally uniformly on Δ.

Since the center of Δ was arbitrary, we can conclude that:

$\frac{F'}{F} = \sum_{k=1}^{\infty} \frac{g_k'}{g_k}$ with the latter converging locally uniformly on $U - F^{-1}(\{0\})$. ■

In 1734, as part of writing the original "proof" that $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$, Euler claimed that:

$$\sin(\pi z) = \pi z \prod_{k=1}^{\infty}\left(1 - \frac{z^2}{k^2}\right). \; (*)$$

We shall rigorously prove this product converges absolutely and locally uniformly. But first a homework problem will be useful.

Math 220b Set 2 Problem 5: Show that if $f, g : G \to \mathbb{C}$ are holomorphic functions in a simply connected region G that have the same zeros with the same multiplicty, then there exists a holomorphic function $h : G \to \mathbb{C}$ such that $f = e^h g$.

Note if both f and g are the constant zero functions, then the exercise is trivial as we can just set $h \equiv 1$. So, we may assume all zeros of f and g finite multiplicity.

Consider the function f/g. Since f and g have the same zeros with the same multiplicty, we can holomorphically extend f/g to the points w where $g(w) = 0$. Furthermore, our extension of f/g never vanishes on G.

$f(z)$ is never zero when g isn't. So we don't get a zero that way. Also suppose $f(w) = 0 = g(w)$ and let ℓ be the multiplicity of w for both functions.

After writing $f = (z - w)^{\ell} \widetilde{f}$ and $g = (z - w)^{\ell} \widetilde{g}$ where $\widetilde{f}, \widetilde{g}$ are holomorphic functions on G not vanishing at w, we have that:

$$\lim_{z \to w} \frac{f(z)}{g(z)} = \frac{\tilde{f}(w)}{\tilde{g}(w)} \neq 0$$

Yet now by (Conway) Theorem IV.6.17 (on *page 416*), we know there exists $h \in O(G)$ such that $e^h = {}^f/_g$. Or in other words, $f = e^h g$.

(I was a bit of clumbsy with the last step so I'll add this note:
Technically e^h equals the extension of the function ${}^f/_g$. That said, this is enough to tell us that $e^{h(z)} = {}^{f(z)}/_{g(z)}$ when $g(z) \neq 0$. And then as $ge^h = f$ on a set with a limit point, we can conclude that they equal everywhere on G.) ■

As a side note, the assumption that G is simply connected is necessary.

Let $G = \mathbb{C} - \{0\}$ and then set $f(z) = z + 1$ and $g(z) = z^{-1} + 1$. Now f and g are both holomorphic functions on G whose only zero is at $z = -1$. Also $f'(-1) = 1$ and $g'(-1) = -1$. So, we can see that -1 is a zero with multiplicity 1 for both f and g.

Yet note that if $h : G \to \mathbb{C}$ satisfies that $f = e^h g$, then we'd have that:

$$z + 1 = e^h(\tfrac{1}{z} + 1) \Longrightarrow e^h = \frac{z+1}{z^{-1}+1} = \frac{z(z+1)}{1+z} = z$$

This says that h would have to be a branch of the logarithm defined on $\mathbb{C} - \{0\}$. Yet we know no such branch exists.

Returning to the subject of proving $(*)$:

Claim 1: $F(z) := \pi z \prod_{k=1}^{\infty}(1 - \frac{z^2}{k^2})$ absolutely converges locally uniformly on $\mathbb{C}$.

Set $g_k(z) = \frac{-z^2}{k^2}$. Then it's easy to see that $\sum_{k=1}^{\infty} g_k$ converges normally (and thus absolutely locally uniformly) on $\mathbb{C}$. For a refresher on what normal convergence is, check *page 530*.

As a result $\prod_{k=1}^{\infty}(1 + g_k)$ converges locally uniformly.

Claim 2: $\sin(\pi z)$ and $F(z)$ have the same zeros (with the same multiplicities) on $\mathbb{C}$.

$\sin(\pi z) = 0$ iff $z \in \mathbb{Z}$, and furthermore $\pi \cos(\pi z) = \pm\pi$ for all $z \in \mathbb{Z}$. This tells us that the all zeros of $\sin(\pi z)$ have multiplicity 1.

Meanwhile, by the theorem before we covered logarithmic derivatives, we also have that F also has zeros of multiplicity 1 at each integer and no other zeros.

By the prior homework problem, we know there exists an entire function $h : \mathbb{C} \to \mathbb{C}$ such that $\sin(\pi z) = e^{h(z)}F(z) = e^{h(z)} \cdot \pi z \prod_{k=1}^{\infty}(1 - \frac{z^2}{k^2})$. Taking the logarithmic derivative of both sides, we get that:

$$\pi \cot(\pi z) = \frac{\pi \cos(\pi z)}{\sin(\pi z)} = \frac{e^{h(z)}h'(z)}{h'(z)} + \frac{\pi}{\pi z} + \sum_{k=1}^{\infty} \frac{-\frac{2}{k^2}z}{(1-\frac{z^2}{k^2})} = h'(z) + \frac{1}{z} + \sum_{k=1}^{\infty} \frac{2z}{z^2-k^2}$$

Claim (Conway exercise V.2.8): $\pi \cot(\pi z) = \frac{1}{z} + \sum_{k=1}^{\infty} \frac{2z}{z^2-k^2}$.

Proof:
Consider the polygonal path;

$$\gamma = [m + \tfrac{1}{2} + mi, -m - \tfrac{1}{2} + mi, -m - \tfrac{1}{2} - mi, m + \tfrac{1}{2} - mi, m + \tfrac{1}{2} + mi].$$

For a drawing of the polygonal path, see the next page. Our first goal will be to evaluate the path integral $\int_\gamma \frac{\pi \cot(\pi z)}{z^2-a^2}dz$ using the residue theorem while assuming $a \notin \mathbb{Z}$ and that m is large enough so that $n(\gamma; a) = 1$.

Firstly, note that the poles of $\frac{\pi\cot(\pi z)}{z^2-a^2}$ whose winding number relative to γ is nonzero are precisely $0, \pm 1, \ldots, \pm m$, and $\pm a$. Furthermore, the winding numbers of those points are all 1. So, we get that:

$$\frac{1}{2\pi i}\int_\gamma \frac{\pi\cot(\pi z)}{z^2-a^2}\mathrm{d}z = \operatorname{Res}(\tfrac{\pi\cot(\pi z)}{z^2-a^2}, a) + \operatorname{Res}(\tfrac{\pi\cot(\pi z)}{z^2-a^2}, -a) + \sum_{k=-m}^{m}\operatorname{Res}(\tfrac{\pi\cot(\pi z)}{z^2-a^2}, k)$$

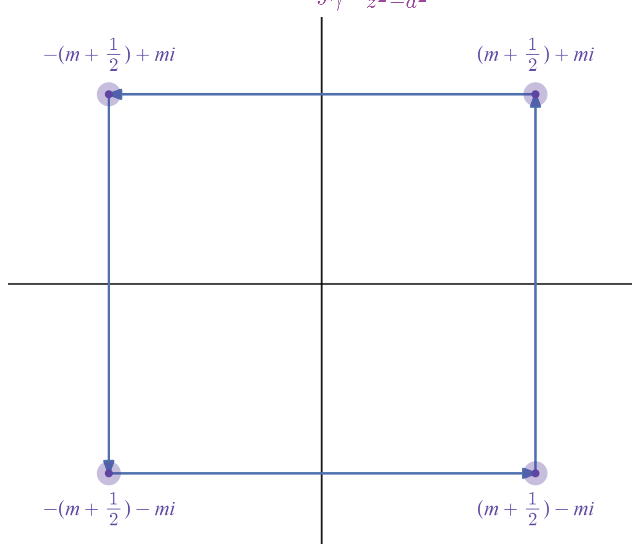
- Suppose $k \in \mathbb{Z}$. Then, we can check (using the fact that $\cot(\pi z)$ is 1-periodic) that:
$$\begin{aligned}\lim_{z\to k}(z-k)\pi\cot(\pi z) &= \lim_{z\to 0}\pi z\cot(\pi z)\\ &= \lim_{w\to 0} w\cot(w) = (\lim_{w\to 0}\tfrac{\sin(w)}{w})^{-1}\cdot\lim_{w\to 0}\cos(w) = 1\end{aligned}$$
In turn, $\lim\limits_{z\to k}(z-k)\frac{\pi\cot(\pi z)}{z^2-a^2} = \frac{1}{k^2-a^2}$. And thus, we can conclude that:
$$\operatorname{Res}(\tfrac{\pi\cot(\pi z)}{z^2-a^2}, k) = \tfrac{1}{k^2-a^2} \text{ for all } k \in \mathbb{Z}.$$

- $(z-a)\frac{\pi\cot(\pi z)}{z^2-a^2} = \frac{\pi\cot(\pi z)}{z+a} \to \frac{\pi\cot(\pi a)}{2a}$ as $z \to a$. Hence:
$$\operatorname{Res}(\tfrac{\pi\cot(\pi z)}{z^2-a^2}, a) = \tfrac{\pi\cot(\pi a)}{2a}.$$

- $(z+a)\frac{\pi\cot(\pi z)}{z^2-a^2} = \frac{\pi\cot(\pi z)}{z-a} \to \frac{\pi\cot(-\pi a)}{-2a} = \frac{\pi\cot(\pi a)}{2a}$ as $z \to -a$. Hence:
$$\operatorname{Res}(\tfrac{\pi\cot(\pi z)}{z^2-a^2}, a) = \tfrac{\pi\cot(\pi a)}{2a} \text{ as well.}$$

At last, we can conclude that:
$$\begin{aligned}\frac{1}{2\pi i}\int_\gamma \frac{\pi\cot(\pi z)}{z^2-a^2}\mathrm{d}z &= \frac{\pi\cot(\pi a)}{2a} + \frac{\pi\cot(\pi a)}{2a} + \sum\nolimits_{k=-m}^{m}\frac{1}{k^2-a^2}\\ &= \frac{\pi\cot(\pi a)}{a} - \frac{1}{a^2} - \sum\nolimits_{k=1}^{m}\frac{2}{a^2-k^2}\end{aligned}$$

But now to finish off, we take the limit of $\int_\gamma \frac{\pi\cot(\pi z)}{z^2-a^2}\mathrm{d}z$ as $m \to \infty$.



Unfortunately we now need to bound $\cot(\pi z)$ on the path of γ. So what follows is a bunch of trig manipulations:

To start off, we have that:
$$\begin{aligned}\cos(x+iy) &= \sin(x)\sin(iy) - \cos(x)\cos(iy)\\ &= i\sin(x)\sinh(y) - \cos(x)\cosh(y)\end{aligned}$$

Therefore, $|\cos(x+iy)|^2 = \sin^2(x)\sinh^2(y) + \cos^2(x)\cosh^2(y)$. By noting that $\cosh^2(y) = 1 + \sinh(y)$, we can then conclude that:
$$\begin{aligned}|\cos(x+iy)|^2 &= \sin^2(x)\sinh^2(y) + \cos^2(x) + \cos^2(x)\sinh^2(y)\\ &= \cos^2(x) + 1\sinh^2(y)\end{aligned}$$

Similarly, we have that:
$$\begin{aligned}\sin(x+iy) &= \sin(x)\cos(iy) + \cos(x)\sin(iy)\\ &= \sin(x)\cosh(y) + i\cos(x)\sinh(y)\end{aligned}$$

And therefore $|\sin(x+iy)|^2 = \sin^2(x)\cosh^2(y) + \cos^2(x)\sinh^2(y)$. By again applying the fact that $\cosh^2(y) = 1 + \sinh(y)$, we can conclude that:
$$\begin{aligned}|\sin(x+iy)|^2 &= \sin^2(x) + \sin^2(x)\sinh^2(y) + \cos^2(x)\sinh^2(y)\\ &= \sin^2(x) + 1\sinh^2(y)\end{aligned}$$

At last, we now know that $|\cot(x+iy)|^2 = \frac{\cos^2(x)+\sinh^2(y)}{\sin^2(x)+\sinh^2(y)}$.

Note that $\sinh$ is an odd function satisfying that $\sinh(x) \to \infty$ as $x \to \infty$. Therefore for m large enough, we have that:
$$|\cot(\pi(x \pm mi))|^2 = \frac{\cos^2(\pi x)+\sinh^2(\pm m\pi)}{\sin^2(\pi x)+\sinh^2(\pm m\pi)} = \frac{\cos^2(\pi x)+\sinh^2(m\pi)}{\sin^2(\pi x)+\sinh^2(m\pi)} \leq \frac{\sinh^2(m\pi)+1}{\sinh^2(m\pi)-1}$$
where the latter bound goes to 1 as $m \to \infty$.

Meanwhile:
$$|\cot(\pi(\pm(m+\tfrac{1}{2})+iy))|^2 = \frac{\cos^2(\pm\pi m \pm \frac{\pi}{2})+\sinh^2(\pi y)}{\sin^2(\pm\pi m \pm \frac{\pi}{2})+\sinh^2(\pi y)} = \frac{\sinh^2(\pi y)}{1+\sinh^2(\pi y)} \leq 1$$

It follows that there is some M_1 such that for all $m \geq M_1$ we have that:
$|\cot(\pi z)| \leq 2$ on the trace of the path γ.

Also note that $\frac{2\pi}{|z^2-a^2|} \leq \frac{2\pi}{|z|^2-|a|^2} \leq \frac{2\pi}{m^2-a^2}$ for all $z \in \{\gamma\}$ Therefore:
$$\lim_{m\to\infty} \left|\int_\gamma \frac{\pi\cot(\pi z)}{z^2-a^2}\mathrm{d}z\right| \leq \lim_{m\to\infty} \int_\gamma \frac{2\pi}{m^2-a^2}\mathrm{d}|z| = \lim_{m\to\infty} \frac{2\pi\cdot(2(2m)+2(2m+1))}{m^2-a^2}$$

Note on notation:
If $\gamma : [a,b] \to \mathbb{C}$ is a piecewise C^1 path, then I'm writing $\int_\gamma g(z)\mathrm{d}|z|$ to mean $\int_a^b g(\gamma(t))|\gamma(t)|\mathrm{d}t$.

In the expression $\frac{2\pi\cdot(2(2m)+2(2m+1))}{m^2-a^2}$, the numerator depends linearly on m while the denominator depends quadratically. Therefore, the expression goes to 0. And this proves that $\lim\limits_{m\to\infty} \int_\gamma \frac{\pi\cot(\pi z)}{z^2-a^2}\mathrm{d}z = 0$.

Combining this with the work we did using the residue theorem, we have that:
$0 = \frac{\pi\cot(\pi a)}{a} - \frac{1}{a^2} - \sum_{k=1}^\infty \frac{2}{a^2-k^2}$ for all $a \notin \mathbb{Z}$.

Finally, we can rearrange to get that:

$$\pi\cot(\pi a) = \frac{1}{a} + \sum_{k=1}^{\infty}\frac{2a}{a^2-k^2}. \blacksquare$$

By applying what we just proved, we get that $h' \equiv 0$. So, we've proven that there is a constant $\alpha \in \mathbb{C} - \{0\}$ such that:

$$\sin(\pi z) = \alpha\pi z\prod_{k=1}^{\infty}(1 - \frac{z^2}{k^2}).$$

Finally, note that:

$$\lim_{z\to 0}\alpha\prod_{k=1}^{\infty}(1-\frac{z^2}{k^2}) = \lim_{z\to 0}\frac{\sin(\pi z)}{\pi z} = 1 = \prod_{k=1}^{\infty}(1-\frac{(0)^2}{k^2}) = \lim_{z\to 0}\prod_{k=1}^{\infty}(1-\frac{z^2}{k^2}).$$

Therefore, we can solve that $\alpha = 1$. And this finished the proof of $(*)$.

Remarks:

- We can plug in different values of z into $\sin(\pi z) = \pi z\prod_{k=1}^{\infty}(1-\frac{z^2}{k^2})$ to get some interesting identities. For example:

  Let $z = \frac{1}{2}$. Then:

  $$1 = \frac{\pi}{2}\cdot\prod_{k=1}^{\infty}(1-\frac{1}{4k^2}) = \frac{\pi}{2}\cdot\prod_{k=1}^{\infty}\frac{4k^2-1}{4k^2} = \frac{\pi}{2}\cdot\prod_{k=1}^{\infty}\frac{(2k-1)(2k+1)}{2k\cdot 2k}$$

  Upon rearranging we get that: $\frac{\pi}{2} = \prod_{k=1}^{\infty}\frac{2k}{2k-1}\cdot\frac{2k}{2k+1}$. Or in other words:

  $$\frac{\pi}{2} = \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdots$$

  Note that since z^{-1} is continuous on $\mathbb{C} - \{0\}$, if $\prod_{k=1}^{\infty}(1+a_k)$ absolutely converges to a nonzero value then:

  $$\begin{aligned}(\prod_{k=1}^{\infty}(1+a_k))^{-1} &= (\lim_{n\to\infty}\prod_{k=1}^{n}(1+a_k))^{-1}\\ &= \lim_{n\to\infty}(\prod_{k=1}^{n}(1+a_k))^{-1} = \lim_{n\to\infty}\prod_{k=1}^{n}(1+a_k)^{-1}\end{aligned}$$

  Also note that $(1+a_k)^{-1} = 1 - \frac{a_k}{1+a_k}$. And by a similar trick to when we invented the constant δ many pages ago (recall $(*)$), we can show using comparison test that $\sum_{k=1}^{\infty}a_k$ converges absolutely iff $\sum_{k=1}^{\infty}\frac{a_k}{1+a_k}$ converges absolutely. (Remember that no $a_k = -1$ since our original product is nonzero).

  Therefore $\prod_{k=1}^{\infty}(1+a_k)^{-1}$ converges absolutely to $(\prod_{k=1}^{\infty}(1+a_k))^{-1}$.

  Manwhile, if we let $z = i$ then $\sin(i\pi) = i\pi\prod_{k=1}^{\infty}(1+\frac{1}{k^2})$. Therefore:

  $$\prod_{k=1}^{\infty}(1+\frac{1}{k^2}) = \frac{\sinh(\pi)}{\pi} = \frac{e^{\pi}-e^{-\pi}}{2\pi}$$

- Here's the original way Euler solved the Basel problem:

  Note that $\pi z\prod_{k=1}^{\infty}(1-\frac{z^2}{k^2}) = \sin(\pi z) = \pi z - \frac{\pi^3 z^3}{6} + \cdots$. Then dividing both sides by πz gives us:

  $$\prod_{k=1}^{\infty}(1-\frac{z^2}{k^2}) = 1 - \frac{\pi^2}{6}z^2 + \cdots$$

  But note that if we tried foiling out $\prod_{k=1}^{\infty}(1-\frac{z^2}{k^2})$, we'd get that:

$$\prod_{k=1}^{\infty}(1-\frac{z^2}{k^2}) = 1-(\sum_{k=1}^{\infty}\frac{1}{k^2})z^2+\cdots$$

By matching the z^2 coefficients, we get that $\sum_{k=1}^{\infty}\frac{1}{k^2}=\frac{\pi^2}{6}$.

(I have no idea if being able to foil that infinite product can be justified.)

- We can now save a lot of effort trying to get a product expression for $\cos(\pi z)$.

  Note by the double angle identity that:

  $$\cos(\pi z) = \frac{\sin(2\pi z)}{2\sin(\pi z)} = \frac{2\pi z\prod_{k=1}^{\infty}(1-\frac{4z^2}{k^2})}{2\pi z\prod_{\ell=1}^{\infty}(1-\frac{z^2}{\ell^2})} = \frac{\prod_{k=1}^{\infty}(1-\frac{4z^2}{k^2})}{\prod_{\ell=1}^{\infty}(1-\frac{z^2}{\ell^2})}$$

  Next, as $\prod_{k=1}^{\infty}(1-\frac{4z^2}{k^2})$ converges absolutely, we can write:

  $$\begin{aligned}\prod_{k=1}^{\infty}(1-\frac{4z^2}{k^2}) &= \prod_{\ell=1}^{\infty}(1-\frac{4z^2}{(2\ell-1)^2})\cdot\prod_{\ell=1}^{\infty}(1-\frac{4z^2}{(2\ell)^2})\\ &= \prod_{\ell=1}^{\infty}(1-\frac{4z^2}{(2\ell-1)^2})\cdot\prod_{\ell=1}^{\infty}(1-\frac{z^2}{\ell^2})\end{aligned}$$

  It's incredibly late and I don't want to spend time proving you can split the product up like how I just did. The general idea of the proof is that we already know we can slit up the sum of $\sum_{k=N}^{\infty}\mathrm{Log}(1-\frac{4z^2}{k^2})$. Then, we just need to exponentiate that.

  As a result, we get that $\cos(\pi z)=\prod_{\ell=1}^{\infty}(1-\frac{4z^2}{(2\ell-1)^2})$.

And now I am fully caught up to Math 220b. God I hope my left hand heels fast cause I'm so far behind due to my stupid injury.

Math 200b Lecture Notes:

Here are two more important examples of A-modules:

- If $\{M_i\}_{i\in I}$ is a family of A-modules, then we define their external direct sum to be:

  $\oplus_{i\in I}M_i := \{(x_i)_{i\in I} : x_i\in M_i$ for all i and $x_i=0$ except for finitely many $i\}$.

  To turn $\oplus_{i\in I}M_i$ into an A-module, we interpret all the elements of $\oplus_{i\in I}M_i$ as finite formal sums of the elements of the M_i. Equivalently, note that $\oplus_{i\in I}M_i$ is a submodule of $\prod_{i\in I}M_i$

- Suppose M and N are two A-modules.

  We say $\phi : M\to N$ is an A-module homomorphism if $\phi(x+x')=\phi(x)+\phi(x')$ for all $x,x'\in M$ and $\phi(ax)=a\phi(x)$ for all $a\in A$ and $x\in M$. Equivalently, we say that ϕ is an A-linear map.

Next, we denote $\mathrm{Hom}_A(M,N)$ as the set of all A-modules homomorphisms from M to N. If A is commutative, we can turn $\mathrm{Hom}_A(M,N)$ into an A-module by setting:

$$(\phi_1+\phi_2)(x) \coloneqq \phi_1(x)+\phi_2(x) \text{ and } (a\cdot\phi)(x) \coloneqq \phi(ax).$$

If M,N are merely left A-modules, then $\mathrm{Hom}_A(M,N)$ would be a right A-module with $(\phi\cdot a)(x) \coloneqq \phi(ax)$ Similiarly, if M,N are merely right A-modules, then $\mathrm{Hom}_A(M,N)$ would be a left A-module with $(a\cdot\phi)(x) \coloneqq \phi(xa)$.

Short Lemma: If M,N are A-modules and $\phi\in\mathrm{Hom}_A(M,N)$ is bijective, then $\phi^{-1}\in\mathrm{Hom}_A(N,M)$.

Proof:
Let $x_1,x_2\in N$ and suppose $\phi^{-1}(x_i)=y_i$ for both i. Then for any $a,b\in A$ we have that:

$$\phi(ay_1+by_2)=a\phi(y_1)+b\phi(y_2)=ax_1+bx_2.$$

In turn, $\phi^{-1}(ax_1+bx_2)=ay_1+by_2=a\phi^{-1}(x_1)+b\phi^{-1}(x_2)$. ■

Let $\{M_i\}_{i\in I}$ be a family of A-modules. Then we define:

- $j_i : M_i \to \prod\limits_{i'\in I} M_{i'}$ by $((j_i(x))_{i'})_{i'\in I} = \begin{cases} x & \text{if } i'=i \\ 0 & \text{if } i'\neq i\end{cases}$
- $P_i : \prod\limits_{i'\in I} \to M_i$ by $P_i((x_{i'})_{i'\in I}) = x_i$.

It's easy to check j_i and P_i are A-module homomorphism. Also, by restricting the codomain of j_i or the domain of P_i, we can obtain instead view j_i as mapping into $\oplus_{i'\in I}M_{i'}$ and P_i as mapping from $\oplus_{i'\in I}M_{i'}$.

Math 200b Set 2 Problem 3: Prove that:

(a) $\mathrm{Hom}_A(\oplus_{i\in I}M_i,N)\cong\prod_{i\in I}\mathrm{Hom}_A(M_i,N)$

(b) $\mathrm{Hom}_A(N,\prod_{i\in I}M_i)\cong\prod_{i\in I}\mathrm{Hom}_A(N,M_i)$

as A-modules.

Proof of (b):
Suppose $(\phi_i)_{i\in I}\in\prod_{i\in I}\mathrm{Hom}_A(N,M_i)$. Then define:

$$\Phi : N\to\prod_{i\in I}M_i \text{ by } \Phi(x)=(\phi_i(x))_{i\in I}.$$

Clearly $\Phi\in\mathrm{Hom}_A(N,\prod_{i\in I}M_i)$ since for all $a,b\in A$ and $x_1,x_2\in N$:

$$\begin{aligned}\Phi(ax_1+bx_2) &= (\phi_i(ax_1+bx_2))_{i\in I}\\ &= (a\phi_i(x_1)+b\phi_i(x_2))_{i\in I} = a(\phi_i(x_1))_{i\in I}+b(\phi_i(x_2))_{i\in I} = a\Phi(x_1)+b\Psi(x_2)\end{aligned}$$

It follows that the map $(\phi_i)_{i\in I}\mapsto\Phi$ from $\prod_{i\in I}\mathrm{Hom}_A(N,M_i)$ to $\mathrm{Hom}_A(N,\prod_{i\in I}M_i)$ is well-defined.

We also show this map is invertible. After all, note that $(\phi_i)_{i\in I}=(P_i\circ\Phi)_{i\in I}$ holds for the homomorphisms we were working with before. Also suppose we are given an A-module homomorphism $\Psi : N\to\prod_{i\in I}M_i$. Then after defining $\psi_i \coloneqq P_i\circ\Psi$, we have that $\psi_i\in\mathrm{Hom}_A(N,M_i)$ and that our original mapping sends $(\psi_i)_{i\in I}$ to Ψ.

Finally, we need to show that $(\phi_i)_{i\in I} \mapsto \Phi$ is an A-linear map.

Let $(\phi_i)_{i\in I}$ and $(\psi_i)_{i\in I}$ be elements of $\prod_{i\in I} \mathrm{Hom}_A(N, M_i)$. Then set:

$$\Phi(x) := (\phi_i(x))_{i\in I} \text{ and } \Psi(x) = (\psi_i(x))_{i\in I}.$$

Now for any $a, b \in A$ we have that:

$$\begin{aligned}((a \cdot \Phi) + (b \cdot \Psi))(x) &= \Phi(ax) + \Psi(bx) \\ &= (\phi_i(ax))_{i\in I} + (\psi_i(bx))_{i\in I} \\ &= ((a \cdot \phi_i)(x))_{i\in I} + ((b \cdot \psi_i)(x))_{i\in I} \\ &= \left(((a \cdot \phi_i) + (b \cdot \psi_i))(x)\right)_{i\in I}\end{aligned}$$

Proof of (a):

For all $\Phi \in \mathrm{Hom}_A(\oplus_{i\in I} M_i, N)$ and $i \in I$ define $\phi_i := \Phi \circ j_i$ for all i. Note that ϕ_i being the composition of two A-module homomorphism is also an A-module homomorphism. Next consider the mapping $\Phi \mapsto (\phi_i)_{i\in I} \in \prod_{i\in I} \mathrm{Hom}_A(M_i, N)$.

To show that that mapping is invertible, first note that:

$$\Phi((x_i)_{i\in I}) = \sum_{i\in I} x_i = \sum_{i\in I} \phi_i((x_i)_{i\in I}) \text{ for all } (x_i)_{i\in I} \in \oplus_{i\in I} M_i.$$

Meanwhile, suppose $(\psi_i)_{i\in I} \prod_{i\in I} \mathrm{Hom}_A(M_i, N)$ and define $\Psi : \oplus_{i\in I} M_i \to N$ by $\Psi((x_i)_{i\in I}) := \sum_{i\in I} \psi_i(x_i)$. I claim that Ψ is an A-module homomorphism. After all, suppose $a, b \in A$ and $(x_i)_{i\in I}, (y_i)_{i\in I} \in \oplus_{i\in I} M_i$. Then:

$$\begin{aligned}\Psi(a(x_i)_{i\in I} + b(y_i)_{i\in I}) &= \Psi((ax_i + by_i)_{i\in I}) \\ &= \sum_{i\in I} \psi_i(ax_i + by_i) \\ &= \sum_{i\in I} (a\psi_i(x_i) + b\psi_i(y_i)) \\ &= a \sum_{i\in I} \psi_i(x_i) + b \sum_{i\in I} \psi_i(y_i) = a\Psi((x_i)_{i\in I}) + b\Psi((y_i)_{i\in I})\end{aligned}$$

By the way, note that all these sums are well-defined because only finitely many of the terms being summed over are nonzero.

Furthermore, note that $\Psi \circ j_i = \psi_i$ for all i. Hence, our original mapping sends Ψ back to $(\psi_i)_{i\in I}$. And with that we've proven there is an inverse function to our original mapping.

Finally, we need to show that $\Phi \mapsto (\phi_i)_{i\in I}$ is an A-linear map.

Suppose $\Phi, \Psi \in \mathrm{Hom}_A(\oplus_{i\in I} M_i, N)$ and let $a, b \in A$. Then for any $i \in I$ and $x \in M_i$, we have that:

$$\begin{aligned}((a \cdot \Phi) + (b \cdot \Psi))(j_i(x)) &= \Phi(aj_i(x)) + \Psi(bj_i(x)) \\ &= \Phi(j_i(ax)) + \Psi(j_i(bx)) = (a \cdot (\Phi \circ j_i) + b \cdot (\Psi \circ j_i))(x)\end{aligned}$$

In other words $((a \cdot \Phi) + (b \cdot \Psi)) \circ j_i = (a \cdot (\Phi \circ j_i)) + (b \cdot (\Psi \circ j_i))$ for all i. This proves that each component of the map $\Phi \mapsto (\phi_i)_{i\in I}$ is A-linear. So, the entire map is A-linear. ■

Note that $\oplus_{i \in I} M_i$ satisfying identity (a) from the prior result is equivalent to saying that $\oplus_{i \in I} M_i$ is the category theoretic coproduct of the M_i in the category of A-modules.

Similarly, $\prod_{i \in I} M_i$ satisfying identity (b) from the prior result is equivalent to saying that $\prod_{i \in I} M_i$ is the category theoretic product of the M_i in the category of A-modules.

Suppose $\phi : M \to N$ is an A-module homomorphism. Then we set
$$\ker(\phi) \coloneqq \{x \in M : \phi(x) = 0\} \text{ and } \operatorname{im}(\phi) \coloneqq \{\phi(x) : x \in M\}.$$

You can easily show that $\ker(\phi)$ is a submodule of M and $\operatorname{im}(\phi)$ is a submodule of N.

As a side note: $0 \cdot x = (0 + 0) \cdot x = (0 \cdot x) + (0 \cdot x) \Longrightarrow 0 = (0 \cdot x)$.

Suppose $N \subseteq M$ is a submodule. Since $(M, +)$ is an abelian group, we know that $(N, +) \lhd (M, +)$. In turn, $(M/N, +)$ is a well-defined abelian quotient group. We can also turn M/N into an A-module by defining $a \cdot (x + N) \coloneqq ax + N$ for all $a \in A$ and $x + N \in M/N$.

To see that this is well defined, note that if $x + N = x' + N$, then $x - x' \in N$. In turn $a(x - x') \in N$ and so $ax + N = ax' + N$.

Hopefully now all the algebraic properties of scalar multiplication are obvious.

The 1st. Isomorphism Theorem For A-modules: Suppose $\phi : M \to N$ is an A-module homomorphism. Then we have the following commuting diagram.

$$\begin{tikzcd}
x \arrow[rrr, mapsto, bend left] \arrow[ddd, mapsto] & & & \phi(x) \\
& M \arrow[r, "\phi"] \arrow[d, two heads, "\pi"'] & N & \\
& M/\ker(\phi) \arrow[r, "\sim", "\overline{\phi}"'] & \operatorname{im}(\phi) \arrow[u, hook] & \\
x+\ker(\phi) \arrow[rrr, mapsto, bend right] & & & \overline{\phi}(x+\ker(\phi))=\phi(x) \arrow[uuu, mapsto]
\end{tikzcd}$$

Proof:
By the group theory 1st isomorphism theorem, we know that $\overline{\phi}$ is a well-defined bijective group homomorphism. Also, for all $a \in A$ we have that:
$$\overline{\phi}(a \cdot (x + \ker(\phi))) = \overline{\phi}(ax + \ker(\phi)) = \phi(ax) = a\phi(x) = a\overline{\phi}(x + \ker(\phi))$$

Thus $\overline{\phi}$ is an A-module homomorphism. ■

Suppose X is an arbitrary set, and then for all $x \in X$ define $Ax \coloneqq \{ax : a \in A\}$ where ax is a formal expression. We can turn Ax into an obvious A-module by defining $ax + a'x = (a+a')x$ and $a \cdot (a'x) = (aa')x$.

We claim $\mathcal{F}(X) \coloneqq \oplus_{x \in x} Ax$ is the free A-module of X.

Proof:

Let $i : X \hookrightarrow \mathcal{F}(X)$ be given by $x \mapsto j_x(x)$ (recall the definition of j three pages ago). Now suppose we are given some function $f : X \to N$ where N is an A-module. Next, define $\phi_x : Ax \to N$ by $\phi_x(ax) \coloneqq af(x)$ for all $a \in A$ and $x \in X$. Clearly ϕ_x is the unique A-module homomorphism from Ax to N satifsying that $\phi_x(x) = f(x)$.

But now by the homework problem I did prior, there is a unique A-module homomorphism $\widehat{f} \in \mathrm{Hom}_A(\oplus_{x \in X} Ax, N)$ such that $\widehat{f}(j_x(ax)) = \phi_x(ax) = af(x)$ for all $a \in A$ and $x \in X$.

Thus, the following diagram holds:

$$\begin{tikzcd} & \mathcal{F}(X) \arrow[d, dashed, "\exists! \widehat{f}"] \\ X \arrow[ur, hook, "i"] \arrow[r, "f"'] & N \end{tikzcd}$$

In particular, note that when $|X| = n$, then $\mathcal{F}(X) \cong A^n$.

I'm struggling to focus on the notes so I'm going to do a homework problem and then read a book for a different class.

Math 200b Set 2 Problem 1: Prove that the following polynomials are irreducible:

(a) $f(x) \coloneqq x^{p-1} + x^{p-2} + \cdots + 1$ in $\mathbb{Q}[x]$ where p is a prime number;

To start off, define $\overline{f}(x) \coloneqq f(x+1)$. Then note that $\overline{f}$ is not irreducible iff f is not irreducible. After all, note for any ring A and polynomial $h(x) \in A[x]$ that:

$$\deg(h(x+a)) = \deg(h(x)).$$

This is a because of the fact that $a_k(x+a)^k = a_k x^k + (\text{ lower degree terms })$ for all k.

As a result, we can see that $f(x) = g_1(x)g_2(x)$ where $\deg(g_1), \deg(g_2) > 0$ if and only if $f(x+1) = g_1(x+1)g_2(x+1)$ where $\deg(g_1(x+1)), \deg(g_2(x+2)) > 0$.

With that, we shall now prove that f is irreducible by proving that $\overline{f}$ is irreducible. Observe that:

$$\begin{aligned} \overline{f}(x) = f(x+1) = \tfrac{(x+1)^p - 1}{(x+1)-1} = \tfrac{(x+1)^p - 1}{x} &= \textstyle\sum_{k=1}^{p} \binom{p}{k} x^{k-1} \\ &= x^{p-1} + \binom{p}{p-1} x^{p-2} + \cdots + \binom{p}{1} \end{aligned}$$

Claim: $p \mid \binom{p}{r}$ if p is prime and $0 < r < p$.

> This is because $r!\binom{p}{r} = p(p-1)\cdots(p-r+1)$. So $p \mid r!\binom{p}{r}$. That said, we also know that $p \not| \, r!$ since p is coprime to all k such that $0 < k \leq r < p$. So, we must have that $p \mid \binom{p}{r}$.

At the same time though, note that $\binom{p}{1} = p$. So p^2 doesn't divide the constant term of $\overline{f}$. It now follows by Eisenstein's criterion that $\overline{f}$ is irreducible.

(b) $g(x,y) := x^{p-1} + q_2(y)x^{p-2} + \cdots + q_{p-1}(y) \in \mathbb{Q}[x,y]$ where p is prime and $q_i(y) \in \mathbb{Q}[y]$ with $q_i(1) = 1$ for all i;

To start off, note that $\mathbb{Q}[y]$ is a P.I.D. and $y-1$ is irreducible in $\mathbb{Q}[y]$. It follows that $\langle y-1 \rangle$ is maximal among principal ideals and thus is itself a maximal ideal in $\mathbb{Q}[y]$.

Next, suppose there exists polynomials $g_1, g_2 \in (\mathbb{Q}[y])[x]$ such that $g(x,y) = g_1(x,y)g_2(x,y)$ and the (x-)degree of g_1 and g_2 are both less than $p-1$. By applying the modulo map $\pi : (\mathbb{Q}[y])[x] \to (\frac{\mathbb{Q}[y]}{\langle y-1 \rangle})[x]$ (and noting that $\frac{\mathbb{Q}[y]}{\langle y-1 \rangle} \cong \mathbb{Q}$), we get that:

$$x^{p-1} + x^{p-2} + \cdots + 1 = x^{p-1} + q_2(1)x^{p-2} + \cdots + q_{p-1}(1) = \widetilde{g}_1(x)\widetilde{g}_2(x)$$

where $\widetilde{g}_i(x) = \pi(g_i)$ is a polynomial in $\mathbb{Q}[x]$ of degree less than $p-1$ for each i. By part (a) though, this is a contradiction.

> As a side note, if $f(x) \in A[x]$ and $a \in A$ for some ring A, then by the long division theorem there exists $q(x) \in A[x]$ and $\alpha \in A$ such that $f(x) = (x-a)q(x) + \alpha$. And by evaluating at $x = a$ we then get that $f(a) = \alpha$. This is why modding $\mathbb{Q}[x]$ by $\langle x - a \rangle$ is equivalent to evaluating a polynomial in $\mathbb{Q}[x]$ at a.

Hence, we've now shown if $g(x,y) = g_1(x,y)g_2(x,y)$ then the (x-)degree of one of the g_i must be $p-1$ and the (x-)degree of the other is 0. In other words, without loss of generality there exists polynomials $r_0(y), \ldots, r_{p-1}(y), s(y) \in \mathbb{Q}[y]$ such that:

$$g_1(x,y) = r_1(y)x^{p-1} + \ldots + r_{p-1}(y) \text{ and } g_2(x,y) = s(y).$$

But now as $g = g_1g_2$, we can compare coefficients to get that $r_1(y)s(y) = 1$. Since $\mathbb{Q}[y]$ is a domain, this implies that $s(y)$ is a constant polynomial. Or in other words, g_2 is just some constant. This proves that $g(x,y)$ is irreducible in $\mathbb{Q}[x,y]$.

(c) $k(x,y) := x^n - y$ in $F[x,y]$ where F is a field;

Let $D = F[y]$ and denote the field of fractions of D as $Q(D)$. Importantly, D is a U.F.D. Also, y is irreducible in D, and if we view:

$$k(x,y) = 1x^n + 0x^{n-1} + \cdots + 0x - y \text{ as a polynomial in } D[x],$$

then $y \not| \, 1$, $y^2 \not| \, -y$, and y does divide all the nonleading coefficients.

By Eisenstein's criterion, we can conclude that $k(x,y)$ is irreducible in $(Q(D))[x]$. Yet also note that as the leading term of $k(x,y)$ is a unit in D, we have that $k(x,y)$ is a primitive polynomial in $(Q(D))[x]$. So, we can further conclude that $k(x,y)$ is irreducible in $D[x] = (F[y])[x]$.

Finally, as F and in turn $F[y]$ is a domain, we have that $(F[y])[x]^\times = F[y]^\times = F$. Therefore, if $k(x,y) = k_1(x,y)k_2(x,y)$ then one of the k_i must be a constant in $F^\times$ since $k(x,y)$ is irreducible in $(F[y])[x]$. Obviously $F^\times \subseteq F[x,y]^\times$ (and I'm too tired to prove the other inclusion). So, we've shown that $k(x,y)$ is irreducible in $F[x,y]$ as well.

(d) $p(x,y) := x^2 + y^2 - 2$ in $F[x,y]$ where F is a field and its characteristic is not 2;

Again consider p as a polynomial in $(F[y])[x]$. Importantly, $F[y]$ is a U.F.D. Also, we claim that there is an irreducible factor $g(y) \in F[y]$ such that $g^2 \not| \, y^2 - 2$.

If $y^2 - 2$ is irreducible in $F[y]$, then we are done. Otherwise, since $F[y]$ is a U.F.D. we know there are irreducible polynomials $y - a$ and $y - b$ in $F[y]$ such that $(y-a)(y-b) = y^2 - 2$. We need to show that $a \neq b$. That way $(y-a)^2 \not| \, y^2 - 2$.

Suppose $(y-a)^2 = y^2 - 2$. Then $y^2 - 2ay + a^2 = y^2 - 2$. By comparing terms we get that $-2a = 0$. Also, as F is an integral domain and $-2 \neq 0$ (since F has characteristic $\neq 2$), we must have that $a = 0$. But now we also have that $a^2 = 0^2 = -2$. This is a contradiction.

But now by applying Eisenstein's criterion using that irreducible factor $g(y)$ we showed exists, we have that $p(x,y)$ is irreducible in $(Q(F[y]))[x]$ (where $Q(F[x])$ is the field of fractions of $F[x]$). the rest of the proof of the irreducibility of $p(x,y)$ follows as in part (c).

I need to take a break from Algebra to do my complex analysis homework. I will return to this later.

1/22/2026

Math 220b Homework 2:

Set 2 Problem 1: We define the Riemann Zeta function $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ (here $n^s := \exp(s\,\mathrm{Log}(n))$).

(i) Show that $\zeta(s)$ is a well-defined holomorphic function in s in the region $\{\mathrm{Re}(s) > 1\}$.

Let $f_m(s) = \sum_{n=1}^{m} \frac{1}{n^s}$ for all $m \in \mathbb{N}$. Then $f_m \in O(\mathbb{C})$ for all $m \in \mathbb{N}$. Our first claim is that $\zeta(s) = \lim_{m\to\infty} f_m(s)$ exists if $\mathrm{Re}(s) > 1$.

Note that $|n^s| = |e^{s\,\mathrm{Log}(n)}| = e^{\mathrm{Re}(s)\log(n)} = n^{\mathrm{Re}(s)}$. Therefore, $\sum_{n=1}^{\infty} \frac{1}{n^s}$ converges absolutely iff $\sum_{n=1}^{\infty} \frac{1}{n^{\mathrm{Re}(s)}}$ converges. But from a first course in real analysis, we know that $\sum_{n=1}^{\infty} \frac{1}{n^{\mathrm{Re}(s)}}$ converges iff $\mathrm{Re}(s) > 1$.

The other claim we make is that f_m converges to ζ locally uniformly as $m \to \infty$.

Let K be any compact set in $\{s \in \mathbb{C} : \mathrm{Re}(s) > 1\}$. Then as $z \mapsto \mathrm{Re}(z)$ is a continuous function, we can conclude by the extreme value theorem that there is some $\alpha \in \mathbb{R}$ such that $1 < \alpha = \min_{s\in K}(\mathrm{Re}(s))$.

But now for all $n \in \mathbb{N}$ and $s \in K$, we have that $\left|\frac{1}{n^s}\right| = \frac{1}{n^{\operatorname{Re}(s)}} \leq \frac{1}{n^\alpha}$ and $\sum_{n=1}^{\infty} \frac{1}{n^\alpha} < \infty$. By the Weierstrass M-test we can conclude that $f_m \to \zeta$ uniformly on K.

Now it's a simple corollary of the Weierstraß convergence theorem that ζ is holomorphic.

(ii) Let $p_1 < p_2 < \cdots$ be the sequence of positive prime integers. Show that the infinite product $\prod_{n=1}^{\infty}(1 - \frac{1}{p_n^s})$ converges absolutely and locally uniformly for $\operatorname{Re}(s) > 1$ to $\frac{1}{\zeta(s)}$.

We can easily conclude by the comparison test with the series $\sum_{n=1}^{\infty} \frac{1}{n^{\operatorname{Re}(s)}}$ that:

$$\sum_{n=1}^{\infty} \left|\frac{-1}{p_n^s}\right| = \sum_{n=1}^{\infty} \frac{1}{p_n^{\operatorname{Re}(s)}} \text{ converges if } \operatorname{Re}(s) > 1.$$

Going a step further, if K is any compact set in $\{s \in \mathbb{C} : \operatorname{Re}(s) > 1\}$ then we can show via identical reasoning as in part (i) that $\sum_{n=1}^{\infty} \frac{1}{p_n^s}$ converges uniformly on K. Therefore, we can conclude that $\sum_{n=1}^{\infty} \frac{1}{p_n^s}$ converges absolutely locally uniformly.

It follows that $\prod_{n=1}^{\infty}(1 - \frac{1}{p_n^s})$ also converges absolutely and locally uniformly on $\{s \in \mathbb{C} : \operatorname{Re}(s) > 1\}$.

Next note for all $n \in \mathbb{N}$ and $s \in \mathbb{C}$ with $\operatorname{Re}(s) > 1$, we have that:

$$|p_n^{-s}| = p_n^{\operatorname{Re}(-s)} < p_n^{-1} \leq \tfrac{1}{2} < 1.$$

As a result:

$$\left(\prod_{n=1}^{\infty}(1 - \tfrac{1}{p_n^s})\right)^{-1} = \prod_{n=1}^{\infty} \frac{1}{1 - p_n^{-s}} = \prod_{n=1}^{\infty}\left(\sum_{k=0}^{\infty}(p_n^{-s})^k\right) = \prod_{n=1}^{\infty}\left(\sum_{k=0}^{\infty}(\tfrac{1}{p_n^k})^s\right)$$

where $\sum_{k=0}^{\infty}(\frac{1}{p_n^k})^s$ converges absolutely for each n.

By considering successive Cauchy products of the series (i.e. foiling them) and applying Merten's theorem (see my math 140a notes), we can conclude for any $N \in \mathbb{N}$ that:

$$\prod_{n=1}^{N}\left(\sum_{k=0}^{\infty}(\tfrac{1}{p_n^k})^s\right) = \sum_{k \in S^{(N)}} \frac{1}{k^s}$$

where $S^{(N)}$ is the set of positive integers whose only prime factors are among $\{p_1, \ldots, p_N\}$.

Finally, as the series $\sum_{k=1}^{\infty} \frac{1}{k^s}$ converges absolutely and $\{S^(n)\}_{n \in \mathbb{N}}$ is an increasing sequence of sets whose union is all of $\mathbb{N}$, we have that:

$$\sum_{k \in S^{(N)}} \frac{1}{k^s} \to \sum_{k=1}^{\infty} \frac{1}{k^s} = \zeta(s) \text{ as } N \to \infty.$$

All of this proves that $\left(\prod_{n=1}^{\infty}(1 - \frac{1}{p_n^s})\right)^{-1} = \zeta(s)$. Or in other words:

$$\prod_{n=1}^{\infty}(1 - \tfrac{1}{p_n^s}) = \frac{1}{\zeta(s)}. \blacksquare$$

Set 2 Problem 3: This question introduces the entire function G, which is closely related to the Γ-function.

(i) Possibly using Taylor expansion, show that if $|w| \leq \frac{1}{2}$ then $|\operatorname{Log}(1+w) - w| \leq 2|w|^2$.

Note that $\operatorname{Log}(1+w) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} w^k$. In turn:

$$\operatorname{Log}(1+w) - w = w^2 \sum_{k=2}^{\infty} \frac{(-1)^{k+1}}{k} w^{k-2}.$$

And finally, we get when $|w| < 1$ that:

$$\begin{aligned} |\operatorname{Log}(1+w) - w| &= |w|^2 \cdot |\tfrac{-1}{2} + \tfrac{1}{3}w - \tfrac{1}{4}w^2 + \tfrac{1}{5}w^3 - \cdots| \\ &\leq |w|^2 \cdot (\tfrac{1}{2} + \tfrac{1}{3}|w| + \tfrac{1}{4}|w|^2 + \tfrac{1}{5}|w|^3 + \cdots) \\ &\leq |w|^2 \cdot (1 + |w| + |w|^2 + |w|^3 + \cdots) = |w|^2 \cdot \tfrac{1}{1-|w|} \end{aligned}$$

In particular, when $|w| \leq \frac{1}{2}$ then $\frac{1}{1-|w|} \leq 2$. So, if $|w| \leq \frac{1}{2}$ then:

$$|\operatorname{Log}(1+w) - w| \leq 2|w|^2.$$

(ii) Let $a, b \in \mathbb{C}$ have positive real parts. Show that $\operatorname{Log}(ab) = \operatorname{Log}(a) + \operatorname{Log}(b)$.

Write $a = r_1 e^{i\theta_1}$ and $b = r_2 e^{i\theta_2}$ where $\theta_1, \theta_2 \in (\frac{-\pi}{2}, \frac{\pi}{2})$. Then $ab = r_1 r_2 e^{i(\theta_1+\theta_2)}$ and importantly $\theta_1 + \theta_2 \in (-\pi, \pi)$. It follows that:

$$\begin{aligned} \operatorname{Log}(ab) &= \log(r_1 r_2) + i(\theta_1 + \theta_2) \\ &= (\log(r_1) + i\theta_1) + (\log(r_2) + i\theta_2) = \operatorname{Log}(a) + \operatorname{Log}(b). \end{aligned}$$

(iii) Let $r > 0$. Show that there exists N such that for all $n \geq N$ and $z \in \Delta(0, r)$:

$1 + \frac{z}{n}$ and $e^{-\frac{z}{n}}$ have positive real parts.

Since $\operatorname{Re}(1+z)$ and $\operatorname{Re}(e^{-z})$ are continuous functions that evaluate to 1 at $z = 0$, we know there exists $\delta > 0$ such that $|\operatorname{Re}(1+w) - 1| < 1$ and $|\operatorname{Re}(e^{-z}) - 1| < 1$ whenever $|w| < \delta$. Next, we set N large enough so that $\frac{r}{N} < \delta$.

Now for all $n \geq N$ and $w \in \Delta(0, r)$, we have that $|\frac{z}{n}| < \frac{r}{n} < \delta$. In turn, $|\operatorname{Re}(1 + \frac{z}{n}) - 1| < 1$ and $|\operatorname{Re}(e^{-\frac{z}{n}}) - 1| < 1$, This proves what we wanted.

(iv) (Taking Logs and arguing that the series of Logs converges absolutely and locally uniformly), show that the product $G(z) := \prod_{n=1}^{\infty} ((1 + \frac{z}{n}) e^{-\frac{z}{n}})$ converges to an entire function.

Suppose $K \subseteq \mathbb{C}$ is compact. Then we know there is some $r > 0$ such that $K \subseteq \Delta(0, r)$, and by part (iii) we know there exists N such that $\operatorname{Re}(1 + \frac{z}{n}) > 0$ and $\operatorname{Re}(e^{-\frac{z}{n}}) > 0$ for all $z \in \Delta(0, r)$ and $n \geq N$. Also as we'll see soon, it will be convenient to make N small enough so that $\frac{r}{N} \leq \frac{1}{2}$.

By part (ii) we have that:

$$\begin{aligned} \sum_{n=N}^{\infty} \operatorname{Log}((1 + \tfrac{z}{n}) e^{-\frac{z}{n}}) &= \sum_{n=N}^{\infty} \left(\operatorname{Log}(1 + \tfrac{z}{n}) + \operatorname{Log}(e^{-\frac{z}{n}})\right) \\ &= \sum_{n=N}^{\infty} \left(\operatorname{Log}(1 + \tfrac{z}{n}) - \tfrac{z}{n}\right) \end{aligned}$$

In turn, by the Weierstraß M-test we have that $\sum_{n=N}^{\infty} \operatorname{Log}((1 + \frac{z}{n}) e^{-\frac{z}{n}})$ converges absolutely and uniformly on K. After all, for all $z \in \Delta(0, r)$ we have by part (i) that:

$$\sum_{n=N}^{\infty} |\operatorname{Log}(1+\tfrac{z}{n}) - \tfrac{z}{n}| \leq \sum_{n=N}^{\infty} 2|\tfrac{z}{n}|^2 = 2r^2 \sum_{n=N}^{\infty} \tfrac{1}{n^2} < \infty$$

This proves that $G(z)$ congerges absolutely on K. Furthermore, as

$$\lim_{m\to\infty} \prod_{n=1}^{m} ((1+\tfrac{z}{n})e^{-\frac{z}{n}}) = \lim_{m\to\infty} \left(\exp(\sum_{n=N}^{m} \operatorname{Log}((1+\tfrac{z}{n})e^{-\frac{z}{n}})) \prod_{n=1}^{N-1} ((1+\tfrac{z}{n})e^{-\frac{z}{n}})\right),$$

we can conclude that $\lim_{m\to\infty} \prod\limits_{n=1}^{m} ((1+\frac{z}{n})e^{-\frac{z}{n}})$ converges uniformly on K.

Now it is a simple consequence of the Weierstraß convergence theorem that G is entire.

What are the zeros of G?

$G(z) = 0$ iff $(1+\frac{z}{n})e^{-\frac{z}{n}} = 0$ for some $n \in \mathbb{N}$. But since exponentials are never zero, we in turn know that $G(z) = 0$ iff $(1+\frac{z}{n}) = 0$ for some $n \in \mathbb{N}$. So the zeros of G are precisely the negative integers. ■

This is not all of the problems on the problem set. But unfortunately I need to turn my attention back to algebra because I'm even more behind in that class.

Math 200b Notes:

If M is an A-module and $\{M_i\}_{i\in I}$ is a collection of submodules of M satisfying that:

$$\textstyle\sum_{i\in I} x_i = 0 \implies \text{ all } x_i \text{ are zero for all } (x_i)_{i\in I} \in \oplus_{i\in I} M_i,$$

then we say the submodule generated by $\bigcup_{i\in I} M_i$ forms an internal direct sum.

Lemma: The smallest submodule containing $\bigcup_{i\in I} M_i$ is $\{\sum_{i\in I} x_i : (x_i)_{i\in I} \in \oplus_{i\in I} M_i\}$. Sometimes, we'll denote this submodule as $\sum_{i\in I} M_i$.

Hopefully this is obvious.

Proposition: The following are equivalent:

1. $\sum_{i\in I} M_i$ forms an internal direct sum.
2. If $\sum_{i\in I} x_i = \sum_{i\in I} x_i'$ (where all but finitely many x_i and x_i' are zero), then $x_i = x_i'$ for all i.
3. $f : \oplus_{i\in I} M_i \to \sum_{i\in I}$ given by $(x_i)_{i\in I} \mapsto \sum_{i\in I} x_i$ is an isomorphism.
4. $\forall i \in I$, $M_i \cap (\sum_{\substack{i'\in I\\ i'\neq i}} M_{i'}) = \{0\}$.

$(1. \implies 2.)$
$\sum_{i\in I} x_i - \sum_{i\in I} x_i' = \sum_{i\in I}(x_i - x_i') = 0 \implies \forall i \in I, x_i - x_i' = 0 \implies \forall i \in I, x_i = x_i'$.

$(2. \implies 3.)$
Note that f is an A-module homomorphism as:

- $$\begin{aligned} f((x_i)_{i\in I} + (x_i')_{i\in I}) &= f((x_i + x_i')_{i\in I}) \\ &= \textstyle\sum_{i\in I}(x_i + x_i') \end{aligned}$$

$$= \sum_{i\in I} x_i + \sum_{i\in I} x_i' = f((x_i)_{i\in I}) + f((x_i')_{i\in I})$$

- $f(a(x_i)_{i\in I}) = f((ax_i)_{i\in I}) = \sum_{i\in I} ax_i = a\sum_{i\in I} x_i = af((x_i)_{i\in I})$.

By the definition of $\sum_{i\in I} M_i$, it's clear f is surjective. Also, (2.) implies f is injective.

$(3. \implies 1.)$
$\sum x_i = 0 \implies f((x_i)_{i\in I}) = 0$ In turn, since f is injective we have that $(x_i)_{i\in I} = 0$. So, $x_i = 0$ for all i.

$(1. \implies 4.)$
$x_i = \sum_{i'\in I-\{i\}} x_{i'} \implies -x_i + \sum_{i'\in I-\{i\}} x_{i'} = 0 \implies x_i = 0$ and $x_{i'} = 0$ for all i'.

$(4. \implies 1.)$
Suppose $\sum_{i\in I} x_i = 0$ and some $x_{i'} \neq 0$. Then that would imply that:

$0 \neq -x_i \in M_{i'} \cap (\sum_{\substack{i\in I \\ i\neq i'}} M_i)$, contradicting (4.). ■

As a result of the isomorphism given in the last proposition, internal direct sums are often denoted with the $\oplus$ symbol just like external direct sums are. But when we need to distinguish between an interal direct sum and an external direct sum, we shall use the $\sum$ symbol for the interal sum.

Using Zorn's lemma, if V is a vector space over F then V has an F-basis $\mathcal{B}$ (meaning the elements of $\mathcal{B}$ are F-linearly independent and $V = \langle \mathcal{B} \rangle$ as an F-vector space).

(Note that $\langle \mathcal{B} \rangle = \mathrm{span}(\mathcal{B})$. Also as a reminder, the elements of $\mathcal{B}$ are F-linearly independent if $a_1m_1 + \ldots + a_nm_n = 0 \implies a_1, \ldots, a_n = 0$ for any finite linear combination of elements in $\mathcal{B}$. Hopefully it's clear how we define A-linear independence if A is a general ring instead of a field.)

If M is an A-module, we define:

- $\mathrm{rank}(M) \coloneqq \max\{n : \exists m_1, \ldots, m_n\ s.t.$ all m_i are linearly independent$\}$.
- $d(M) \coloneqq \min\{n \in \mathbb{N} : M = \langle m_1, \ldots, m_n \rangle$ for some $m_1, \ldots, m_n\}$

Note: We are allowing $\mathrm{rank}(M), d(M)$ to be ∞. Also by the next lemma, we have that if $\mathrm{rank}(M) = n$ then there is an injective A-module homomorphism $A^k \to M$ for all $k \leq n$. By a similar argument, we have that if $d(M) = n$ then there is a surjective A-module homomorphism $A^k \to M$ for all $k \geq n$.

Lemma: If $m_1, \ldots, m_n \in M$ are linearly independent then $\theta : A^n \to \langle m_1, \ldots, m_n \rangle$ given by $(a_1, \ldots, a_n) \mapsto a_1m_1 + \cdots + a_nm_n$ is an A-module isomorphism.

Proof:
Every element of $\langle m_1, \ldots, m_n \rangle$ is an A-linear combination of the m_i. So θ is surjective. Also θ is clearly and A-module homomorphism. And finally, if $(a_1, \ldots, a_n) \in \ker(\theta)$ then $a_1m_1 + \ldots + a_nm_n = 0$. But as $m_1, \ldots, m_n$ are linearly independent, we thus know that $a_i = 0$ for all i. So, θ is injective. ■

Side note: One difference between linear algebra and general module theorey is that if $m \neq 0$ is in M, then $\{m\}$ is not necessarily linearly independent. For an example of this, consider $\mathbb{Z}/6\mathbb{Z}$ as a module with itself for scalars and set $m = 2 + 6\mathbb{Z}$. Then $\{m\}$ is not linearly independent as $3 + 6\mathbb{Z} \neq 0 + 6\mathbb{Z}$ and yet $(3 + 6\mathbb{Z})m = 0 + 6\mathbb{Z}$.

Proposition: If $\theta : A^m \to A^n$ is a surjective A-module homomorphism, then $m \geq n$.

Proof:

Set $e_i = (0, \ldots, 0, \underbrace{1}_{i\text{th. index}}, 0, \ldots, 0) \in A^m$ for all $i \in \{1, \ldots, m\}$. Then:

$$A^m = Ae_1 \oplus \cdots \oplus Ae_m$$

Next, let $a_{i,j} \in A$ be such that $\theta(e_j) = (a_{1,j}, \ldots, a_{n,j})$ for all i, j. And let

$$T = \begin{bmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,m} \end{bmatrix} \in M_{n,m}(A).$$

As a side note, the professor likes this physics notation:

$$|(r_1, \ldots, r_m)\rangle \coloneqq \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix} \text{ and } \langle (r_1, \ldots, r_m)| \coloneqq \begin{bmatrix} r_1 & \cdots & r_m \end{bmatrix}$$

We claim that θ is given by matrix-vector multiplication with T. After all:

$$\begin{aligned} |\theta((c_1, \ldots, c_m))\rangle &= |\theta(\textstyle\sum_{j=1}^m c_j e_j)\rangle \\ &= |\textstyle\sum_{j=1}^m c_j \theta(e_j)\rangle = \textstyle\sum_{j=1}^m c_j |\theta(e_j)\rangle \\ &= [|\theta(e_1)\rangle, \cdots, |\theta(e_m)\rangle]\, |c_1, \ldots, r_m\rangle \\ &= T|c_1, \ldots, r_m\rangle \end{aligned}$$

But now let $\mathfrak{m}$ be a maximal ideal in A and consider the map $T \pmod{\mathfrak{m}} \in M_{n,m}(A/\mathfrak{m})$. In other words, we are considering the map $|v\rangle \in (A/\mathfrak{m})^m \mapsto [T \pmod{m}]|v\rangle$. Since the matrix T is surjective, we know that $T \pmod{\mathfrak{m}}$ is also surjective. Also $A/\mathfrak{m}$ is a field. So by our knowledge of linear algebra, we must have that $m \geq n$. ■

Note that if $|d_1, \ldots, d_n\rangle = T|c_1, \ldots, c_m\rangle$, then:

$$|d_1 \pmod{\mathfrak{m}}, \ldots, d_n \pmod{\mathfrak{m}}\rangle = (T \pmod{\mathfrak{m}})|c_1 \pmod{\mathfrak{m}}, \ldots, c_m \pmod{\mathfrak{m}}\rangle.$$

This is why T being surjective means $T \pmod{\mathfrak{m}}$ is surjective.

Corollary: If $A^n \cong A^m$ then $n = m$.

Set 2 Problem 1 continued:

(e) $q(x) \coloneqq x^4 + 12x^3 - 9x + 6 \in (\mathbb{Q}[i])[x]$ (where $\mathbb{Q}[i] = \{x + iy : x, y \in \mathbb{Q}\}$);

My first claim I want to make is that the field of fractions of $\mathbb{Z}[i]$ is isomorphic to $\mathbb{Q}[i]$. After all, if $a + bi, c + di \in \mathbb{Z}[i]$ with $c + di \neq 0$, then we can identify $\frac{a+bi}{c+di}$ with $\frac{ac+bd}{c^2+d^2} + \frac{-ad+bc}{c^2+d^2} i \in \mathbb{Q}[i]$. Conversely, we can identify if $\frac{a}{b}, \frac{c}{d} \in \mathbb{Q}$ then we can identify $\frac{a}{b} + i\frac{c}{d}$ with $\frac{ad+bci}{bd} \in Q(\mathbb{Z}[i])$. By standard algebraic arguments, these identifications invert each other and preserve addition and multiplication.

Next, note that $\mathbb{Z}[i]$ is a U.F.D., and by *problem 7 on homework set 1*, a positive prime integer p is irreducible in $\mathbb{Z}[i]$ iff $x^2 \equiv -1 \pmod{p}$ doesn't have a solution. Using this fact, we can easily show that $p = 3$ is irreducible in $\mathbb{Z}[i]$.

$$0^2 = 0 \not\equiv -1 \pmod 3,\ 1^2 = 1 \not\equiv -1 \pmod 3, \text{ and } 2^2 = 4 \equiv 1 \not\equiv -1 \pmod 3.$$

As a result of Eisenstein's criterion using $p = 3$, we can thus see that $q(x)$ is irreducible in $(Q(\mathbb{Z}[i]))[x] \cong (\mathbb{Q}[i])[x]$.

(f) $r(x) \coloneqq (x-1)(x-2)\cdots(x-n) + 1$ in $\mathbb{Q}[x]$ where n is an odd integer.

Suppose to the contrary that there exists $r_1(x), r_2(x) \in \mathbb{Q}[x]$ of positive degree such that $r(x) = r_1(x)r_2(x)$. By Gauss' lemma v3, we can assume without loss of generality that $r_1, r_2 \in \mathbb{Z}[x]$. And importantly this will have the consequence that $r_1(k), r_2(k) \in \mathbb{Z}$ for all $k \in \mathbb{Z}$.

Next note that by plugging in any $j \in \{1, \ldots, n\}$ we get that $1 = r(j) = r_1(j)r_2(j)$. Since $r_1(j), r_2(j) \in \mathbb{Z}$, this implies that $r_1(j) = r_2(j) = \pm 1$.

But now if we consider the polynomials $r_1(x)^2 - 1$ and $r_2(x)^2 - 1$, we get that both polynomials have zeros at all $j \in \{1, \ldots, n\}$. By the long division theorem, this let's us conclude that $\deg(r_i(x)^2 - 1) \geq n$ for both i. Going a step further, we must have that $\deg(r_i(x)^2 - 1) = 2\deg(r_i(x))$ is even for both i. Therefore, since n is an odd integer we can go a step further and say that $\deg(r_i(x)^2 - 1) \geq n + 1$ for both i.

Finally, we can now conclude that $\deg(r_i) \geq \frac{n+1}{2}$ for both i. This is a contradiction because it suggests that:

$$n = \deg(r) = \deg(r_1)\deg(r_2) \geq \tfrac{n+1}{2} + \tfrac{n+1}{2} = n + 1. \blacksquare$$

Set 2 Problem 2: Suppose p is a prime in $\mathbb{Z}$, $a \in \mathbb{Z}$ and $p \not| a$. Prove that $x^{p^n} - x + a$ does not have zero in $\mathbb{Q}$.

By the rational root theorem plus the fact that $x^{p^n} - x + a$ is monic, we can conclude that any rational zero of our polynomial must be in $\mathbb{Z}$ and divide a. So, for the sake of contradiction suppose such an integer k exists such that $k^{p^n} - k + a = 0$.

Upon rearranging and modding by p, we get that $k - k^{p^n} \equiv a \not\equiv 0 \pmod{p}$. Yet also note by Fermat's little theorem that:

$$k^{p^n} = (k^{p^{n-1}})^p \equiv k^{p^{n-1}} = (k^{p^{n-2}})^p \equiv \cdots \equiv k^p \equiv k \pmod{p}.$$

Thus we have a contradiction as $k - k^{p^n} \equiv k - k \equiv 0 \pmod{p}$. So, we conclude no such k can exist. $\blacksquare$

Proposition: If D is an integral domain, then $\operatorname{rank}(D^n) = n$.

Proof:

Note that $e_1, \ldots, e_n$ are D-linearly independent (see two pages ago). Therefore $\operatorname{rank}(D^n) \geq n$. To show the other inequality, we must prove that if $v_1, \ldots, v_{n+1} \in D^n$ then the v_i are D-linearly dependent.

Let $F = Q(D)$ and view the v_i as elements of F^n. From linear algebra, we know that $v_1, \ldots, v_{n+1}$ are F-linearly dependent. So, there exists $c_1, \ldots, c_n \in F$ such that $c_i \neq 0$ for some i and $c_1 v_1 + \ldots + c_{n+1} v_{n+1} = 0$. But by multiplying by some $d \in D - \{0\}$, we can guarentee that $dc_j \in D$ for all j. Hence $(dc_1)v_1 + \ldots + (dc_{n+1})v_{n+1} = 0$ where each $dc_j \in D$. Also $dc_i \neq 0$. So, $v_1, \ldots, v_{n+1}$ are D-linearly dependent. ■

(Disclaimer: I'm doing this this after the assignment was due.)

Set 2 Problem 4: Suppose A is a commutative unital ring. Given a matrix $[a_{i,j}] \in M_n(A)$, let $\det[a_{i,j}] := \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}$ where S_n is the symmetric group and $\operatorname{sgn} : s_n \to \{\pm 1\}$ is the sign function.

The (ℓ, k)-minor of $x := [a_{i,j}]$ is the determinant of the $(n-1)$x$(n-1)$ matrix $x(\ell, k)$ obtained after removing the ℓ-th. row and the k-th. column of x.

Let $\operatorname{adj}(x) := [(-1)^{i+j} \det(x(j,i))] \in M_n(A)$ be the adjugate of A. We'll need the following properties (which I'll only spend time proving if I think they aren't obvious):

(a) $\det$ is multi-linear with respect to the columns and rows.

(b) $\det(I) = 1$ (where $I = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}$).

(c) For all $x, y \in M_n(A)$, $\det(xy) = \det(x)\det(y)$.

Proof:

Step 1: Suppose F is a field. Then for all $x, y \in M_n(F)$, $\det(xy) = \det(x)\det(y)$.

I go over the general idea of this in my MITx nxn systems of equations notes from highschool. Essentially, you can write y as a product of element row operation matrices and then show that the determinant multiplicative property holds when one of the two matrices being multiplied is an elementary row operation.

Step 2: Consider the polynomial ring $R = \mathbb{Z}[x_{i,j}, y_{i,j} : i,j \in \{1, \ldots, n\}]$ in $2n^2$ variables. As R is a domain, we can view R as a subring of its field of fractions $Q(R)$. Hence if x', y' are matrices in $M_n(R) \subseteq M_n(Q(R))$, then we have that:

$$\det(x')\det(y') = \det(x'y').$$

Step 3: Note the two following (hopefully obvious) lemmas:

- Lemma A: Suppose $f : A \to B$ is a unitary ring homomorphism. Then $g : M_n(A) \to M_n(B)$ defined by $[a_{i,j}] \mapsto [f(a_{i,j})]$ is a well-defined ring homomorphism.

- Lemma B: Let f and g be the ring homomorphisms described in the prior lemma. Then $\det(g(A)) = f(\det(A))$.

<u>Final Step:</u> Let $x = [a_{i,j}]$ and $y = [b_{i,j}]$ be matrices in $M_n(A)$. Then consider the evaluation homomorphism $f : R \to A$ satisfying that $f(x_{i,j}) = a_{i,j}$ and $f(y_{i,j}) = b_{i,j}$. Then let $g : M_n(R) \to M_n(A)$ be the ring homomorphism given by lemma (a).

If we let $x' = [x_{i,j}]$ and $y' = [y_{i,j}]$ be matrices in $M_n(R)$, then $g(x') = x$ and $g(y') = y$. By lemma (b), we have that:

$$\begin{aligned}\det(xy) = \det(g(x'y')) = f(\det(x'y'))&\\ = f(\det(x')\det(y')) = f(\det(x'))f(\det(y'))&\\ = \det(g(x'))\det(g(y'))&\\ = \det(x)\det(y). \blacksquare&\end{aligned}$$

(d) If $x \in M_n(A)$ has two identical columns or rows, then $\det x = 0$.

<u>Case 1:</u> Suppose $2 \neq 0$ in A. Then the determinant of any matrix $y \in M_n(A)$ performing a row swap or column swap operation is -1. Thus if $x = xy$, then $\det(x) = \det(xy) = -\det(x) \implies 2\det(x) = 0 \implies \det(x) = 0$.

<u>Case 2:</u> Suppose $2 = 0$ in A. Then as $-1 = +1$, our formula for the determinant of $x = [a_{i,j}]$ simplifies to $\det x = \sum_{\sigma \in S_n} \prod_{i=1}^n a_{i,\sigma(i)}$.

But now suppose the i_1th and i_2th columns are repeated. Then if $\tau \in S_n$ is the transposition of i_1 and i_2 and we consider the orbits of $\langle\tau\rangle \curvearrowright S_n$, we get that:

$$\det(x) = \sum_{\langle\tau\rangle\cdot\sigma \in S_n/\langle\sigma\rangle} 2\prod_{i=1}^n a_{i,\sigma(i)} = 0.$$

As for the case that two rows are repeated then just note as $\sigma \mapsto \sigma^{-1}$ is a bijection on S_n and $\prod_{i=1}^n a_{i,\sigma^{-1}(i)} = \prod_{i=1}^n a_{\sigma(i),i}$, we have that:

$$\sum_{\sigma\in S_n} \prod_{i=1}^n a_{i,\sigma(i)} = \sum_{\sigma\in S_n} \prod_{i=1}^n a_{i,\sigma^{-1}(i)} = \sum_{\sigma\in S_n} \prod_{i=1}^n a_{\sigma(i),i}$$

And now we can do analogous reasoning to the case where we had repeated columns.

(e) $\operatorname{adj}(x)x = x\operatorname{adj}(x) = \det(x)I$

Let $x = [a_{i,j}]$ and recall the Laplace expansion formula for a matrix determinant. Namely, if $\ell, k \in \{1, \ldots, n\}$ then we have that:

- $\det(x) = \sum_{j=1}^{n} a_{\ell,j} \cdot (-1)^{\ell+j}\det(x(\ell, j)) =$ the (ℓ,ℓ)th. entry of $x\operatorname{adj}(x)$.
- $\det(x) = \sum_{i=1}^{n} a_{i,k} \cdot (-1)^{i+k}\det(x(i, k)) =$ the (k,k)th. entry of $\operatorname{adj}(x)x$.

For a reminder on how to prove this, see my MITx nxn systems of equations notes from highschool. The proof definitely requires A to be commutative.

Also by combining property (d) with the formula for Laplace's expansion, we can also show that every other entry in $x\operatorname{adj}(x)$ and $\operatorname{adj}(x)x$ is 0.

Let $e_1, \ldots, e_n$ be the standard basis for A^n as described on *page 565*. We already saw on that page that an A-module homomorphism $\phi : A^n \to A^n$ can be associated with a matrix $x_\phi \in M_n(A)$. Furthermore, $x_\phi x_\psi = x_{\phi \circ \psi}$.

Now the actual problem:

(a) Prove x is a unit in $M_n(A)$ iff $\det(x) \in A^\times$.

$(\Longrightarrow)$
If $x \in (M_n(A))^\times$, then $1 = \det(I) = \det(x)\det(x^{-1})$. Hence, $\det(x) \in A^\times$.

$(\Longleftarrow)$
Suppose $\det(x) \in A^\times$. Then:

$$\begin{aligned} \operatorname{adj}(x)x = x\operatorname{adj}(x) = \det(x)I \Longrightarrow \\ (\det(x)^{-1}\operatorname{adj}(x))x = x(\det(x)^{-1}\operatorname{adj}(x)) = I \\ \Longrightarrow \det(x)^{-1}\operatorname{adj}(x) = x^{-1}. \end{aligned}$$

So, $x \in M_n(A)^\times$.

As a side note, note that the fact that $\frac{1}{\det(x)}\operatorname{adj}(x) = x^{-1}$ is equivalent to *theorem 2.14 on page 86*.

(b) Suppose $\phi : A^n \to A^n$ is an A-module homomorphism. Prove the following are equivalent.

i. ϕ is surjective;

ii. For all maximal ideals $\mathfrak{m}$ of A, the induced $(A/\mathfrak{m})$-linear map:

$$\overline{\phi} : (A/\mathfrak{m})^n \to (A/\mathfrak{m})^n \text{ defined by } \overline{\phi}(x + \mathfrak{m}^n) = \phi(x) + \mathfrak{m}^n$$

is a well-defined bijection;

iii. ϕ is bijective.

$(i. \Longrightarrow ii.)$

Let $\mathfrak{m}$ be any ideal of A (not necessarily maximal). To show that $\overline{\phi}$ is well a defined map, let $x_\phi \in M_n(A)$ be denoted by $[a_{i,j}]$. Then supposing that:

$$|(c_1, \ldots, c_n)\rangle + \mathfrak{m}^n = |(d_1, \ldots, d_n)\rangle + \mathfrak{m}^n,$$

we know that $|(c_1 - d_1, \ldots, c_n - d_n)\rangle \in \mathfrak{m}^n$. Or equivalently, $c_i - d_i \in \mathfrak{m}$ for all i.

But now $\sum_{j=1}^n a_{i,j}(c_j - d_j) \in \mathfrak{m}$ for all i. Hence:

$$\phi(|(c_1, \ldots, c_n)\rangle - |(d_1, \ldots, d_n)\rangle) = x_\phi|(c_1 - d_1, \ldots, c_n - d_n)\rangle \in \mathfrak{m}^n$$

It follows that $\phi(|(c_1, \ldots, c_n)\rangle) + \mathfrak{m}^n = \phi(|(d_1, \ldots, d_n)\rangle) + \mathfrak{m}^n$

So we've shown that $\overline{\phi}$ is a well defined map.

Since $x_{\overline{\phi}}$ is easily seen to be given by the matrix $x_\phi \pmod{\mathfrak{m}}$, all of this reasoning is just a special case of Lemma A from two pages ago.

In the particular case that $\mathfrak{m}$ is a maximal ideal, then we know that $A/\mathfrak{m}$ is a field. Hence, $\overline{\phi}$ is a linear map on a vector space $(A/\mathfrak{m})^n$. Also importantly, $\overline{\phi}$ is surjective since ϕ is as well. By our knowledge of linear algebra (i.e. the rank nullity theorem), we can conclude that $\overline{\phi}$ is a bijection.

$(ii. \implies iii.)$
Suppose ϕ is not a bijection. Then we know that x_ϕ is not invertible, and by part (a) that means that $\det(x_\phi) \notin A^\times$. But now we know there exists a maximal ideal $\mathfrak{m} \lhd A$ containing $\langle \det(x_\phi) \rangle$ (see *pages 456-457*). If we consider that:

$$0 + \mathfrak{m} = \det(x_\phi) + \mathfrak{m} = \det(x_{\overline{\phi}}) \notin (A/\mathfrak{m})^\times,$$

we know by part (a) that $x_{\overline{\phi}}$ is not invertible in $M_n(A/\mathfrak{m})$. It follows that $\overline{\phi}$ is not a bijection for that $\mathfrak{m}$.

$(iii. \implies i.)$
This is trivial. ■

Set 2 Problem 5 was proved as a proposition in lecture.

Set 2 Problem 6: An A-module M is called Noetherian if the following equivalent statements hold:

1. Every chain $\{N_i\}_{i \in I}$ of submodules of M (ordered by inclusion) has a maximum.
2. Every non-empty family of submodules of M has a maximal element.
3. The ascending chain condition holds in M. That means that if $N_1 \subseteq N_2 \subseteq \cdots$ are submodules of M then there exists $i_0 \in \mathbb{N}$ such that $N_{i_0} = N_{i_0+1} = \cdots$.
4. All the submodules of M are finitely generated.

The way that one proves that all these statements are equivalent is identical to how we showed the analgous statements are equivalent for Noetherian rings (see *page 488*). The only small lemma you need to show is as follows.

Suppose $\mathcal{C}$ is a chain of A-submodules of M (simply-ordered by inclusion). Then $\bigcup_{N \in \mathcal{C}} N$ is also an A-submodule of M.

Also notice that a ring A is Noetherian if and only if it is a Noetherian A-module.

Now the actual problem:

(a) Suppose N is a submodule of M. Prove that M is Noetherian if and only if M/N and N are Noetherian.

Submodule Correspondance Lemma: There is a bijective correspondance from the submodules of M containing N to the submodules of M/N given by $L \mapsto L/N$.

Proof:
From math 100a we already know $L \mapsto L/N$ is a bijective correspondance between subgroups of $(M/N, +)$ and subgroups of $(M, +)$ containing N. Therefore, it now suffices to show that L is a submodule iff L/N is a submodule. Or in other words, we need to show that for any $a \in A$ and $x \in L$ we have that $ax \in L$ iff $ax + N \in L/N$.

The $\implies$ direction is trivial. As for the other direction, note that:

$$ax + N \in L/N \implies ax \in \ell + N \text{ for some } \ell \in L.$$

But then as $N \subseteq L$, we have that $ax \in \ell + L = L$.

$(\Longrightarrow)$
Now all submodules of N are also submodules of M and thus finitely generated. Hence, N is Noetherian. Meanwhile, suppose L/N is a submodule of M/N. Then as L is generated by some $x_1, \ldots, x_n$ in M, we have that L/N is generated by $x_1 + N, \ldots, x_n + N$. So, all submodules of M/N are also finitely generated and we've proven that M/N is Noetherian.

$(\Longleftarrow)$
Suppose $L_1 \subseteq L_2 \subseteq \cdots$ is an increasing sequence of submodules of M.

Since it is easy to see that intersections of modules are modules, we have that $L_1 \cap N \subseteq L_2 \cap N \subseteq \cdots$ is an increasing sequence of submodules of N. As N is Noetherian, we can conclude that there exists j such that $L_j \cap N = L_{j+k} \cap N$ for all $k \in \mathbb{N}$.

Next, note that as the image of a submodule under a module homomorphism is another submodule, we have that $\frac{L_1+N}{N} \subseteq \frac{L_2+N}{N} \subseteq \cdots$ is another increasing of submodules in M/N. As M/N is Noetherian, we similarly know there exists j' such that $\frac{L_{j'}+N}{N} = \frac{L_{j'+k}+N}{N}$ for all $k \in \mathbb{N}$.

Finally, set $i_0 = \max(j, j')$. I claim $L_{i_0+k} = L_{i_0}$ for all $k \in \mathbb{N}$.

Suppose for the sake of contradiction that $y \in L_{i_0+k}$ but also $y \notin L_{i_0}$.

It's immediately clear that y can't be in N. But also note that $y + N \in \frac{L_{i_0}+N}{N}$. It follows that there exists $\ell \in L_{i_0}$ such that $y + N = \ell + N$. Yet this implies that $y - \ell \in L_{i_0+k} \cap N$. At the same time though, $y - \ell \notin L_{i_0}$ since that would contradict that $y = \ell + (y - \ell) \notin L_{i_0}$. So, we have shown that $y - \ell$ is contained in $L_{i_0+k} \cap N$ but not contained in $L_{i_0} \cap N$. This is a contradiction.

(b) Suppose A is a Noetherian ring and M is a finitely generated A-module. Prove that M is Noetherian.

Suppose $A = \operatorname{span}(x_1, \ldots, x_n)$. Then, there exists a surjective A-module homomorphism $\phi : A^n \to M$ given by $|(a_1, \ldots, a_n)\rangle \mapsto \sum_{i=1}^n a_i x_i$.

Next note that $A^n = \oplus_{i=1}^n A$. Importantly, as A is a Noetherian ring we also have that A is a Noetherian A-module. Also, we can show as follows that if $M_1, \ldots, M_n$ is a finite collection of Noetherian A-modules, then so is $\oplus_{i=1}^n M_i$.

It suffices to show that if M_1, M_2 are Noetherian then so is $M_1 \oplus M_2$. After all, in the general case we have that:

$$M_1 \oplus \cdots \oplus M_{n-1} \oplus M_n \cong (M_1 \oplus \cdots \oplus M_{n-1}) \oplus M_n$$

So, we can just conclude by our simple case plus an inductive hypothesis that $\oplus_{i=1}^n M_i$.

Fortuantely, if we view M_1 as a submodule of $M_1 \oplus M_2$ (i.e. identify M_1 with the internal direct sum $M_1 \oplus \{0\} \subseteq M_1 \oplus M_2$), then we know by assumption that M_1 is Noetherian. Furthermore, it is easily seen that $\frac{M_1 \oplus M_2}{M_1} \cong M_2$ which is also Noetherian. By part (a) of the problem, we thus know that $M_1 \oplus M_2$ is Noetherian.

So, we have a surjective A-module homomorphism $\phi : A^n \to M$ where A_n is Noetherian. By part (a), we know that $A_n/\ker(\phi)$ must also be Noetherian and by the first isomorphism theorem we know that $M \cong A^n/\ker(\phi)$. So, M is Noetherian.

Set 2 Problem 7: Suppose A is a unital commutative ring and $\phi : A^m \to A^n$ is an injective A-module homomorphism.

If A is a domain, then we can give a fairly simple argument for why $m \leq n$.

Recall that for a domain D, we showed on *page 567* that $\operatorname{rank}(D^n) = n$. But we can also easily see that for $\phi : D^m \to D^n$ to be injective, we must have that $\{\phi(e_1), \ldots, \phi(e_m)\}$ is linearly independent. Therefore, we must have that $m \leq \operatorname{rank}(D^n) = n$.

In this problem we shall prove that $m \leq n$ even if A is not a domain.

(a) Suppose A is a Noetherian ring. Prove that $m \leq n$.

Suppose to the contrary that $m > n$. Then note that $A^m \cong A^n \oplus A^{m-n}$. So, we may define $\widehat{\phi} : A^m \to A^m$ by $\widehat{\phi}(x) := (\phi(x), 0_{m-n})$ and $M := \{0\}^n \oplus A^{m-n} \subseteq A^m$.

By the definition of $\widehat{\phi}$ plus property 4 of the proposition on *page 563*, we have that $\widehat{\phi}(A^m) + M \subseteq A^m$ is an internal direct sum.

But also note that as ϕ is injective, we must have that $\widehat{\phi}$ is injective. As a result, since $M \cap \widehat{\phi}(A^m) = \{0\}$, we can conclude that $\widehat{\phi}(M) \cap \widehat{\phi}(\widehat{\phi}(A^m)) = \{0\}$ as well. So, $\widehat{\phi}^2(A^m) + \widehat{\phi}(M)$ is also an internal direct sum in $\widehat{\phi}(A^m)$. Then from there, we can conclude that: $\widehat{\phi}^2(A^m) + \widehat{\phi}(M) + M$ is an internal direct sum in A^m.

After all suppose $x_2 \in \widehat{\phi}^2(A^m)$, $x_1 \in \widehat{\phi}(M)$, and $x_0 \in M$ satisfy that $x_0 + x_1 + x_2 = 0$. Since $\widehat{\phi}(A^m) + M \subseteq A^m$ is an internal direct sum in M, we know that $x_0 = 0 = (x_2 + x_1)$. Next, as $\widehat{\phi}^2(A^m) + \widehat{\phi}(M)$ is an internal direct sum in $\widehat{\phi}(A^m)$, we have that $x_2 + x_1 = 0 \implies x_2 = 0 = x_1$.

By repeating this argument and using an inductive hypothesis, we can show that $\widehat{\phi}^k(A^m) + \widehat{\phi}^{k-1}(M)$ is an internal direct sum in $\widehat{\phi}^{k-1}(A^m)$ for all k. Then in turn, we can show similarly to above that for all $k \in \mathbb{Z}_{\geq 0}$:

$\widehat{\phi}^k(A^m) + \widehat{\phi}^{k-1}(M) + \cdots + \widehat{\phi}(M) + M$ is an internal direct sum in A^m.

But now as we can easily see check that $\widehat{\phi}^k(M) \neq \{0\}$ for all k, we must have that $M \subsetneq M + \widehat{\phi}(M) \subsetneq M + \widehat{\phi}(M) + \widehat{\phi}^2(M) \subsetneq \cdots$. Yet as A and thus A^m is Noetherian, this is a contradiction.

(b) Prove that $m \leq n$ even if A is not Noetherian.

Let $x_\phi \in M_{n,m}(A)$ be the matrix associated with ϕ. Then let A_0 be the subring of A which is generated by 1 and the entires of x_ϕ (Note I wrote subring and not ideal).

Because ϕ is given by matrix multiplication by x_ϕ, we can restrict the domain and codomain of ϕ to get an A_0-module homomorphism $\widetilde{\phi} : A_0^m \to A_0^n$. Also as ϕ is injective, so is $\widetilde{\phi}$.

Importantly though, A_0 is a finitely generated ring and thus Noetherian (see *page 517*) After all, it is simple to see that:

$$A_0 = \{ \sum_{I=(i_1,\ldots,i_n)} c_I a_1^{i_1} \cdots a_n^{i_n} : c_I \in \mathbb{Z}1_A, \text{ each } i_j \in \mathbb{Z}^{\geq 0}, \text{ and all but finitely many } c_I = 0\}.$$

By part (a) applied to $\widetilde{\phi}$ we can thus conclude that $m \leq n$.

Set 2 Problem 8: Prove that $\operatorname{rank}(M) \leq d(M)$.

Suppose $d(M) = n$ and $\operatorname{rank}(M) = m$. Then there exists a surjective A-module homomorphism $\phi : A^n \to M$ and an injective A-module homomorphism $\psi : A^m \to M$. In turn, if $\{e_1, \ldots, e_m\}$ is the standard A-basis of A^m then there exists $v_1, \ldots, v_n \in A^n$ such that $\phi(v_i) = \psi(e_i)$ for all i.

But now let $\theta : A^m \to A^n$ be the unique A-module homomorphism given by $\theta(e_i) = v_i$ for all i. Then the following diagram commutes.

$$\begin{tikzcd} A^m \arrow[d, "\theta"'] \arrow[dr, hook, "\psi"] & \\ A^n \arrow[r, two heads, "\phi"'] & M \end{tikzcd}$$

It follows that $\phi \circ \theta$ and in turn θ is injective. By the prior problem, this is only possible if $m \leq n$. ■

Set 2 Problem 9: Suppose A is a unital commutative ring and M is a finitely generated A-module. Suppose $d(M) = \operatorname{rank}(M) = n$.

(a) Suppose A is Noetherian. Prove that $M \cong A^n$.

Identically to how we proceeded in the prior problem, get homomorphisms satisfying the commuting diagram below:

$$\begin{tikzcd} A^n \arrow[d, "\theta"'] \arrow[dr, hook, "\psi"] & \\ A^n \arrow[r, two heads, "\phi"'] & M \end{tikzcd}$$

By identical reasoning as in the last problem, we can conclude that $\theta : A^n \to A^n$ is injective.

We first claim that $\theta(A^n) + \ker(\phi)$ is an internal direct sum in A^n.

After all suppose $x, y \in A^n$ satisfy that $\theta(x) + y = 0$ and $\phi(y) = 0$. then $\psi(x) = \phi(\theta(x) + y) = \phi(0) = 0$. But since ψ is injective, we in turn know that $x = 0$. In turn, $\theta(x) = 0$ and we also have that $y = 0$.

Next by analogous reasoning to problem 7a of this problem set, we can show for all $k \in \mathbb{Z}_{\geq 0}$ that $\theta^k(A^n) + \theta^{k-1}(\ker(\phi)) + \cdots + \theta(\ker(\phi)) + \ker(\phi)$ is an internal direct sum in A^n. Finally, suppose for the sake of contradiction that $\ker(\phi) \neq \{0\}$. Then by more analogous reasoning to problem 7a, we get that:

$$\ker(\phi) \subsetneq \ker(\phi) + \theta(\ker(\phi)) \subsetneq \ker(\phi) + \theta(\ker(\phi)) + \theta^2(\ker(\phi)) \subsetneq \cdots$$

is a strictly increasing sequence of submodules in A^n. But that is a contradiction because A and thus and A^n is Noetherian.

So, we must have that $\ker(\phi) = \{0\}$, thus proving that ϕ is an injection. Since ϕ is also a surjection, we have proven the existence of an isomorphism between A^n and M.

(b) Prove that $M \cong A^n$ even if A is not Noetherian.

To show the general case, set up this commuting diagram another time:

$$\begin{tikzcd} A^n \arrow[d, hook, "\theta"'] \arrow[dr, hook, "\psi"] & \\ A^n \arrow[r, two heads, "\phi"'] & M \end{tikzcd}$$

This time though, we'll assume for the sake of contradiction that ϕ is not injective. In turn we know there must exist some $y := (b_1, \ldots, b_n) \in \ker(\phi) - \{0\}$.

Let $x_\theta = [a_{i,j}] \in M_n(A)$ be the matrix associated with θ. Next, let A_0 be the subring of A which is generated by 1 as well as all the b_i and $a_{i,j}$. Also let $M_0 := \phi(A_0^n)$.

- As $\theta(A_0^n) \subseteq A_0^n$ due to all the $a_{i,j}$ being in A_0, we know that θ with it's domain and codomained restricted to A_0^n is a well-defined A_0-module homomorphism. Also θ is still injective after restricting it.
- Also it's clear that by restricting the domain of ϕ to A_0^n and codomain of ϕ to M_0, we get that ϕ is a well-defined surjective A_0-module homomorphism.
- Using the fact that $\psi = \phi \circ \theta$, we can get that $\psi(A_0)^n \subseteq M_0$. So, we can restrict the domain and codomain to A_0^n and M_0 respectively without any problems. Also, ψ will still be injective are restricting it.

This gives us the following commuting diagram:

$$\begin{tikzcd} A_0^n \arrow[d, hook, "\theta"'] \arrow[dr, hook, "\psi"] & \\ A_0^n \arrow[r, two heads, "\phi"'] & M_0 \end{tikzcd}$$

But finally $y \in A_0^n$ and $y \in \ker(\phi) - \{0\}$. Since A_0 is a Noetherian ring (by identical reasoning to problem 7a), this contradicts part (a) of this problem. ■

Sorry for getting handwavy by the end of this homework assignment. It's currently 1:26AM, and I know I won't be turning this in. So me doing the homework problems now is purely for the note-taking value.

1/26/2026

Math 220b Notes:

Given a collection of points $a_1, \ldots, a_N \in \mathbb{C}$ and positive integers $m_1, \ldots, m_N$, we can always find an entire function $f : \mathbb{C} \to \mathbb{C}$ which only has zeros at the a_n and with multiplicities given by m_n. After all, we can just consider the polynomial:

$$f(z) = \prod_{n=1}^{n}(z - a_n)^{m_n}.$$

A natural folowup question to ask is if we can find an entire function with infinitely many prescribed zeros of varying multiplicities. As it turns out the answer is yes.

A natural first idea one might have for proving this is to try setting:

$$f(z) = \prod_{a=1}^{\infty}(1 - \tfrac{z}{a_n})^{m_n}.$$

The issue with this approach though is that that product may diverge. As an example, consider setting $a_n = -n$ and $m_n = 1$ for all $n \in \mathbb{N}$. Then the product $f(z) := \prod_{n=1}^{\infty}(1 - \frac{z}{a_n})^{m_n} = \prod_{n=1}^{\infty}(1 + \frac{z}{n})$ diverges to ∞ at $z = 1$. After all, we get that $f(1) = 2(\frac{3}{2})(\frac{4}{3})(\frac{5}{4})\cdots$.

By noting that $\operatorname{ord}(\prod_{k=1}^{\infty} g_k, a) = \sum_{k=1}^{\infty} \operatorname{ord}(g_k, a)$, a natural way of modifying our first idea is to try and multiply each $(1 - \frac{z}{a_n})^{m_n}$ term by some other function with no zeros. That way, we don't add any unwanted zeros to the function we are constructing and we can maybe coerce the product into converging.

It turns out this modified approach will work. Although surprisingly, by the homework problem on *pages 549-550*, we know these other function we're multiplying onto $(1 - \frac{z}{a_n})^{m_n}$ would have to have the form e^{g_n} where $g_n \in O(\mathbb{C})$ for all n.

Given any $p \in \mathbb{Z}_{\geq 0}$, we define the pth. Weierstrass primary/elementary factor to be:

$$E_p(z) = \begin{cases} (1 - z) & \text{if } p = 0 \\ (1 - z)\exp(z + \frac{z^2}{2} + \cdots + \frac{z^p}{z}) & \text{if } p \neq 0 \end{cases}.$$

Note that if $a \in \mathbb{C} - \{0\}$, then $E_p(\frac{z}{a})$ is entire and it's only zero is a a simple one at $z = a$.

For motivation on why we are defining $E_p(z)$, note that the Taylor series for $-\operatorname{Log}(1 - z)$ about 0 is $\sum_{n=1}^{\infty} \frac{z^n}{n}$. Therefore, we'll ideally have that $E_p(z)$ is approximately equal to $1 = \frac{1-z}{1-z}$ when p is large and $|z|$ is small. The next lemma will make this idea more concrete.

Lemma A: $|E_p(z) - 1| \leq |z|^{p+1}$ if $|z| \leq 1$.

Proof:

If $p = 0$, then $|E_p(z) - 1| = |-z| = |z|$. So our proposed inequality trivialy holds.

Suppose $p > 0$ and set $u(z) = z + \frac{z^2}{2} + \cdots + \frac{z^p}{z}$. Then $E_p(z) = (1 - z)e^{u(z)}$ and we want to express $E_p(z)$ as a power series $\sum_{k=0}^{\infty} a_k z^k$ (which will have infinite radius of convergence since $E_p(z)$ is entire).

Claim 1: $a_0 = 1$.

Note that $a_0 = E_p(0) = (1-0)e^{u(0)} = 1$.

Claim 2: $a_1 = a_2 = \cdots = a_p = 0$.

By differentiating $u(z)$, we get that $u'(z) = 1 + z + \ldots + z^{p-1}$. In turn:

$$(1-z)u'(z) = 1 - z^p.$$

But that implies that:

$$E_p'(z) = -e^{u(z)} + (1-z)u'(z)e^{u(z)} = -e^{u(z)} + (1-z^p)e^{u(z)} = -z^p e^{u(z)}$$

Therefore, if we look at the taylor expansion $\sum_{k=1}^{\infty} k a_k z^{k-1}$ for $E_p'(z)$ about 0 and note that $e^{u(0)} = 1 \neq 0$, we must have that the lowest degree term in that power series is the z^p term. In other words, $a_1 = a_2 = \cdots = a_p = 0$.

Claim 3: $a_k \leq 0$ in $\mathbb{R}$ for all $k \geq p+1$.

We showed in claim 2 that:

$$\sum_{k=p+1}^{\infty} k a_k z^{k-1} = -z^p e^{u(z)}.$$

In turn, $\sum_{k=p+1}^{\infty} k a_k z^{k-(p+1)} = -e^{u(z)} = -\exp(z + \frac{z^2}{2} + \cdots + \frac{z^p}{p})$. Or in other words, we have that a_k equals $\frac{-1}{k}$ times the $(k-(p+1))$th. coefficient in the Taylor expansion of $\exp(z + \frac{z^2}{2} + \cdots + \frac{z^p}{p})$ for all $k \geq p+1$.

If we can show that all coefficients in the Taylor expansion of $\exp(z + \frac{z^2}{2} + \cdots + \frac{z^p}{p})$ are positive, we will be done. Fortunately, note that:

$$\exp(z + \tfrac{z^2}{2} + \cdots + \tfrac{z^p}{p}) = e^z e^{\frac{z^2}{2}} \cdots e^{\frac{z^p}{p}} = \prod_{n=1}^{p} \left(\sum_{\ell=0}^{\infty} \frac{1}{\ell! n^\ell} z^{n\ell} \right).$$

By taking successive Cauchy products of those series (i.e. foiling), we'll get that the coefficients of the power series for $\exp(z + \frac{z^2}{2} + \cdots + \frac{z^p}{p})$ are all sums of products of the $\frac{1}{\ell! n^\ell}$ (which are all positive). Hence, those coefficients are positive.

Unnecessary tangent: Here's how we can actually calculate the power series expansion $\sum_{k=0}^{\infty} c_k z^k$ for $e^{u(z)}$.

To start off, the derivative of $e^{u(z)}$ is $(1 + z + \ldots + z^{p-1})e^{u(z)}$. Therefore, we must have that $\sum_{k=0}^{\infty}(k+1)c_{k+1}z^k = (1 + z + \ldots + z^{p-1}) \sum_{k=0}^{\infty} c_k z^k$. Also, when you consider that $e^{u(0)} = 1$, we thus get the following recurrence relation:

- $c_0 = 1$
- $c_{k+1} = \frac{1}{k+1} \sum\limits_{\ell=0}^{\min(k,p-1)} c_{k-\ell}$

Next, we can easily calculate from that relation that $c_1 = \cdots = c_p = 1$. Furthermore, note that for any $k \geq p-1$ that:

$$c_{k+2} - \tfrac{k+1}{k+2} c_{k+1} = \tfrac{1}{k+2} c_{k+1} - \tfrac{1}{k+2} c_{k-p+1}.$$

In other words, $c_{k+2} = c_{k+1} - \frac{1}{k+2} c_{k-p+1}$. So finally (and this is how much simplified I'm capable of getting it before my attention span runs out), we have that the coefficients c_k are given by the following recurrence relation:

- $c_0 = c_1 = \cdots = c_p = 1$;
- $c_{k+1} = c_k - \frac{1}{k+1} c_{k-p}$ if $k \geq p$.

Claim 4: $\sum_{k=p+1}^{\infty} |a_k| = 1$

Note that $0 = E_p(1) = 1 + \sum_{k=p+1}^{\infty} a_k$. Therefore:

$$\sum_{k=p+1}^{\infty} |a_k| = -\sum_{k=p+1}^{\infty} a_k = 1$$

Finally, note that when $|z| \leq 1$, we have that:

$$|E_p(z) - 1| = |z^{p+1} \sum_{k=p+1}^{\infty} a_k z^{k-p-1}|$$

$$\leq |z|^{p+1} \sum_{k=p+1}^{\infty} |a_k||z^{k-p-1}| \leq |z|^{p+1} \sum_{k=p+1}^{\infty} 1|a_k| = |z|^{p+1}. \blacksquare$$

Lemma B: Given any sequence $\{a_n\}_{n\in\mathbb{N}} \subseteq \mathbb{C} - \{0\}$ such that $|a_n| \to \infty$, we have for all $r > 0$ that $\sum_{n=1}^{\infty} (\frac{r}{|a_n|})^n < \infty$.

Proof:

Since $|a_n| \to \infty$, we know there exists N such that $|a_n| \geq 2r$ for all $n \geq N$. In turn, for all $n \geq N$ we have that $(\frac{r}{|a_n|})^n \leq \frac{1}{2^n}$. And since $\sum_{n=0}^{\infty} \frac{1}{2^n} < \infty$, we can conclude by the comparison test that $\sum_{n=1}^{\infty} (\frac{r}{|a_n|})^n < \infty$. $\blacksquare$

With that we are ready to prove our desired theorem:

Weierstraß Factorization Problem: Let $\{a_n\}_{n\in\mathbb{N}}$ be any sequence of distinct elements in $\mathbb{C}$ with $|a_n| \to \infty$, and also let $\{m_n\}_{n\in\mathbb{N}} \subseteq \mathbb{Z}_{>0}$.

Note that the condition that $|a_n| \to \infty$ is equivalent to guarenteeing that the set $\{a_n : n \in \mathbb{N}\}$ has no limit points in $\mathbb{C}$.

We claim there exists an entire function $f : \mathbb{C} \to \mathbb{C}$ which only has zeros at the a_n and with multiplicity given by m_n.

Proof:

To start off, it will be convenient (for the sake of notation) to replace $\{a_n\}_{n\in\mathbb{N}}$ with the sequence:

$$\underbrace{a_1, \ldots, a_1}_{m_1 \text{ times}}, \underbrace{a_2, \ldots, a_2}_{m_2 \text{ times}}, \underbrace{a_3, \ldots, a_3}_{m_3 \text{ times}} \ldots$$

Importantly, doing this relabeling won't change the fact that $|a_n| \to \infty$ as $n \to \infty$. One other note is that we can without loss of generality assume $a_n \neq 0$ for any n. After all, if not we can just remove those terms from our sequence, solve the factorization problem to get an entire function $h(z)$, and then finally set $f(z) = z^m h(z)$ where m is the number of zero terms we removed from the sequence.

Now our goal is to pick a sequence $\{p_n\}_{n\in\mathbb{N}} \subseteq \mathbb{Z}_{\geq 0}$ such that $f(z) = \prod_{n=1}^{\infty} E_{p_n}(\frac{z}{a_n})$ converges absolutely and locally uniformly. Equivalently, we need $\sum_{n=1}^{\infty} |E_{p_n}(\frac{z}{a_n}) - 1|$ to converge locally uniformly.

Set $p_n = n-1$ for all n. This will guarentee that $\sum_{n=1}^{\infty} |E_{p_n}(\frac{z}{a_n}) - 1|$ converges normally (and hence locally uniformly).

Indeed let $K \subseteq \mathbb{C}$ be compact. Then there exists $r > 0$ such that $K \subseteq \overline{\Delta}(0,r)$. Since $|a_n| \to \infty$, there exists N such that $|a_n| \geq r$ if $n \geq N$. And now by lemma A, we have that:

$$|E_{p_n}(\tfrac{z}{a_n}) - 1| \leq |\tfrac{z}{a_n}|^{p_n+1} \leq (\tfrac{r}{|a_n|})^n \text{ for all } z \in K$$

Finally normal converge follows from lemma B.

It follows that $f(z) := \prod_{n=1}^{\infty} E_{p_n}(\frac{z}{a_n})$ is an entire function with a zero at each a_n. ■

Note that by no means did we show that the p_n used in the above proof are unique. On the contrary, it is easy to justify that we can always modify at least finitely many of the p_n.

By combining the above proof with the homework problem on *pages 549-550*, we can rewrite our prior result in the following more versatile form:

Theorem: If f is entire with countably infinite zeros, then $f(z) = z^m e^{h(z)} \prod_{n=1}^{\infty} E_{p_n}(\frac{z}{a_n})$ for some $h \in O(\mathbb{C})$, $\{a_n\}_{n\in\mathbb{N}} \subseteq \mathbb{C} - \{0\}$, $m \in \mathbb{Z}_{\geq 0}$ and $\{p_n\}_{n\in\mathbb{N}} \subseteq \mathbb{Z}_{\geq 0}$.

Taking a break from the lecture and looking at (Conway) Theorem VII.5.15, we can also solve the Weierstraß factorization problem on an arbitrary region $G \subseteq \mathbb{C}$.

Theorem VII.5.15: Let G be a region and let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence of distinct points in G with no limit point in G. Also let $\{m_n\}_{n\in\mathbb{N}}$ be a sequence of positive integers. Then there is a function $f \in O(G)$ whose only zeros are at the a_n and with multiplicity given by m_n

Proof:

Part 1: Suppose there exists $R > 0$ such that $\{z \in \mathbb{C} : |z| > R\} \subseteq G$ and $|a_n| \leq R$ for all n.

Note that this guarentees that G^{C} is a nonempty compact set. After all, G^{C} is clearly closed and bounded. Also, G^{C} can't be empty as G^{C} must contain at least one limit point of the a_n in the comapct set $\overline{\Delta}(0,R)$.

Then we claim there is a function $f \in H(G)$ solving the Weierstraß problem and which has the further property that $\lim_{z\to\infty} f(z) = 1$.

Like in the proof of the Weierstraß problem for entire functions, we replace $\{a_n\}_{n\in\mathbb{N}}$ with the sequence:

$$\underbrace{a_1, \ldots, a_1}_{m_1 \text{ times}}, \underbrace{a_2, \ldots, a_2}_{m_2 \text{ times}}, \underbrace{a_3, \ldots, a_3}_{m_3 \text{ times}} \ldots$$

Doing this will not change the fact that $\{a_n\}_{n\in\mathbb{N}}$ has no limit points. Next as G^{C} is compact, we know for each $n \in \mathbb{N}$ that there exists a point $b_n \in G^{\mathsf{C}}$ with:

$$|b_n - a_n| = \min_{w\in G^{\mathsf{C}}} |w - a_n|.$$

Also $\min_{w\in G^{\mathsf{C}}} |w - a_n| \to 0$ as $n \to \infty$ or else we'd know the a_n have a limit point in G. Therefore, $|a_n - b_n| \to 0$ as $n \to \infty$.

Now set $p_n = n$ and define $f(z) := \prod_{n=1}^{\infty} E_{p_n}(\frac{a_n - b_n}{z - b_n})$

Note that:

$$E_p(\tfrac{a_n-b_n}{z-b_n}) = (1 - \tfrac{a_n-b_n}{z-b_n})\exp(\tfrac{a_n-b_n}{z-b_n} + \tfrac{1}{2}\left(\tfrac{a_n-b_n}{z-b_n}\right)^2 + \cdots + \tfrac{1}{p}\left(\tfrac{a_n-b_n}{z-b_n}\right)^p).$$

The exponential factor contributes no zeros. Meanwhile $(1 - \frac{a_n-b_n}{z-b_n})$ only has a zero at $z = a_n$. And since $(1 - \frac{a_n-b_n}{z-b_n})' = \frac{-(a_n-b_n)}{(z-b_n)^2}$, we also know that zero is simple.

Based on the above reasoning, we know that if f converges absolutely locally uniformly on G, then the only zeros of f will be at the a_n and with the multiplicity we want. In other words, it suffices to show that if $K \subseteq G$ is compact, then:

$\sum_{n=1}^{\infty} |E_{p_n}(\frac{a_n-b_n}{z-b_n}) - 1|$ converges uniformly on K.

Let $\varepsilon = \inf\{|z-w| : z \in K, w \in G^{\mathsf{C}}\}$ and note that $\varepsilon > 0$. Then:

$$|\tfrac{a_n-b_n}{z-b_n}| \leq |a_n - b_n|\varepsilon^{-1}.$$

Next as $|a_n - b_n| \to 0$, we know there exists N such that $|a_n - b_n| < \varepsilon/2$ for all $n \geq N$. So, by *lemma A*, we have for $n \geq N$

$$|E_{p_n}(\tfrac{a_n-b_n}{z-b_n}) - 1| \leq (\tfrac{\varepsilon}{2} \cdot \varepsilon^{-1})^{p_n+1} = \tfrac{1}{2^{n+1}}.$$

By the Weierstrass M-test, we can conclude that $\sum_{n=1}^{\infty} |E_{p_n}(\frac{a_n-b_n}{z-b_n}) - 1|$ converges uniformly on K.

To finish this part, we need to show $f(z) = \prod_{n=1}^{\infty} E_{p_n}(\frac{a_n-b_n}{z-b_n}) \to 1$ as $|z| \to \infty$. Therefore, consider any $\varepsilon > 0$ and let $\kappa_1 > 0$ satisfy that $|e^w - 1| < \varepsilon$ when $|w| < \kappa_1$.

Like on page 546 (see $(*)$), we can choose some $\delta \in (0, 1)$ such that:

$\frac{1}{2}|z| \leq |\operatorname{Log}(1+z)| \leq \frac{3}{2}|z|$ when $|z| < \delta$.

Next note that if we fix $0 < \kappa_2 < \delta$, then we can set $R_1 > 0$ large enough so that $2R < \kappa_2(R_1 - R)$. And since $a_n, b_n \in \overline{\Delta}(0, R)$ for all n we know that:

$|\frac{a_n-b_n}{z-b_n}| \leq \frac{2R}{R_1-R} < \kappa_2$ for all $n \in \mathbb{N}$ and $|z| \geq R_1$.

In particular, this has the advantage that since $|E_{p_n}(\frac{a_n-b_n}{z-b_n}) - 1| < \kappa_2^{p_n+1} < 1$, we know $\operatorname{Log}(E_{p_n}(\frac{a_n-b_n}{z-b_n}))$ is well-defined for all $n \in \mathbb{N}$ and $|z| \geq R_1$. So, we can write $f(z) = \exp(\sum_{n=1}^{\infty} \operatorname{Log}(E_{p_n}(\frac{a_n-b_n}{z-b_n})))$ when $|z| \geq R_1$ and then note for all $|z| \geq R_1$ that:

$$\begin{aligned} |\sum_{n=1}^{\infty} \operatorname{Log}(E_{p_n}(\tfrac{a_n-b_n}{z-b_n}))| &\leq \sum_{n=1}^{\infty} |\operatorname{Log}(E_{p_n}(\tfrac{a_n-b_n}{z-b_n}))| \\ &\leq \tfrac{3}{2}\sum_{n=1}^{\infty} |E_{p_n}(\tfrac{a_n-b_n}{z-b_n}) - 1| \leq \tfrac{3}{2}\sum_{n=1}^{\infty} |\tfrac{a_n-b_n}{z-b_n}|^{p_n+1} \leq \tfrac{3}{2}\sum_{n=1}^{\infty} \kappa_2^{n+1} \\ &= \tfrac{3}{2} \cdot \tfrac{\kappa_2^2}{1-\kappa_2} \end{aligned}$$

By choosing κ_2 small enough, we can guarentee that $\frac{3}{2} \cdot \frac{\kappa_2^2}{1-\kappa_2} < \kappa_1$. Therefore, $|f(z) - 1| < \varepsilon$ when $|z| \geq R_1$.

Part 2:
Now suppose G is an arbitrary region in $\mathbb{C}$ with $\{a_n\}_{n\in\mathbb{N}}$ a sequence of distinct points in G with no limit point, and let $\{m_n\}_{n\in\mathbb{N}}$ be a sequence of positive integers. Then there is some disk $\overline{\Delta}(a, r) \subseteq G$ with $\Delta(a, r)$ containing none of the a_n. So, consider the Möbius transformation $Tz = (z - a)^{-1}$ and put $G_1 = T(G)$ and $b_n = Ta_n = (a_n - a)^{-1}$ for all n.

Part 1 clearly applies to the region G_1 and the sequences $\{b_n\}_{n\in\mathbb{N}}$ and $\{m_n\}_{n\in\mathbb{N}}$.
We know $\{b_n\}_{n\in\mathbb{N}}$ has no limit points in G_1 because T has a continuous inverse Möbius transformation. Suppose $b_{n_k} \to \beta$ where there exists $\alpha \in G$ with $T\alpha = \beta$. Then $a_{n_k} = T^{-1}(b_{n_k}) \to T^{-1}\beta = \alpha$. But that's a contradiction.

So, we can pick a function $g \in O(G_1)$ whose only zeros are at the b_n and with multiplicity m_n. Next, we set $f(z) = g(Tz)$. Then $f \in O(G - \{a\})$ and has only zeros at the a_n and with multiplicity m_n.

Finally, $f(z) = g(\frac{1}{z-a}) \to 1$ as $z \to a$, we know f has a removable singularity at a. So, we can extend f to be a function in $O(G)$ that has exactly the zeros we want. ■

One example of the significance of the Weierstraß factorization problem is as follows:
Suppose that $G \subseteq \mathbb{C}$ is a region and then note that $O(G)$ is an integral domain.
To see why, note that G cannot be countable as any open ball contained in G is not meager (recall the Baire category theorem). That said, any set S without a limit point in G must be countable. This is because G is σ-compact. As a consequence, if $f, g \in O(G)$ are not the constant zero functions, then fg can only have countably many zeros and is thus also not the constant zero function.

Claim: The field of fractions of $O(G)$ is precisely the set of all meromorphic functions on G.

Proof:
It's easy to see that if $g, h \in O(G)$ with $h \not\equiv 0$, then g/h is meromorphic in G. So, the field of fractions of $O(G)$ can be embedded in the set of meromorphic functions on G.

Conversely, suppose f is meromorphic in G and let S be the set of poles of f. Then there exists a function $h \in O(G)$ with zeros only at the points in S and which satisfies that the multiplicity of any zero in S of h in equal to the order of the corresponding pole of f.

Next set $g = fh$ on $G - S$. Then as all the singularities of g are removable, we can holomorphically extend g to being a function in $O(G)$. At last, we have found two holomorphic functions $g, h \in O(G)$ with $f = \frac{g}{h}$. ■

Another interpretation goes as follows:

A divisor on G is a formal sum $D = \sum_{p \in G} n_p[p]$ where $n_p \in \mathbb{Z}$ for all $p \in G$ and the set $\{p \in G : n_p \neq 0\}$ does not have a limit point.

We say a divisor $D = \sum_{p \in G} n_p[p]$ is effective (denoted $D \geq 0$) if $n_p \geq 0$ for all p.

Note that divisors can be added formally.

In particular, if S_1 and S_2 are sets in G without a limit point, then $S_1 \cup S_2$ also can't have a limit point. After all, if p was a limit point of $S_1 \cup S_2$, then any open set U containing p would have to contain infinitely many points in $S_1 \cup S_2$. But by making U small enough, we can guarentee that U contains only finitely many points in S_1 or S_2.

It follows that the divisors form an abelian group which we'll denote $(\mathrm{Div}, +)$.

A divisor D is principal if there exists a meromorphic function f on G with:

$$D = \sum_{p \in G} \mathrm{ord}(f, p)[p].$$

Note: We define $\mathrm{ord}(f, p)$ to be a negative integer if f has a pole at p and $\mathrm{ord}(f, p)$ is a positive integer if f has a zero at p.

We shall denote $\sum_{p \in G} \mathrm{ord}(f, p)[p] = \mathrm{div}(f)$.

Importantly, note that $\mathrm{div}(fg) = \mathrm{div}(f) + \mathrm{div}(g)$. Therefore, the collection of principal divisors forms a subgroup $(\mathrm{Prin}, +)$ of $(\mathrm{Div}, +)$.

Finally, we define the divisor class subgroup to be quotient group $(\mathrm{Cl}, +) := (\mathrm{Div}/\mathrm{Prin}, +)$.

Corollary of Weierstraß: $\mathrm{Cl} = \{0\}$.

Proof:

We show $\mathrm{Div} = \mathrm{Prin}$. To start off, if D is a divisor, then we can write $D = D_+ - D_-$ where both D_+ and D_- are effective. Next, by Weierstraß there exists $g, h \in O(G)$ with $D_+ = \mathrm{div}(g)$ and $D_- = \mathrm{div}(h)$. Finally, $\mathrm{div}(g/h) = D$. ∎

You may wonder why mathematicians went to the effort of defining Cl. Essentially, it's because we can more generally define holomorphic and meromorphic functions on objects called Riemann surfaces.

One more note from the lecture. Suppose an entire function f is given with countably infinite zeros. Then by the theorem on page 578, we may write f in the form:

$$f(z) = z^m e^{g(z)} \prod_{n=1}^{\infty} E_{p_n}\left(\tfrac{z}{a_n}\right)$$

A way of trying to make this form canonical is as follows:

Assume there exists $h \geq 0$ such that $\sum \frac{1}{|a_n|^{h+1}} < \infty$. Then pick the smallest $h \in \mathbb{Z}_{\geq 0}$ with $\sum \frac{1}{|a_n|^{h+1}} < \infty$. Now we claim that the following product (which we call the canonical product of genus h): $\prod_{n=1}^{\infty} E_h\left(\frac{z}{a_n}\right)$, converges absolutely and locally uniformly.

To show that it converges and absolutely locally uniformly, just note for any $r > 0$ that there exists N with $|a_n| > r$ for all $n \geq N$. In turn:

$\sum_{n=N}^{\infty} \left|E_h\left(\frac{z}{a_n}\right) - 1\right| \leq r^{h+1} \sum_{n=N}^{\infty} \frac{1}{|a_n|^{h+1}} < \infty$ for all $z \in \overline{\Delta(0, r)}$.

Math 220b Homework:

Set 3 Problem 1 (Greatest Common Divisor): Assume that f and g are entire functions. Show that there exists entire functions h, F, and G such that:

$f(z) = h(z)F(z)$, $g(z) = h(z)G(z)$, and F and G have no zommon zeros.

Proof:
If either f or g is the constant zero function, then the problem is trivial. After all, suppose $f \equiv 0$. Then $f = g \cdot 0$ and $g = g \cdot 1$. A similar trivial solution holds if both f and g are the constant zero function.

Now suppose that neither f nor g is the constant zero function. Then if S is a multiset containing all the zeros shared by both f and g, we may let $h \in O(\mathbb{C})$ be a solution to the Weierstraß problem based on that S. To put that into other words, we are picking an entire function h satisfying that $\operatorname{ord}(h, a) = \min(\operatorname{ord}(f, a), \operatorname{ord}(g, a))$ for all $a \in \mathbb{C}$.

Next, we claim that all the singularities of the functions $\frac{f}{h}$ and $\frac{g}{h}$ are removable. After all, if $h(a) = 0$ then we know that we can pick integers $m_1, m_2 > 0$ as well as entire functions $\widetilde{f}, \widetilde{g}, \widetilde{h}$ approaching nonzero limits as $z \to a$ such that:

$f(z) = (z-a)^{m_1}\widetilde{f}(z)$, $g(z) = (z-a)^{m_2}\widetilde{g}(z)$, and $h(z) = (z-a)^{\min(m_1,m_2)}\widetilde{h}(z)$.

Now $\frac{f}{h} = (z-a)^{m_1 - \min(m_1,m_2)}\frac{\widetilde{f}(z)}{\widetilde{h}(z)} \to$ something finite as $z \to a$. And a similar story holds for $\frac{g}{h}$. So, we may holomorphically extend $\frac{f}{h}$ and $\frac{g}{h}$ to being entire functions $F(z)$ and $G(z)$.

Finally, $F(z)$ and $G(z)$ have no common zeros. After all, $F(a) = 0$ if and only if $\operatorname{ord}(f, a) > \operatorname{ord}(g, a)$ so that $\operatorname{ord}(h, a) \neq \operatorname{ord}(f, a)$. Similarly, $G(a) = 0$ if and only if $\operatorname{ord}(g, a) > \operatorname{ord}(f, a)$ so that $\operatorname{ord}(h, a) \neq \operatorname{ord}(g, a)$. ■

Set 3 Problem 2: Let f be an entire function and $n \geq 1$. Show that there exists an entire function g such that $g^n = f$ if and only if the orders of all zeros of f are divisible by n.

$(\Longrightarrow)$
Suppose $g^n = f$. Then $\operatorname{ord}(f, a) = n \cdot \operatorname{ord}(g, a)$ for all $a \in \mathbb{C}$. Hence, all order of all zeros of f are divisible by n.

$(\Longleftarrow)$
Let h be an entire function whose zeros are precisely the same as f and which satisfies for each zero a that $\operatorname{ord}(h, a) = \frac{1}{n} \cdot \operatorname{ord}(f, a)$. Then h^n will have precisely the same zeros as f and with the same multiplicities as f. Hence, we know there is some entire function u such that $f(z) = \exp(u(z))h^n(z)$. Finally, set $g(z) = \exp(\frac{1}{n}u(z))h(z)$. Then g is entire and $g^n(z) = \exp(u(z))h^n(z) = f(z)$. ■.

Set 3 Problem 3 (The Weierstraß σ and ζ functions): Let $\omega_1, \omega_2 \in \mathbb{C} - \{0\}$ satisfy that $\omega_2/\omega_1 \notin \mathbb{R}$. Then consider the lattice $\Lambda = \{m\omega_1 + n\omega_2 : m, n \in \mathbb{Z}\}$.

(i) Show that $\sum_{\lambda \in \Lambda - \{0\}} \frac{1}{|\lambda|^3}$ converges.

Consider the function $f : \mathbb{R}^2 \to \mathbb{C}$ given by $f(x, y) = \frac{|x\omega_1 + y\omega_2|}{|x| + |y|}$. Notably, f is continuous on $\mathbb{R}^2 - \{0\}$. Also, because ω_1 and ω_2 are $\mathbb{R}$-linearly independent, we have that $|x\omega_1 + y\omega_2| = 0$ if and only if $x = y = 0$. As a result, f is nonzero on its domain.

Now as the set $S^1 := \{(x,y) : x^2+y^2=1\} \subseteq \mathbb{R}^2$ is compact, we know that f achieves a minimum c on S^1 (which must be strictly positive.) In other words, we know that if (x,y) is a unit vector in $\mathbb{R}^2$, then:

$$|x\omega_1 + y\omega_2| \geq c(|x|+|y|)$$

By multiplying both sides by any $r>0$, we can extend this result to applying to any vector in $\mathbb{R}^2 - \{0\}$. And finally, this result is trivially true when $x=y=0$.

As for the significance of this, note that there are $4k$ many integer solutions of $|m|+|n|=k$.

If we require both m and n to be strictly, positive, then there are $k-1$ solutions. By flipping the sign of m and n, we can then get that there are $4(k-1)$ solutions to $|m|+|n|=k$ where $m,n\neq 0$. Finally, we can manually count that there are 4 more solutions when we let either m or n be 0.

So at last we get that:

$$\sum_{\lambda\in\Lambda-\{0\}} \frac{1}{|\lambda|^3} \leq c^{-3} \sum_{\substack{m,n\in\mathbb{Z} \\ (m,n)\neq 0}} \frac{1}{(|m|+|n|)^3} = c^{-3}\sum_{k=1}^{\infty} \frac{4k}{k^3} = \frac{2\pi^2}{3c^3} < \infty$$

We now define the Weierstraß σ-function as the infinite product:

$$\sigma(z) = z\prod_{\lambda\in\Lambda-\{0\}} (1-\tfrac{z}{\lambda})\exp(\tfrac{z}{\lambda}+\tfrac{z^2}{2\lambda^2}) = z\prod_{\lambda\in\Lambda-\{0\}} E_2(\tfrac{z}{\lambda})$$

(ii) Show that σ is an entire function with simple zeros only at the points of Λ.

To show that σ is entire, it suffices to show that $\sum\limits_{\lambda\in\Lambda-\{0\}} |E_2(\frac{z}{\lambda})-1|$ converges locally uniformly on $\mathbb{C}$.

Suppose $K\subseteq\mathbb{C}$ is compact and choose $r>0$ such that $K\subseteq\Delta(0,r)$. Then if c is the constant from part (i), and $\min(|n|,|m|)\geq\frac{r}{c}$, we know that $|m\omega_1+n\omega_2|\geq r$. Therefore, for all $z\in K$ we have that:

$$\begin{aligned}
\sum_{\lambda\in\Lambda-\{0\}} |E_2(\tfrac{z}{\lambda})-1| &= \sum_{\substack{|m|,|n|<\frac{r}{c} \\ (m,n)\neq(0,0)}} |E_2(\tfrac{z}{m\omega_1+n\omega_2})-1| + \sum_{|m|,|n|\geq\frac{r}{c}} |E_2(\tfrac{z}{m\omega_1+n\omega_2})-1| \\
&\leq \sum_{\substack{|m|,|n|<\frac{r}{c} \\ (m,n)\neq(0,0)}} |E_2(\tfrac{z}{m\omega_1+n\omega_2})-1| + \sum_{|m|,|n|\geq\frac{r}{c}} \left|\tfrac{z}{m\omega_1+n\omega_2}\right|^3 \\
&\leq \sum_{\substack{|m|,|n|<\frac{r}{c} \\ (m,n)\neq(0,0)}} \max_{z\in K} |E_2(\tfrac{z}{m\omega_1+n\omega_2})-1| + r^3\sum_{\lambda\in\Lambda-\{0\}} \tfrac{1}{|\lambda|^3}
\end{aligned}$$

By the Weierstraß M-test, we can conclude that $\sum_{\lambda\in\Lambda-\{0\}} |E_2(\frac{z}{\lambda})-1|$ converges uniformly on K. So, the product $\sigma(z)$ also converges absolutely and locally uniformly. And by the Weierstraß convergence theorem, we can conclude that σ is entire.

It's also clear that σ has zeros only at the points of Λ and all those zeros are simple since each $E_2(\frac{z}{\lambda})$ has only a simple zero at λ and z has a simple zero at 0.

Weierstraß also defined the function ζ (not to be confused with the Reimann zeta function) by taking the logarithmic derivative $\zeta(z) = \frac{\sigma'(z)}{\sigma(z)}$.

(iii) Show that $\zeta(z) = \frac{1}{z} + \sum_{\lambda\in\Lambda-\{0\}}(\frac{1}{z-\lambda} + \frac{1}{\lambda} + \frac{z}{\lambda^2})$ and conclude that ζ is a meromorphic function with poles at $\lambda \in \Lambda$ with residue equal to 1.

Let $\{\lambda_n\}_{n\in\mathbb{N}}$ be any enumeration of the points in $\Lambda - \{0\}$. Then by a prior theorem in class (see *page 549*), we have that:

$\zeta(z) = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{(E_2(\frac{z}{\lambda_n}))'}{E_2(\frac{z}{\lambda_n})}$ where the latter converges locally uniformly on $\mathbb{C} - \Lambda$.

Yet note that that same theorem also shows that for any permutation $\sigma : \mathbb{N} \to \mathbb{N}$, we have that:

$\zeta(z) = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{(E_2(\frac{z}{\lambda_{\sigma(n)}}))'}{E_2(\frac{z}{\lambda_{\sigma(n)}})}$ where the latter converges locally uniformly on $\mathbb{C} - \Lambda$.

It follows that $\sum_{n=1}^{\infty} \frac{(E_2(\frac{z}{\lambda_n}))'}{E_2(\frac{z}{\lambda_n})}$ converges absolutely on $\mathbb{C} - \Lambda$. So, we can say without any problems that:

$$\zeta(s) = \frac{1}{z} + \sum_{\lambda\in\Lambda-\{0\}} \frac{(E_2(\frac{z}{\lambda}))'}{E_2(\frac{z}{\lambda})}$$

More generally, this reasoning shows that the series $\sum \frac{g_k'}{g_k}$ in the theorem statement on *page 549* converges absolutely on it's domain. I mention that here cause I may need that fact later (maybe).

Next, note that if $u(z) = z + \frac{z^2}{2}$, then

$$\frac{(E_2(\frac{z}{\lambda}))'}{E_2(\frac{z}{\lambda})} = \frac{\frac{-1}{\lambda}(\frac{z}{\lambda})^2 \exp(u(\frac{z}{\lambda}))}{(1-\frac{z}{\lambda})\exp(u(\frac{z}{\lambda}))} = \frac{\frac{-z^2}{\lambda^3}}{1-\frac{z}{\lambda}} = \frac{\frac{-z^2}{\lambda^3}}{\frac{\lambda-z}{\lambda}} = \frac{1}{\lambda^2}\cdot\frac{z^2}{z-\lambda} = \frac{1}{\lambda^2}\cdot\frac{\lambda^2+z^2-\lambda^2}{z-\lambda} = \frac{1}{z-\lambda} + \frac{z+\lambda}{\lambda^2}$$

Therefore, we've shown that $\zeta(z) = \frac{1}{z} + \sum_{\lambda\in\Lambda-\{0\}}(\frac{1}{z-\lambda} + \frac{1}{\lambda} + \frac{z}{\lambda^2})$.

Finally, ζ is meromorphic on $\mathbb{C}$ because it is the quotient of two holomorphic functions where the denominator is not the constant zero function. ζ clearly has poles precisely at each $\lambda \in \Lambda$ since that's where σ evaluates to zero.

Also, suppose $\gamma(t) = \lambda + re^{it}$ *(with $t \in [0, 2\pi]$) where r is small enough so that no other $\lambda' \in \Lambda$ are in $\overline{\Delta}(\lambda, r)$. Then by the residue theorem, we have that:

$$\mathrm{Res}(\zeta,\lambda) = \frac{1}{2\pi i}\int_\gamma \zeta(z) = \frac{1}{2\pi i}\int_\gamma \left(\frac{1}{z} + \sum_{\lambda'\in\Lambda-\{0\}}(\frac{1}{z-\lambda'} + \frac{1}{\lambda'} + \frac{z}{(\lambda')^2})\right)$$

Since $\{\gamma\}$ is compact and the series defining ζ converges locally uniformly, we can equivalently say that:

$$\mathrm{Res}(\zeta,\lambda) = \frac{1}{2\pi i}\int_\gamma \zeta(z) = \frac{1}{2\pi i}\left(\int_\gamma \frac{1}{z} + \sum_{\lambda'\in\Lambda-\{0\}} \int_\gamma(\frac{1}{z-\lambda'} + \frac{1}{\lambda'} + \frac{z}{(\lambda')^2})\right)$$

Finally, by applying the residue theorem to each of those smaller integral seperately, we get that:

$$\frac{1}{2\pi i}\int_\gamma \zeta(z) = \frac{1}{2\pi i}\left(\int_\gamma \frac{1}{z} + \sum_{\lambda'\in\Lambda-\{0\}} \int_\gamma(\frac{1}{z-\lambda'} + \frac{1}{\lambda'} + \frac{z}{(\lambda')^2})\right) = 1. \blacksquare$$

Set 3 Problem 4 (The Γ function): In set 2 problem 3 (see *pages 562-563*), we introduced the entire function G defined by $G(z) = \prod_{n=1}^{\infty}((1+\frac{z}{n})e^{-\frac{z}{n}}) = \prod_{n=1}^{\infty} E_1(\frac{z}{-n})$

(i) Using the definition of G, show that $G(z)G(-z) = \frac{\sin(\pi z)}{\pi z}$.

Since G converges absolutely, we have that:

$$G(z)G(-z) = \left(\prod_{n=1}^{\infty}((1+\tfrac{z}{n})e^{-\frac{z}{n}})\right) \cdot \left(\prod_{n=1}^{\infty}((1-\tfrac{z}{n})e^{\frac{z}{n}})\right)$$
$$= \prod_{n=1}^{\infty}(1-\tfrac{z^2}{n^2}) = \frac{\sin(\pi z)}{\pi z}$$

(ii) By inspecting the locations of the zeros, show that $G(z-1) = zG(z)e^{\gamma(z)}$ for some entire function γ.

Note that both $G(z-1)$ and $zG(z)$ have only the nonpositive integers as zeros (and those zeros are simple for both functions). Therefore, as $\mathbb{C}$ is simply connected we can conclude by problem 5 on the second homework set (see *page 549*) that $G(z-1) = zG(z)e^{\gamma(z)}$ for some entire function γ.

(iii) Using logarithmic derivatives, show $\gamma'(z) = 0$. Thus, γ is a constant and $G(z-1) = zG(z)e^{\gamma}$.

By taking logarithmic derivatives, we get that:

$$\frac{G'(z-1)}{G(z-1)} = \frac{1}{z} + \frac{G'(z)}{G(z)} + \gamma'(z)$$

Next note that $E_1'(z) = -ze^z$. In turn, we can say that:

$$\frac{E_1'(\frac{z}{-n})}{E_1(\frac{z}{-n})} = \frac{\frac{1}{-n}\cdot\frac{-z}{-n}e^{\frac{z}{-n}}}{(1+\frac{z}{n})e^{\frac{z}{-n}}} = \frac{\frac{-z}{n^2}}{1+\frac{z}{n}} = \frac{1}{n}\cdot\frac{-z}{z+n} = \frac{1}{n}\cdot\frac{-z+n-n}{z+n} = \frac{-1}{n} + \frac{1}{z+n}$$

But now $\frac{G'(z)}{G(z)} = \sum_{n=1}^{\infty}\left(\frac{1}{z+n} - \frac{1}{n}\right)$. Similarly, $\frac{G'(z-1)}{G(z-1)} = \sum_{n=1}^{\infty}\left(\frac{1}{z-1+n} - \frac{1}{n}\right)$.
Therefore, we can conclude that:

$$\gamma'(z) = \sum_{n=1}^{\infty}\left(\frac{1}{z-1+n} - \frac{1}{n}\right) - \sum_{n=1}^{\infty}\left(\frac{1}{z+n} - \frac{1}{n}\right) - \frac{1}{z}.$$

Finally, note that $\sum_{n=1}^{\infty}\left(\frac{1}{z-1+n} - \frac{1}{n}\right) = \frac{1}{z} + \sum_{n=1}^{\infty}\left(\frac{1}{z+n} - \frac{1}{n}\right)$

This is because the limit of the following partial sums are the same:

$$\left(\frac{1}{z-1+(1)} - \frac{1}{(1)}\right) + \left(\frac{1}{z-1+(2)} - \frac{1}{(2)}\right) + \cdots$$
$$= \frac{1}{z-1+(1)} + \left(-\frac{1}{(1)} + \frac{1}{z-1+(2)}\right) + \left(-\frac{1}{(2)} + \frac{1}{z-1+(3)}\right) + \cdots$$

Note that I'm trying to be careful because $\sum_{n=1}^{\infty}\frac{1}{z-1+n}$ and $\sum_{n=1}^{\infty}\frac{1}{n}$ are divergent series.

It follows that $\gamma'(z) = 0$ for all z.

(iv) Show $G(0) = 1$ and use (iii) to conclude that $G(1) = e^{-\gamma}$.

From the definition of G we have that $G(0) = \prod_{n=1}^{\infty}((1+0)e^0) = 1$. Then by plugging in $z = 1$ into the identity shown in part (iii), we get that $G(1) = e^{-\gamma}$.

(v) Using the definition of G as an infinite product and (iv), show that $\gamma = \lim_{n\to\infty}(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log(n))$. (This is called the Euler-Mascheroni constant).

Firstly, by the definition of G we can see that $G(1)$ is a positive real number. It follows by part (iv) that $\gamma = -\log(G(1)) = \log(\frac{1}{G(1)})$.

But now:

$$\begin{aligned}\log(\tfrac{1}{G(1)}) &= \log(\lim_{N\to\infty} \textstyle\prod_{n=1}^N (1+\frac{1}{n})^{-1} e^{1/n}) \\ &= \lim_{N\to\infty} \log(\textstyle\prod_{n=1}^N (1+\frac{1}{n})^{-1} e^{1/n}) \\ &= \lim_{N\to\infty} \left(\textstyle\sum_{n=1}^N \frac{1}{n} + \log(n) - \log(n+1)\right) \\ &= \lim_{N\to\infty} \left((\textstyle\sum_{n=1}^N \frac{1}{n}) - \log(N) + \log(N) - \log(N+1)\right) \\ &= \lim_{N\to\infty} \left((\textstyle\sum_{n=1}^N \frac{1}{n}) - \log(N)\right) + \lim_{N\to\infty} (\log(N) - \log(N+1)) \\ &= \lim_{N\to\infty} \left((\textstyle\sum_{n=1}^N \frac{1}{n}) - \log(N)\right) + 0\end{aligned}$$

(vi) Define the meromorphic function $\Gamma(z) := \frac{e^{-\gamma z}}{z} \frac{1}{G(z)}$. Show that Γ has poles only at the non-positive integers and that these poles are simple.

Γ only has a pole at z satisfying that $zG(z) = 0$. But that precisely happens at the non-positive integers. Also, all the zeros of z and $G(z)$ are simple, and z and $G(z)$ have no shared zeros. Hence, all the poles are also simple.

(vii) Show $\Gamma(1) = 1$. Then using part (iii) show that $\Gamma(z+1) = z\Gamma(z)$. In particular, this means that $\Gamma(n+1) = n!$. Thus, the Γ function is a generalization of the factorial.

By part (iii), we know that when z is not a non-positive integer then:

$$\Gamma(z+1) = \frac{e^{-\gamma(z+1)}}{(z+1)G(z+1)} = \frac{e^{-\gamma z}}{e^{\gamma}(z+1)G(z+1)} = \frac{e^{-\gamma z}}{G(z)} = z\Gamma(z)$$

Also, $\Gamma(1) = \frac{e^{-\gamma}}{G(1)} = 1$ by part (iv). By induction we can thus conclude that $\Gamma(n+1) = n!$ for all $n \in \mathbb{Z}_{\geq 0}$.

(viii) Find the residue of Γ at $z = 0$. Then using (vii) find the residues of Γ at all non-positive integers.

Since all the poles of Γ are simple, we have that $\operatorname{Res}(\Gamma, -n) = \lim_{z\to n}(z+n)\Gamma(z)$ for all $n \in \mathbb{Z}_{\geq 0}$. In particular, this means that:

$$\operatorname{Res}(\Gamma, 0) = \lim_{z\to 0} z\Gamma(z) = \lim_{z\to 0} \Gamma(z+1) = \Gamma(1) = 1.$$

Meanwhile, if $n \in \mathbb{Z}_{>0}$, then on some punctured disk about n we have that:

$$\Gamma(z) = \frac{\Gamma(z+1)}{z} = \cdots = \frac{\Gamma(z+n+1)}{z(z+1)\cdots(z+n-1)(z+n)}$$

Therefore:

$$\operatorname{Res}(\Gamma, -n) = \lim_{z\to -n} (z+n)\Gamma(z) = \lim_{z\to -n} \frac{\Gamma(z+n+1)}{z(z+1)\cdots(z+n-1)} = \frac{\Gamma(1)}{(-n)(-n+1)\cdots(-2)(-1)} = \frac{(-1)^n}{n!}.$$

(ix) Using the definition (vi) as well as the results in (i), (vii), show that $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$.

$$\Gamma(z)\Gamma(1-z) = \frac{e^{-\gamma z}}{zG(z)} \cdot \frac{e^{+\gamma z - \gamma}}{(1-z)G(1-z)} = \frac{1}{zG(z)\cdot e^{\gamma}(1-z)G(1-z)} = \frac{1}{zG(z)G(-z)} = \frac{\pi z}{z\sin(\pi z)} = \frac{\pi}{\sin(\pi z)}$$

(x) Deduce that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Note that $\Gamma(\frac{1}{2})^2 = \Gamma(\frac{1}{2})\Gamma(1-\frac{1}{2}) = \frac{\pi}{\sin(\frac{\pi}{2})} = \pi$. Thus, we have that $\Gamma(\frac{1}{2}) = \pm\sqrt{\pi}$. Furthermore as $G(\frac{1}{2})$ is positive (since it's a product of positive things), we can conclude from the definition of Γ that $\Gamma(\frac{1}{2})$ is also positive. Therefore, we have that:
$$\Gamma(\tfrac{1}{2}) = +\sqrt{\pi}$$

A different definition of the Γ function seen in Math 240 is given by $\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}\mathrm{d}t$. That other definition is only defined when $\mathrm{Re}(z) > 0$. Also, it can be shown (although I won't here) that both definitions agree with each other on their shared domain. If I ever show their equivalence, it will be on *page* _______.

I'm going to switch to cramming algebra now.

Math 200b Notes:

Note that if A is a unital commutative ring and $\mathfrak{a} \lhd A$, then $\mathfrak{a}$ is an A-module.

Lemma: A is a P.I.D. if and only if for all $\mathfrak{a} \lhd A$ with $\mathfrak{a} \neq \{0\}$, $\mathfrak{a} \cong A$ as an A-module.

($\Longrightarrow$)
Let $\mathfrak{a} \lhd A$ with $\mathfrak{a} \neq \{0\}$. Since A is a P.I.D., $\mathfrak{a} = Ax$ for some $x \in A - \{0\}$. Now, define $\theta : A \to \mathfrak{a}$ by $\theta(c) = cx$. Clearly θ is a surjective A-module homomorphism. Also, $c \in \ker(\theta) \Longleftrightarrow cx = 0 \Longleftarrow c = 0$ since A is a domain and $x \neq 0$.

($\Longleftarrow$)
Suppose $0 \neq \mathfrak{a} \lhd A$. Then there exists an A-module isomorphism $\theta : A \to \mathfrak{a}$. In turn, for all $x \in \mathfrak{a}$ there exists $c \in A$ such that $\theta(c) = x$. In particular, $x = \theta(c) = c\theta(1)$. Thus $x \in A\theta(1)$ for all $x \in \mathfrak{a}$. Or in other words $\mathfrak{a} \subseteq A\theta(1)$. Since $\theta(1) \in \mathfrak{a}$, we also know $A\theta(1) \subseteq \mathfrak{a}$. Hence, we have proven that $\mathfrak{a}$ is principal and equals $\langle\theta(1)\rangle$.

With that we've shown that every ideal of A is principal. To finish off, we need to show that A is a domain. So, suppose for the sake of contradiction that $ab = 0$ but $a \neq 0$ and $b \neq 0$. Then let $\mathfrak{a} = Aa \lhd A$. By assumption, there exists an A-module isomorphism $\theta : \mathfrak{a} \to A$. In particular, $\mathrm{im}(\theta) = \{\theta(ca) : c \in A\} = \{c\theta(a) : c \in A\} = A\theta(a)$.

Since $\mathrm{im}(\theta) = A$, this means that $\theta(a) = A^\times$. In particular, this means that $\theta(a)$ is not a zero divisor of A. Yet now we have a contradiction as:
$$0 = \theta(0) = \theta(ba) = b\theta(a) \Longrightarrow b = 0. \blacksquare$$

Theorem: Suppose D is a P.I.D. and $\{0\} \neq M \subseteq D^n$ is a D-submodule. Then:

1. M is a free D-module;
2. There exists $x_1, \ldots, x_n \in D^n$ and $a_1, \ldots, a_m \in D - \{0\}$ (with $m \leq n$) such that:

   2-a. $D^n = Dx_1 \oplus \cdots \oplus Dx_n$, 2-b. $a_1 \mid a_2 \mid \cdots \mid a_m$,

2-c. $M = Da_1x_1 \oplus \cdots \oplus Da_mx_m$.

Proof:
Let $\Sigma := \{\phi(M) : \phi \in \mathrm{Hom}_D(D^n, D)\}$. Importantly, note for all $\phi \in \mathrm{Hom}_D(D^n, D)$ that $\phi(M)$ is an ideal of D. This is because $\phi(M)$ is a submodule of D. As a result, for each $\phi \in \mathrm{Hom}_D(D^n, D)$ we can write $\phi(M) = \langle a_\phi \rangle$ for some $a_\phi \in D$ (since D is a P.I.D.).

Since P.I.D.s are Noetherian, we know Σ has a maximal element, say $\phi_1(M) = \langle a_1 \rangle$.

Claim 0: $a_1 \neq 0$.

Proof:
Since $M \neq \{0\}$, we know there exists some $y = (y^{(1)}, \ldots, y^{(n)}) \in M$ with $y^{(i)} \neq 0$ for some i. In turn, if we consider the projection homomorphism $P_i : D^n \to D$ onto the ith. coordinate, we have that $0 \neq y^{(i)} \in P_i(M) \in \Sigma$. So, $\{0\}$ is not a maximal ideal in Σ and in turn we can't possibly have that $a_1 = 0$.

Next pick $y_1 \in M$ such that $\phi_1(y_1) = a_1$.

Claim 1: For all $\phi \in \mathrm{Hom}_D(D^n, D)$, we have that $a_1 \mid \phi(y_1)$.

Proof:
Since D is a P.I.D., write $\langle a_1, \phi(y_1)\rangle = \langle d \rangle$ where $d \in D$. Then note that there exists $r, s \in D$ such that $d = ra_1 + s\phi(y_1) = (r\phi_1 + s\phi)(y_1)$. But note that $r\phi_1 + s\phi \in \mathrm{Hom}_D(D^n, D)$. Also, we have that:

$$(r\phi_1 + s\phi)(M) \supseteq \langle (r\phi_1 + s\phi)(y_1)\rangle = \langle d \rangle \supseteq \langle a_1 \rangle$$

Since $(r\phi_1 + s\phi)(M)$, $\langle a_1 \rangle \in \Sigma$ and $\langle a_1 \rangle$ is maximal in Σ, we can conclude that $(r\phi_1 + s\phi)(M) = \langle a_1 \rangle$. In particular, this let's us conclude that $\langle a_1 \rangle = \langle d \rangle$. So, we can conclude that $a_1 \mid \phi(y_1)$.

Corollary of claim 1: $y_1 = a_1x_1$ for some $x_1 \in D^n$.

Proof:
Consider the projection homomorphism $P_i : D^n \to D$ which maps $(c^{(1)}, \ldots, c^{(n)}) \mapsto c^{(i)}$. By claim 1, we get that $y_1^{(i)} = a_1x_1^{(i)}$ for some $x_1^{(i)} \in D$. By doing this for all i, we get that $x_1 = (x_1^{(1)}, \ldots, x_1^{(n)})$ satisfies that $a_1x_1 = y_1$.

Also note that $a_1 = \phi_1(y_1) = \phi_1(a_1x_1) = a_1\phi_1(x_1)$. Since $a_1 \neq 0$ and D is a domain, we can conclude that $\phi_1(x_1) = 1$. This also means $x_1 \neq 0$ since $\phi_1(0) = 0 \neq 1$

Claim 2: $D^n = Dx_1 \oplus \ker(\phi_1)$ and $M = Da_1x_1 \oplus (\ker(\phi_1) \cap M)$.

Proof:
For all $v \in D^n$:

$$\phi_1(v - \phi_1(v)x_1) = \phi_1(v) - \phi_1(v)\phi_1(x_1) = \phi_1(v) - \phi_1(v) = 0.$$

Therefore, $w = v - \phi_1(v)x_1 \in \ker(\phi_1)$ and $v = \phi_1(v)x_1 + w \in Dx_1 + \ker(\phi_1)$. Also, if $v \in Dx_1 \cap \ker(\phi_1)$ then $v = cx_1$ and $\phi_1(v) = 0$. But that implies that $0 = \phi_1(cx_1) = c$. So, $v = cx_1 = 0$.

This proves that $D = Dx_1 \oplus \ker(\phi_1)$ where the latter is an internal direct sum.

Next suppose $v \in M$. Then like before we may write $v = \phi_1(v)x_1 + w$ where $w \in \ker(\phi_1)$. But now $\phi_1(v) \in \phi_1(M) = \langle a_1 \rangle$. So, $\phi_1(v) = a_1 c$ for some $c \in D$. In turn, $v = ca_1x_1 + w$. Since both ca_1x_1 and v are in M, we must have that w is in M as well. Hence, we can say that $v \in Da_1x_1 + (\ker(\phi_1) \cap M)$.

The above reasoning shows that $M \subseteq Da_1x_1 + (\ker(\phi_1) \cap M)$. Since $Da_1x_1 = Dy_1 \subseteq M$, we can also see that $M \supseteq Da_1x_1 + (\ker(\phi_1) \cap M)$. Finally, note that $Da_1x_1 \cap (\ker(\phi_1) \cap M) \subseteq Dx_1 \cap \ker(\phi_1) = \{0\}$. This completes the proof that $M = Da_1x_1 \oplus (\ker(\phi_1) \cap M)$ where the latter is an internal direct sum.

We are now ready to prove the theorem statement. Firstly, we proceed by induction on $\operatorname{rank}(M)$ to prove part 1 of the theorem. As a side note, $\operatorname{rank}(M)$ will be finite because D^n is a Noetherian module and so $d(M)$ must be finite (see *set 2 problem 6 on page 570*). In turn, by *set 2 problem 8 on page 573*, we have that $\operatorname{rank}(M) \leq d(M)$ is finite. So, we run into no issues while doing induction.

First note that if $0 \neq v \in D^n$ and D is a domain, then $\{v\}$ is a D-linearly independent set. After all, if $v = (c^{(1)}, \ldots, c^{(n)})$ and $c^{(i)} \neq 0$ for some i, then $dc^{(i)} \neq 0$ unless $d = 0$. In turn, $dv \neq 0$ unless $d = 0$. The significance of this result is that any submodule $M \neq \{0\}$ of D^n must have strictly positive rank.

Next, recall from claim 2 that $M = Da_1x_1 \oplus (\ker(\phi_1) \cap M)$. For our base case, note that the following are equivalent:

(i) $\operatorname{rank}(M) = 1$;
(ii) $Da_1x_1 = M$;
(iii) $(\ker(\phi_1) \cap M) = \{0\}$.

(i $\implies$ ii)
If $Da_1x_1 \neq M$, then we must have that $(\ker(\phi_1) \cap M) \neq \{0\}$ as $M = Da_1x_1 \oplus (\ker(\phi_1) \cap M)$. But now if we pick $0 \neq v \in \ker(\phi_1) \cap M$, we have that $\{a_1x_1, v\}$ is a D-linearly independent set in M. After all, $a_1x_1 \neq 0$ and $v \neq 0$. Therefore $ca_1x_1 = 0 = c'v$ iff $c = c' = 0$. At the same time, by the definition of an internal direct sum we have that $ca_1x_1 = c'v$ iff $ca_1x_1 = 0 = c'v$. Hence, $\operatorname{rank}(M) \geq 2$.

(ii $\iff$ iii)
Because $M = Da_1x_1 \oplus (\ker(\phi_1) \cap M)$, we must have that
$(\ker(\phi_1) \cap M) = \{0\}$ if and only if $M = Da_1x_1$.

(ii $\implies$ i)
If $Da_1x_1 = M$ then we know that $d(M) = 1$. By set 2 problem 8 (on *page 573*), we in turn know that $\operatorname{rank}(M) \leq d(M) = 1$. And as $\operatorname{rank}(M) \geq 1$, we can conclude that $\operatorname{rank}(M) = 1$.

Now it is clear that if $\operatorname{rank}(M) = 1$ then $M = Da_1x_1$ is a free D-module. Meanwhile, if $\operatorname{rank}(M) > 1$ then we have that:

$$1 \leq \operatorname{rank}(\ker(\phi_1) \cap M) \leq \operatorname{rank}(M) - 1.$$

By an inductive hypothesis, we can conclude there is a D-module isomorphism $\theta : D^r \to \ker(\phi_1) \cap M$. Next, we define:

$$\tilde{\theta}(c_1, \ldots, c_{r+1}) = \theta(c_1, \ldots, c_r) + c_{r+1}(a_1 x_1).$$

Observe that $\tilde{\theta}$ is a D-module homomorphism and that $\mathrm{im}(\theta) \subseteq \mathrm{im}(\tilde{\theta})$ and $Da_1x_1 \subseteq \mathrm{im}(\tilde{\theta})$. It follows that $\tilde{\theta}$ is surjective. Furthermore, suppose $\tilde{\theta}(c_1, \ldots, c_{r+1}) = 0$. Then we must have that:

$$\theta(c_1, \ldots, c_r) = -c_{r+1}a_1x_1 \in Da_1x_1 \cap (\ker(\phi_1) \cap M) = \{0\}.$$

As θ is injective, we thus know that $c_i = 0$ for all i. So, $\tilde{\theta}$ is also injective. And this proves that $M \cong D^{r+1}$.

As for proving part 2 of the theorem, we similarly proceed by induction on $\mathrm{rank}(M)$.

For our base case, again suppose $\mathrm{rank}(M) = 1$. Then by the prior section of the proof we know that $M = Da_1x_1$.

Note that because $D^n = Dx_1 \oplus \ker(\phi_1)$, we can conclude like in the prior section of the proof that $\ker(\phi_1) = \{0\}$ iff $D^n = Dx_1$. But if the latter holds true, we're already done. Thus, suppose $\ker(\phi_1) \neq \{0\}$. Then we know by part 1 of theorem (which we just proved) that there is a D-module isomorphism $\theta : D^r \to \ker(\phi_1)$ where $r \geq 1$.

Like in the prior section, we can next define $\tilde{\theta} : D^{r+1} \to \ker(\phi_1) \oplus Dx_1$ by $\tilde{\theta}(c_1, \ldots, c_{r+1}) = \theta(c_1, \ldots, c_r) + c_{r+1}x_1$ to get another D-module isomorphism. As $D^{r+1} \cong D^n$, we can conclude that $r + 1 = n$. In other words, $\ker(\phi_1) \cong D^r = D^{n-1}$.

Finally, if $e_1, \ldots, e_{n-1}$ is the standard basis of D^{n-1}, then we can now see that $\ker(\phi_1) = D\theta(e_1) \oplus \cdots \oplus D\theta(e_{n-1})$. So at last, we have that a_1 as well as $x_1, \theta(e_1), \ldots, \theta(e_{n-1})$ satisfy part 2 of the theorem.

As for the induction step, assume $\mathrm{rank}(M) > 1$. Then we know $\ker(\phi_1) \cap M \neq \{0\}$ and in turn also that $\ker(\phi_1) \neq \{0\}$. Also by repeating some of the reasoning in our base case, we can see that $\ker(\phi_1) \cong D^{n-1}$

By an inductive hypothesis applied to $\ker(\phi_1)$ and $\ker(\phi_1) \cap M$, we can conclude that there exists $x_2, \ldots, x_n \in \ker(\phi_0)$ as well as $a_2, \ldots, a_m \in D - \{0\}$ (with $m \leq n$) such that:

- $\ker(\phi_1) = Dx_2 \oplus \cdots \oplus Dx_n$,
- $a_2 \mid a_3 \mid \cdots \mid a_m$,
- $\ker(\phi_1) \cap M = Da_2x_2 \oplus \cdots \oplus Da_mx_m$.

It's easy to see from there that $D^n = Dx_1 \oplus \ker(\phi_1) = Dx_1 \oplus Dx_2 \oplus \cdots \oplus Dx_n$ and that $M = Da_1x_1 \oplus (\ker(\phi_1) \cap M) = Da_1x_1 \oplus Da_2x_2 \oplus \cdots \oplus Da_mx_m$. The missing step is to show that $a_1 \mid a_2$.

Let $\langle d\rangle = \langle a_1, a_2\rangle$. Then $d = ra_1 + sa_2$ for some $r, s \in D$. But now consider the D-module homomorphism $\psi : D^n \to D$ such that $\psi(x_1) = r$, $\psi(x_2) = s$, and $\psi(x_i) = 0$ for all $i \geq 3$. We have that:
$$\psi(M) \ni \psi(a_1x_1 + a_2x_2) = a_1\psi(x_1) + a_2\psi(x_2) = a_1r + a_2s = d.$$

Hence, $\langle a_1\rangle \subseteq \langle d\rangle \subseteq \psi(M) \in \Sigma$. Yet $\langle a_1\rangle$ is maximal in Σ. So, $\psi(M) = \langle a_1\rangle$ and in particular this means that $\langle d\rangle = \langle a_1\rangle$. So, $a_1 \mid a_2$. ■

Before moving on, I think it's worth stating for the next theorem that I'll let A^0 denote the module $\{0\}$.

<u>Fundamental Theorem of finitely generated modules over a PID:</u>

(Existence): If D is a P.I.D. and M is a finitely generated D-module, then there exists $r \in \mathbb{Z}_{\geq 0}$ as well as $a_1, \ldots, a_m \in D - \{D^\times \cup \{0\}\}$ such that:

1. $a_1 \mid a_2 \mid \cdots \mid a_m$
2. $M \cong D^r \oplus \frac{D}{\langle a_1\rangle} \oplus \cdots \oplus \frac{D}{\langle a_m\rangle}$.

Proof:
Let $m_1, \ldots, m_n$ be a set of generators for M. Then there exists a surjective D-module homomorphism $\phi : D^n \to M$. By the first isomorphism theorem, we have that $M \cong \frac{D^n}{\ker(\phi)}$. If $\ker(\phi) = \{0\}$, we are done as $M \cong D^n$. Otherwise, by our prior theorem there exists $x_1, \ldots, x_n \in D^n$ and $a_1, \ldots, a_m \in D - \{0\}$ such that:

- $D^n = Dx_1 \oplus \cdots \oplus Dx_n$,
- $a_1 \mid a_2 \mid \cdots \mid a_m$,
- $\ker(\phi) = Da_1x_1 \oplus \cdots \oplus Da_mx_m$.

<u>Lemma:</u> If $\{M_i\}_{i\in I}$ is a family of A-modules with submodules $N_i \subseteq M_i$, then:
$$\frac{\oplus_{i\in I} M_i}{\oplus_{i\in I} N_i} \cong \oplus_{i\in I} \frac{M_i}{N_i}.$$

Proof:
Consider the map $\phi : \frac{\oplus_{i\in I} M_i}{\oplus_{i\in I} N_i} \to \oplus_{i\in I} \frac{M_i}{N_i}$ given by:
$$\phi((x_i)_{i\in I} + \oplus_{i\in I} N_i) = (x_i + N_i)_{i\in I}.$$

This is easily checked to be an A-module isomorphism.

It follows from that lemma that:
$$\frac{D^n}{\ker(\phi)} = \frac{Dx_1 \oplus \cdots \oplus Dx_n}{Da_1x_1 \oplus \cdots \oplus Da_mx_m \oplus \{0\} \oplus \cdots \oplus \{0\}} \cong \frac{Dx_1}{Da_1x_1} \oplus \cdots \oplus \frac{Dx_m}{Da_mx_m} \oplus \underbrace{D \oplus \cdots \oplus D}_{n-m \text{ times}}$$
$$\cong D^{n-m} \oplus \frac{D}{\langle a_1\rangle} \oplus \cdots \oplus \frac{D}{\langle a_m\rangle}.$$

Also if $a_1, \ldots, a_k$ are units (where $k \leq m$), then we can safely remove them since $\frac{D}{\langle u\rangle} \cong \{0\}$ for all $u \in D^\times$.

(Uniqueness): $(r, \langle a_1\rangle, \ldots, \langle a_m\rangle)$ are unique.

We are not ready to prove this yet. But we will be working towards proving it.

Given a P.I.D. D, let $M(r; a_1, \ldots, a_m) \coloneqq D^r \oplus \bigoplus_{i=1}^m \frac{D}{\langle a_i \rangle}$ where $r \in \mathbb{Z}_{\geq 0}$ and $a_i \in D - (D^\times \cup \{0\})$ satisfies that $a_1 \mid a_2 \mid \cdots \mid a_m$.

If M is an A-module, then the annihilator of M is:

$$\mathrm{Ann}_A(M) \coloneqq \{a \in A : am = 0 \text{ for all } m \in M\}.$$

Note that $\mathrm{Ann}_A(M)$ is always an ideal of A. After all:

$$\mathrm{Ann}_A(M) = \bigcap_{m \in M} \ker(a \mapsto am).$$

Similarly, if $m \in M$ then we define $\mathrm{Ann}_A(m) \coloneqq \{a \in A : am = 0\}$. This is also an ideal of A.

Lemma: $\mathrm{rank}(M(r; a_1, \ldots, a_m)) = r$.

Since there is an injection $D^r \hookrightarrow M(r; a_1, \ldots, a_m) =: M$, we know that:

$$r = \mathrm{rank}(D^r) \leq \mathrm{rank}(M).$$

Meanwhile, suppose $v_1, \ldots, v_{r+1} \in M$. Then note that $a_m \in \mathrm{Ann}_D(\bigoplus_{i=1}^m \frac{D}{\langle a_i \rangle}) - \{0\}$. Thus we can get that $a_m v_1, \ldots, a_m v_{r+1}$ are in $D^r \oplus \{0\}$. And since $\mathrm{rank}(D^r) = r$, there exists $c_1, \ldots, c_{r+1}$ not all zero such that $c_1 a_m v_1 + \ldots + c_{r+1} a_m v_r = 0$. Thus, the v_i are D-linearly dependent. And this proves that $\mathrm{rank}(M) < r + 1$. ■

Suppose D is an integral domain and M is a D-module. Then the torsion of M is defined to be:

$$\mathrm{Tor}(M) \coloneqq \{m \in M : \exists a \in D - \{0\} \ s.t.\ am = 0\}$$

In other words, $m \in \mathrm{Tor}(M)$ if $\mathrm{Ann}_D(m) \neq \{0\}$.

Lemma: $\mathrm{Tor}(M)$ is a submodule of M. (note this requires D to be a domain...)

Proof:

Suppose $m_1, m_2 \in \mathrm{Tor}(M)$ and $a_1, a_2 \in D$. We know there exists $c_1, c_2 \in D - \{0\}$ such that $c_1 m_1 = c_2 m_2 = 0$. Also as D is a domain, we know that $c_1 c_2 \neq 0$. That said:

$$c_1 c_2 (a_1 m_1 + a_2 m_2) = c_2 a_1 (c_1 m_1) + c_1 a_2 (c_2 m_2) = 0$$

Therefore $a_1 m_1 + a_2 m_2 \in \mathrm{Tor}(M)$. ■

Remark: If $\theta : M_1 \to M_2$ is a D-module isomorphism, then $\theta(\mathrm{Tor}(M_1)) = \mathrm{Tor}(M_2)$.

Why:

Suppose $\exists c \in D - \{0\}$ such that $cm = 0$. Then $c\theta(m) = \theta(cm) = \theta(0) = 0$. He nce, $\theta(m) \subseteq \mathrm{Tor}(M_2)$

This proves that $\theta(\mathrm{Tor}(M_1)) \subseteq \mathrm{Tor}(M_2)$. By symmetric reasoning, we can see that $\theta^{-1}(\mathrm{Tor}(M_2)) \subseteq \mathrm{Tor}(M_1)$. By applying θ to both sides, we get $\mathrm{Tor}(M_2) \subseteq \theta(\mathrm{Tor}(M_1))$. So, we can conclude that $\theta(\mathrm{Tor}(M_1)) = \mathrm{Tor}(M_2)$.

Consequently, if $\theta : M_1 \to M_2$ is a D-module isomorphism then $\mathrm{Tor}(M_1) \cong \mathrm{Tor}(M_2)$.

Lemma: $\mathrm{Tor}(M(r; a_1, \ldots, a_m)) = \frac{D}{\langle a_1 \rangle} \oplus \cdots \oplus \frac{D}{\langle a_m \rangle}$ (where the latter is an internal direct sum).

Note that we also write $\frac{D}{\langle a_1 \rangle} \oplus \cdots \oplus \frac{D}{\langle a_m \rangle} = M(0, a_1, \ldots, a_m)$.

Proof:
Note that $a_1 \cdots a_m \cdot x = 0$ for any $x \in M(0, a_1, \ldots, a_m)$. Hence, the $\supseteq$ inclusion is clear.

On the other hand, suppose that $(v, x) \in \mathrm{Tor}(M)$ where $v \in D^r$ and $x \in M(0, a_1, \ldots, a_m)$. Then there exists $c \in D - \{0\}$ such that $(cv, cx) = 0$. But $cv = 0 \implies v = 0$ as $c \neq 0$. So, $(v, x) \in M(0, a_1, \ldots, a_m)$. ■

If A is a unital commutative ring, we say $\mathfrak{a}_1, \ldots, \mathfrak{a}_m \lhd A$ are coprime if $\mathfrak{a}_i + \mathfrak{a}_j = A$ whenever $i \neq j$.

As a side note, suppose D is a P.I.D and let $a, b \in D$. Also let $d = \gcd(a, b)$ (see *page 519*). Then we claim that $\langle a \rangle + \langle b \rangle = \langle a, b \rangle$ is precisely equal to $\langle d \rangle$.

To see why, note that because $d \mid a$ and $d \mid b$, we know that $\langle a, b \rangle \subseteq \langle d \rangle$. On the other hand, as D is a P.I.D. we know that $\langle a, b \rangle = \langle c \rangle$ for some $c \in D$. But now as $c \mid a$ and $c \mid b$, we must have that $c \mid d$. Therefore, $\langle d \rangle \subseteq \langle c \rangle = \langle a, b \rangle$.

As a consequence, $\langle a \rangle + \langle b \rangle = A$ iff $\gcd(a, b) \in A^\times$. So this new definition of coprimeness purely generalizes the original definition.

Generalized Chinese Remainder Theorem: Suppose A is a unital commutative ring and $\mathfrak{a}_1, \ldots, \mathfrak{a}_m \lhd A$ are coprime. Then $\frac{A}{\mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_m} \cong \frac{A}{\mathfrak{a}_1} \oplus \cdots \oplus \frac{A}{\mathfrak{a}_m}$ via the A-algebra isomorphism $x + (\mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_m) \mapsto (x + \mathfrak{a}_1, \ldots, x + \mathfrak{a}_m)$

Proof:
Step 1: For all i, we have that $\mathfrak{a}_i + \bigcap_{j \neq i} \mathfrak{a}_j = A$.

Proof:
For all $j \neq i$, there exists $x_j \in \mathfrak{a}_j$ such that $x_j \equiv 1 \pmod{\mathfrak{a}_i}$. Then, $y := \prod_{j \neq i} x_j$ is in $\bigcap_{j \neq i} \mathfrak{a}_j$ and satisfies that $y \equiv 1 \pmod{\mathfrak{a}_i}$. It follows that there exists $x_i \in \mathfrak{a}_i$ with $y + x = 1$. And that shows that $\mathfrak{a}_i + \bigcap_{j \neq i} \mathfrak{a}_j = A$.

Step 2: $\theta : A \to \bigoplus_{i=1}^m \frac{A}{\mathfrak{a}_i}$ given by $\theta(x) = (x + \mathfrak{a}_1, \ldots, x + \mathfrak{a}_m)$ is a surjective A-algebra homomorphism (meaning it is an A-module homomorphism and a ring homomorphism).

As a side note, we can view $\bigoplus_{i=1}^m \frac{A}{\mathfrak{a}_i}$ as a ring by identifying it with the product ring $\prod_{i=1}^m \frac{A}{\mathfrak{a}_i}$. In general, an A-algebra M is an A-module such that M is also a ring and $a(m_1 m_2) = (a m_1) m_2 = m_1 (a m_2)$ for all $a \in A$ and $m_1, m_2 \in M$.

Proof:
The fact that θ is a homomorphism is obvious. What is less obvious is that θ is surjective. To prove this, it is necessary and sufficient to show that:

$$(0, \ldots, 0, 1 + \mathfrak{a}_i, 0, \ldots, 0) \in \mathrm{im}(\theta) \text{ for all } i.$$

Fortunately by step 1, we know there exists $y_i \in \bigcap_{j \neq i} \mathfrak{a}_j$ such that $y_i \equiv 1 \pmod{\mathfrak{a}_i}$. In turn $\theta(y_i) = (0, \ldots, 0, 1 + \mathfrak{a}_i, 0, \ldots, 0) \in \mathrm{im}(\theta)$.

Step 3: $\ker(\theta) = \bigcap_{i=1}^{m} \mathfrak{a}_i$.
Hopefully this is obvious.

To finish, just use the first isomorphism theorem for rings as well as the first isomorphism theorem for A-modules. ■

Corollary: If M is a finitely generated D-module and D is a PID, then M is torsion-free (meaning $\mathrm{Tor}(M) = \{0\}$) if and only if M is free.

$(\Longleftarrow)$
If M is free then we know that $M \cong D^r$ for some positive integer r. But $\mathrm{Tor}(D^r) = \{0\}$ since D is a domain. Therefore, $\mathrm{Tor}(M) = \{0\}$ as well.

$(\Longrightarrow)$
Using the fundamental theorem of finitely generated modules over a PID, write $M \cong D^r \oplus M(0, a_1, \ldots, a_m)$. Then $\{0\} = \mathrm{Tor}(M) = M(0, a_1 \ldots, a_m)$ implies that $M \cong D^r$. So, M is free. ■

Other remarks:

- I used multiple times throughout the proof on *pages 588-591* that if D is an integral domain, then every ideal of D is torsion free.
- Note that if D is not a P.I.D then there exists an ideal $\mathfrak{a} \lhd D$ that is not principle. In turn, $\mathfrak{a}$ is not free.

  To see why, first note that $\mathrm{rank}(\mathfrak{a}) = 1$. After all, we clearly have that $\{a\}$ is D-linearly independent for any $a \in \mathfrak{a} - \{0\}$. Meanwhile suppose b is another element in $\mathfrak{a}$. Then, $b(a) + (-a)b = 0$ but $-a \neq 0$. So, $\{a, b\}$ is not D-linearly independent.

  As isomorphisms preserve the rank of modules and $\mathrm{rank}(D^n) = n$ for all integers n, we must have that if $\mathfrak{a}$ were free then we'd have that $\mathfrak{a} \cong D$ via a D-module homomorphism $\theta : D \to \mathfrak{a}$. But now we'd have a contradiction as $\mathfrak{a} = \langle \theta(1) \rangle$ is principal.

- By combining these remarks, we have shown that the $(\Longrightarrow)$ implication of the above corollary is false if D is merely a domain instead of a P.I.D.

Let D be a P.I.D. and let $\mathcal{P}_0 \subseteq D$ contain precisely one element from each equivalence class of companions containing irreducible elements (see *page 518* and recall that all P.I.Ds are U.F.Ds).

Note that when D is a U.F.D. and $\gcd(a, b) = 1$, then $\langle ab \rangle = \langle a \rangle \cap \langle b \rangle$.
After all, the $\subseteq$ inclusion is trivial. Meanwhile, if $a \mid x$ and $b \mid x$, then because $\gcd(a, b) = 1$, the only way to not violate the unique factorization of x is if $ab \mid x$.
This shows that $\langle a \rangle \cap \langle b \rangle \subseteq \langle ab \rangle$.

By induction we can conclude that $\langle a \rangle = \bigcap_{p\in\mathcal{P}_0} \langle p^{\nu_p(a)}\rangle$. And since D is a P.I.D., we know from two pages ago that all the ideals $\langle p^{\nu_p(a)}\rangle$ are coprime. Hence by the generalized Chinese remainder theorem, we have that:

$$\frac{D}{\langle a \rangle} = \frac{D}{\bigcap_{p\in\mathcal{P}_0}\langle p^{\nu_p(a)}\rangle} \cong \bigoplus_{p\in\mathcal{P}_0} \frac{D}{\langle p^{\nu_p(a)}\rangle}.$$

Now at last we are going to prove the uniqueness part of the theorem on *page 591*.

Theorem: If $M(r; a_1, \ldots, a_m) \cong M(r', b_1, \ldots, b_\ell)$, then $r = r'$, $m = \ell$, and $\langle a_i \rangle = \langle b_i \rangle$ for all i.

Proof:
To start out, we know that $r = \mathrm{rank}(M(r; a_1, \ldots, a_m)) = \mathrm{rank}(M(r', b_1, \ldots, b_\ell)) = r'$.
Also, we know that:

$$\begin{aligned} M(0; a_1, \ldots, a_m) &= \mathrm{Tor}(M(r; a_1, \ldots, a_m)) \\ &\cong \mathrm{Tor}(M(r', b_1, \ldots, b_\ell)) = M(0; b_1, \ldots, b_\ell). \end{aligned}$$

Hence, it now suffices to show that if $M(0; a_1, \ldots, a_m) \cong M(0; b_1, \ldots, b_\ell)$ then $m = \ell$ and $\langle a_i \rangle = \langle b_i \rangle$ for all i.

Part 1: Building a Counting Machine
Consider that:

$$M := M(0; c_1, \ldots, c_m) = \bigoplus_{i=1}^{m} \frac{D}{\langle c_i \rangle} \cong \bigoplus_{i=1}^{m}\left(\bigoplus_{p\in\mathcal{P}_0} \frac{D}{\langle p^{\nu_p(c_i)}\rangle}\right) \cong \bigoplus_{p\in\mathcal{P}_0}\left(\bigoplus_{i=1}^{m} \frac{D}{\langle p^{\nu_p(c_i)}\rangle}\right)$$

We shall denote $M(c_1, \ldots, c_m; p) := \bigoplus_{i=1}^{m} \frac{D}{\langle p^{\nu_p(c_i)}\rangle}$.

Importantly note that as $c_1 \mid c_2 \mid \cdots \mid c_m$, we know that $\nu_p(c_1) \leq \cdots \leq \nu_p(c_m)$ for all $p \in \mathcal{P}_0$. Since $M(c_1, \ldots, c_m; p)$ has that additional structure, we'll try studying it.

- Note that if A is a commutative ring, then for any A-module N and $b \in A$ we have that $\ell_b(m) := bm$ is an A-module homomorphism from N to itself. It follows that $b \cdot M = \mathrm{im}(\ell_b)$ is an A-module.

- If $\{M_i\}_{i\in I}$ is a family of A-modules and $b \in A$, then $b \cdot \bigoplus\limits_{i\in I} M_i = \bigoplus\limits_{i\in I}(b \cdot M_i)$.

- Suppose $\theta : M_1 \to M_2$ is an A-module isomorphism. Then $\theta(b \cdot M_1) = b \cdot M_2$.
  Hopefully this is easy to see.

- In particular, if D is a P.I.D. and $a \in D$, then we have that
  $$b \cdot \frac{D}{\langle a \rangle} = \frac{bD + \langle a \rangle}{\langle a \rangle} = \frac{\langle b, a \rangle}{\langle a \rangle} = \frac{\langle \gcd(a,b)\rangle}{\langle a \rangle}$$

  Consequently, if p_0, p are prime elements and r, k are nonnegative integers, then:
  $$p_0^r \cdot \frac{D}{\langle p^k \rangle} = \begin{cases} \frac{D}{\langle p^k \rangle} & \text{if } p_0 \neq p \\[2ex] \frac{\langle p^{\min(r,k)}\rangle}{\langle p^k \rangle} & \text{if } p_0 = p \end{cases}$$

- $\frac{\langle \gcd(a,b) \rangle}{\langle a \rangle} \cong \frac{D}{\langle \frac{a}{\gcd(a,b)} \rangle}$ as D-modules.

  Proof:

  Let $d = \gcd(a,b)$ and define $\widehat{\theta} : D \to \frac{\langle d \rangle}{\langle a \rangle}$ by $\widehat{\theta}(x) := dx + \langle a \rangle$. Then $\widehat{\theta}$ is a surjective D-module homomorphism.

  Also, $x \in \ker(\widehat{\theta})$ iff $dx \in \langle a \rangle$, and that happens iff $dx = ay$ for some $y \in D$. But now as $d \mid a$ and D is an integral domain, we can say that $x = \frac{a}{d} y$ for some $y \in D$.

  Hence, $\ker(\widehat{\theta}) = \langle \frac{a}{d} \rangle$ and we finish by invoking the first isomorphism theorem.
  ■

Now for all integers $r \geq 0$ and $p_0 \in \mathcal{P}_0$, we have that:

$$\begin{aligned} p_0^r \cdot M &\cong \bigoplus_{p \in \mathcal{P}_0} p_0^r \cdot M(c_1, \ldots, c_m; p) \\ &= \bigoplus_{p \in \mathcal{P}_0} \left(\bigoplus_{i=1}^m p_0^r \cdot \frac{D}{\langle p^{\nu_p(c_i)} \rangle} \right) \\ &\cong \left(\bigoplus_{p \in \mathcal{P}_0 - \{p_0\}} \left(\bigoplus_{i=1}^m \frac{D}{\langle p^{\nu_p(c_i)} \rangle} \right) \right) \oplus \bigoplus_{i=1}^m \frac{\langle {p_0}^{\min(r, \nu_{p_0}(c_i))} \rangle}{\langle {p_0}^{\nu_{p_0}(c_i)} \rangle} \end{aligned}$$

In order to cancel out the first ugly direct sum, we do an algebraic trick (which is referred to as taking a graded filtration). Consider the submodules:

$$M \supseteq p_0 \cdot M \supseteq p_0^2 \cdot M \supseteq \cdots$$

Now by recalling the lemma on *page 591*, we get that:

$$\frac{p_0^r \cdot M}{p_0^{r+1} \cdot M} \cong \frac{\bigoplus_{i=1}^m \frac{\langle {p_0}^{\min(r, \nu_{p_0}(c_i))} \rangle}{\langle {p_0}^{\nu_{p_0}(c_i)} \rangle}}{\bigoplus_{i=1}^m \frac{\langle {p_0}^{\min(r+1, \nu_{p_0}(c_i))} \rangle}{\langle {p_0}^{\nu_{p_0}(c_i)} \rangle}} \cong \bigoplus_{i=1}^m \frac{\left(\frac{\langle {p_0}^{\min(r, \nu_{p_0}(c_i))} \rangle}{\langle {p_0}^{\nu_{p_0}(c_i)} \rangle} \right)}{\left(\frac{\langle {p_0}^{\min(r+1, \nu_{p_0}(c_i))} \rangle}{\langle {p_0}^{\nu_{p_0}(c_i)} \rangle} \right)}.$$

Lemma: Suppose $M \supseteq N \supseteq L$ are D-modules. Then $(M/L)/(N/L) \cong (M/N)$

Proof:

We already know from the third isomorphism theorem for groups that $(x + L) + N/L \mapsto x + N$ is a group isomorphism. Then it's trivial to further check that this group isomorphism is also A-module homomorphism.

So, $\frac{p_0^r \cdot M}{p_0^{r+1} \cdot M} \cong \bigoplus_{i=1}^m \frac{\langle {p_0}^{\min(r, \nu_{p_0}(c_i))} \rangle}{\langle {p_0}^{\min(r+1, \nu_{p_0}(c_i))} \rangle}$.

But now consider that $\frac{\langle {p_0}^{\min(r, \nu_{p_0}(c_i))} \rangle}{\langle {p_0}^{\min(r+1, \nu_{p_0}(c_i))} \rangle} \cong \begin{cases} \{0\} & \text{if } \nu_{p_0}(c_i) \leq r \\ \frac{D}{\langle p_0 \rangle} & \text{if } \nu_{p_0}(c_i) > r \end{cases}$

Therefore, we have that $\frac{p_0^r \cdot M}{p_0^{r+1} \cdot M} \cong \left(\frac{D}{\langle p_0 \rangle} \right)^{\#\{i \,:\, \nu_{p_0}(c_i) > r\}}$ (where $\#S$ denotes the cardinality of the set S).

Lemma: If M is an A-module, $\mathfrak{b} \lhd A$, and $\theta : M \to (A/\mathfrak{b})^k$ is an A-module isomorphism, then we can view M as an $(A/\mathfrak{b})$-module by defining $(a + \mathfrak{b})m = am$. Furthermore, θ is then an $(A/\mathfrak{b})$-module isomorphism when M is equipped with this multiplication operation.

Proof:
Suppose $a_1 \equiv a_2 \pmod{\mathfrak{b}}$. Then $a_1 - a_2 \equiv 0 \pmod{\mathfrak{b}}$. In turn, for any $m \in M$ we must have that

$$\theta((a_1 - a_2)m) = (a_1 - a_2)\theta(m) = 0$$

And because θ is an isomorphism, that implies that $(a_1 - a_2)m = 0$. So, $a_1 m = a_2 m$ and we've proven that our scalar multiplication operation is well-defined. All the needed properties of this scalar multiplication are now easily seen as being inherited from the old scalar multiplication.

At last we get to the key observation. Note that as D is a P.I.D. and p_0 is irreducible, we know that $\frac{D}{\langle p_0 \rangle}$ is a field. Hence, our prior lemma says that $\frac{p_0^r \cdot M}{p_0^{r+1} \cdot M}$ is a $\frac{D}{\langle p_0 \rangle}$-vector space whose dimension is precisely the number of c_i for which $p_0^r \mid c_i$.

<u>Part 2: Counting the irreducible factors of $a_1, \ldots, a_m$ and $b_1, \ldots, b_\ell$</u>
Let $N_1 := M(0; a_1, \ldots, a_m)$ and $N_2 := M(0; b_1, \ldots, b_\ell)$, and suppose $N_1 \cong N_2$.

<u>Remark:</u> If M_1, M_2 are A-modules, $N \subseteq M_1$ is a submodule, and $\theta : M_1 \to M_2$ is an A-module isomorphism, then $\frac{M_1}{N} \cong \frac{M_2}{\theta(N)}$.

Proof:
Let $\pi : M_2 \to M_2/\theta(N)$ be the A-module homomorphism $m \mapsto m + \theta(N)$. Then, consider the map $\phi := \pi \circ \theta$. Clearly ϕ is surjective since both θ and π are. Also, $x \in \ker(\phi)$ iff $\theta(x) \in \theta(N)$. And since θ is injective, that happens iff $x \in N$. To finish off, we invoke the first isomorphism theorem.

As a result of the above remark plus an earlier remark that if $\theta : M_1 \to M_2$ is an A-module isomorphism then $\theta(b \cdot M_1) = b \cdot M_2$, we can now conclude for all for all $p \in \mathcal{P}_0$ and $r \in \mathbb{Z}_{\geq 0}$, we have that:

$$\left(\frac{D}{\langle p \rangle}\right)^{\#\{i \,:\, \nu_p(a_i) > r\}} \cong \frac{p^r \cdot N_1}{p^{r+1} \cdot N_1} \cong \frac{p^r \cdot N_2}{p^{r+1} \cdot N_2} \cong \left(\frac{D}{\langle p \rangle}\right)^{\#\{i \,:\, \nu_p(b_i) > r\}}$$

Since vector space isomorphisms preserve dimension, we can conclude that:

$$\#\{i \;:\; \nu_p(a_i) > r\} = \#\{i \;:\; \nu_p(b_i) > r\} \text{ for all } p \in \mathcal{P}_0 \text{ and } r \in \mathbb{Z}_{\geq 0}$$

This let's us show that $m = \ell$ and that $\langle a_i \rangle = \langle b_i \rangle$ for all i.

To start off, let $M = \max\{\#\{i \;:\; \nu_p(a_i) > 0\} : p \in \mathcal{P}_0\}$.

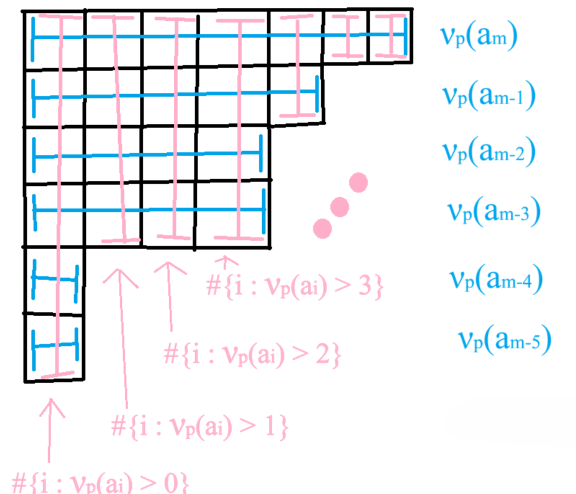
It's clear that $M \leq m$. Meanwhile, because $a_1 \notin A^\times$, there must exists some irreducible element p_0 such that $p_0 \mid a_1$. In turn, as $a_1 \mid a_2 \mid \cdots \mid a_m$, we have that $\#\{i \;:\; \nu_{p_0}(a_i) > 0\} = m$. Hence, we've proven that $M = m$.

By similar reasoning with the quantity $\max\{\#\{i \;:\; \nu_p(b_i) > 0\} : p \in \mathcal{P}_0\}$, we can thus conclude that:

$$m = \max\{\#\{i \;:\; \nu_p(a_i) > 0\} : p \in \mathcal{P}_0\} = \max\{\#\{i \;:\; \nu_p(b_i) > 0\} : p \in \mathcal{P}_0\} = \ell$$

Next, for any fixed $p \in \mathcal{P}_0$ consider the Young diagram such that the number of boxes in the $(r+1)$th column is equal to $\#\{i \;:\; \nu_p(a_i) > r\} = \#\{i \;:\; \nu_p(b_i) > r\}$.

Then the number of boxes in the kth row is precisely equal to $\nu_p(a_{-k+m+1})$ and $\nu_p(b_{-k+m+1})$. Below I've attached a sample diagram to show what I mean.



This proves that $\nu_p(a_i) = \nu_p(b_i)$ for all $p \in \mathcal{P}_0$ and $i \in \{1, \dots, m\}$. The only way this is possible is if a_i, b_i are companions for each i. ■

Suppose M is an $F[x]$-module where F is a field. Then $\ell_x : M \to M$ defined by $\ell_x(m) = xm$ is an F-linear map. Also if $f(x) = c_n x^n + \ldots + c_1 x + c_0 \in F[x]$, then $f(x) \cdot m = \sum_{i=0}^{n} c_i \ell_x^i(m)$. Hence, the F-linear map ℓ_x uniquely determines the $F[x]$-module structure of M.

In a similar vein, if M_1, M_2 are $F[x]$-modules then $\theta : M_1 \to M_2$ is an $F[x]$-module homomorphism if and only if θ is an F-linear map, and $\theta(x \cdot m_1) = x \cdot \theta(m_1)$. To put another θ is an $F[x]$-module homomorphism iff if the following commutative diagram of F-module homomorphisms holds:

$$\begin{array}{ccc} M_1 & \xrightarrow{\theta} & M_2 \\ {\scriptstyle \ell_x}\downarrow & & \downarrow{\scriptstyle \ell_x} \\ M_1 & \xrightarrow{\theta} & M_2 \end{array}$$

$(\Longrightarrow)$
This is obvious.

$(\Longleftarrow)$
Suppose $f(x) = c_n x^n + \ldots + c_1 x + c_0$. Then:

$$\theta(f(x) \cdot m) = \theta(\sum_{i=0}^{n} c_i \ell_x^i(m)) = \sum_{i=0}^{n} c_i \theta(\ell_x^i(m)) = \sum_{i=0}^{n} c_i \ell_x^i(\theta(m)) = f(x) \cdot \theta(m).$$

We're going to use this to study linear algebra. Suppose F is a field and $a \in M_n(F)$. Then we are interested in the map $a : F^n \to F^n$ given by $|v\rangle \mapsto a\,|v\rangle$. So, we define the $F[x]$-

module V_a by taking the F-vector space F^n and defining $\ell_x(v) = a\,|v\rangle$. Hence:

$$(c_n x^n + \cdots + c_1 x + c_0) \cdot v = (c_n a^n + \cdots + c_1 a + c_0 I)|v\rangle$$

Lemma: $V_a \cong V_b$ as $F[x]$-modules if and only if a and b are similar matrices.

Proof:

$V_a \cong V_b$ as $F[x]$-modules iff there exists an $F[x]$ module isomorphism $\theta : V_a \to V_b$. But that happens iff there exists a bijective F-linear map $\theta : V_a \to V_b$ such that the below diagram commutes:

$$\begin{tikzcd} F^n \arrow[r, equal] \arrow[d, "a"] & V_a \arrow[r, "\theta"] \arrow[d, "\ell_x"] & V_b \arrow[r, equal] \arrow[d, "\ell_x"] & F^n \arrow[d, "b"] \\ F^n \arrow[r, equal] & V_a \arrow[r, "\theta"] & V_b \arrow[r, equal] & F^n \end{tikzcd}$$

As a side note, if $a \in M_{n_1}(F)$, $b \in M_{n_2}(F)$, and $n_1 \neq n_2$, then there doesn't exists an F-linear bijection $\theta : F^{n_1} \to F^{n_2}$. Hence, we lose nothing by assuming both V_a and V_b are equal as sets to F^n.

Equivalently, we can say there is some $g \in \mathrm{GL}_n(F)$ with $\theta(v) = g\,|v\rangle$ such that the below diagram commutes:

$$\begin{tikzcd} F^n \arrow[r, "g"] \arrow[d, "a"] & F^n \arrow[d, "b"] \\ F^n \arrow[r, "g"] & F^n \end{tikzcd}$$

In other words, there exists an invertible matrix g such that $ga = bg$. And finally, we get that $b = gag^{-1}$ for some $g \in \mathrm{GL}_n(F)$. Hence, a and b are similar matrices. ■

Suppose $a \in M_{n_1}(F)$ and $b \in M_{n_2}(F)$. Then if we consider the direct sum $V_a \oplus V_b$ note that for any $(v, w) \in V_a \oplus V_b$ we have that:

$$x \cdot (v, w) = (a\,|v\rangle, b\,|v\rangle) = \begin{bmatrix} a & \mathbf{0} \\ \mathbf{0} & b \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix}$$

In particular, if we denote the block matrix $\begin{bmatrix} a & \mathbf{0} \\ \mathbf{0} & b \end{bmatrix}$ as $\mathrm{diag}(a, b)$, then we've shown that $V_a \oplus V_b \cong V_{\mathrm{diag}(a,b)}$.

Hopefully it is clear how this generalizes to larger finite direct sums, and how one would define a block matrix $\mathrm{diag}(a_1, \ldots, a_n)$.

Given a field F, we know that $F[x]$ is a P.I.D. Also if $a \in M_n(F)$ then we trivially have that V_a is finitely generated (because the standard basis for F^n will always generate all of V_a). Therefore, by the *fundamental theorem of finitely generated modules over a P.I.D*, we know that there exists a unique $r \in \mathbb{Z}_{\geq 0}$ as well as unique ideals $\langle f_1 \rangle \supseteq \langle f_2 \rangle \supseteq \cdots \supseteq \langle f_m \rangle$ in $F[x]$ such that:

$$V_a \cong (F[x])^r \oplus \frac{F[x]}{\langle f_1 \rangle} \oplus \cdots \oplus \frac{F[x]}{\langle f_m \rangle} \text{ as } F[x]\text{-modules.}$$

Equivalently, there exists unique monic polynomials $f_1 \mid f_2 \mid \cdots \mid f_m$ in $F[x]$ such that:

$$V_a \cong (F[x])^r \oplus \frac{F[x]}{\langle f_1 \rangle} \oplus \cdots \oplus \frac{F[x]}{\langle f_m \rangle} \text{ as } F[x]\text{-modules.}$$

But now we can immediately deduce that $r = 0$.

To see why compare the dimensions of both sides of the above equation as F-vector spaces. V_a would be an n-dimensional F-vector but $F[x]^r$ would be an infinite-dimensional F-vector space for any $r > 0$.

So, we can further refine our conclusion to the following:

Corollary: Suppose F is a field and $a \in M_n(F)$. There exists unique monic positive degree polynomials $f_1 \mid f_2 \mid \cdots \mid f_m$ such that $V_a \cong \frac{F[x]}{\langle f_1 \rangle} \oplus \cdots \oplus \frac{F[x]}{\langle f_m \rangle}$ as $F[x]$-modules.

We call the polynomials $f_1, \ldots, f_m$ above the invariant factors of the matrix a.

Corollary: If F is a field and $a, b \in M_n(F)$, then a is similar to b iff a and b have the same invariant factors.

Proof:

a is similar to b iff $V_a \cong V_b$ as $F[x]$-modules. But by the *fundamental theorem of finitely generated modules over a P.I.D*, the latter statement is equivalent to both a and b having the same invariant factors. ■

Question? Given any $f \in F[x]$, can we find a matrix $a \in M_n(F)$ such that $V_a \cong \frac{F[x]}{\langle f \rangle}$.

A reason to ask this question is that after finding matrices $a_1, \ldots, a_m$ such that $V_{a_i} \cong \frac{F[x]}{\langle f_i \rangle}$ for each $i \in \{1, \ldots, m\}$, we could then conclude that:

$$V_a \cong \frac{F[x]}{\langle f_1 \rangle} \oplus \cdots \oplus \frac{F[x]}{\langle f_m \rangle} \cong V_{a_1} \oplus \cdots \oplus V_{a_m} \cong V_{\mathrm{diag}(a_1, \ldots, a_m)}$$

The answer to the above question is yes. To show this, first recall from math 100b and 100c that if $f(x) = c_n x^n + \cdots + c_1 x + c_0$ (where $c_n \neq 0$) then $\frac{F[x]}{\langle f \rangle}$ has the following F-basis (where $\overline{x^i}$ is just the equivalence class of $x^i \pmod{\langle f \rangle}$):

$$\mathcal{B} = (\overline{1}, \overline{x}, \ldots, \overline{x^{n-1}})$$

To see why, first note that by the long division theorem we can conclude that every equivalence class in $F[x]/\langle f \rangle$ contains a unique polynomial with degree less than n. It easily follows that $\mathcal{B}$ spans all of $F[x]/\langle f \rangle$. Also, as the unique polynomial of degree less than n equivalent to $0 \pmod{\langle f \rangle}$ is 0, we know that:

$$c_{n-1}\overline{x^{n-1}} + \cdots + c_1\overline{x^1} + c_0\overline{1} \equiv 0 \pmod{\langle f \rangle} \text{ iff all } c_i = 0.$$

Thus, we identify $\frac{F[x]}{\langle f \rangle} \to F^n$ as F-vector spaces via the mapping $\overline{g} \mapsto |\overline{g}\rangle_{\mathcal{B}}$ (where $|\overline{g}\rangle_{\mathcal{B}}$ is just the column vector for $\overline{g}$ with respect to the basis $\mathcal{B}$).

In other words, $a_{m-1}x^{m-1} + \cdots + a_1 x + a_0 + \langle f \rangle \mapsto |(a_0, a_1, \ldots, a_{m-1})\rangle$.

We want to find a matrix $[\ell_x]_{\mathcal{B}} \in M_n(F)$ such that $[\ell_x]_{\mathcal{B}}|\overline{g}\rangle_{\mathcal{B}} = |\ell_x(\overline{g})\rangle_{\mathcal{B}}$. In other words, we want the following diagram to commute:

$$\begin{tikzcd} \frac{F[x]}{\langle f \rangle} \arrow[r, "|\cdot\rangle_{\mathcal{B}}"] \arrow[d, "\ell_x"'] & F^n \arrow[d, "{[\ell_x]_{\mathcal{B}}}"] \\ \frac{F[x]}{\langle f \rangle} \arrow[r, "|\cdot\rangle_{\mathcal{B}}"'] & F^n \end{tikzcd}$$

But this is easy. After all, given any linear map $T : V \to V$ and basis vectors $v_1, \ldots, v_n$, we can always write the matrix of T with respect to that basis as $\begin{bmatrix} T(v_1) & \cdots & T(v_n) \end{bmatrix}$. Hence we must have that:

$$\begin{aligned} [\ell_x]_{\mathcal{B}} &= \begin{bmatrix} |\ell_x(\overline{1})\rangle_{\mathcal{B}} & |\ell_x(\overline{x})\rangle_{\mathcal{B}} & \cdots & |\ell_x(\overline{x^{n-2}})\rangle_{\mathcal{B}} & |\ell_x(\overline{x^{n-1}})\rangle_{\mathcal{B}} \end{bmatrix} \\ &= \begin{bmatrix} |\overline{x}\rangle_{\mathcal{B}} & |\overline{x^2}\rangle_{\mathcal{B}} & \cdots & |\overline{x^{n-1}}\rangle_{\mathcal{B}} & |\overline{x^n}\rangle_{\mathcal{B}} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & \cdots & 0 & -c_0/c_n \\ 1 & 0 & \cdots & 0 & -c_1/c_n \\ 0 & 1 & \cdots & 0 & -c_2/c_n \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -c_{n-1}/c_n \end{bmatrix} \end{aligned}$$

This matrix is called the companion matrix of $f(x)$ and we denote it by c_f. (Note that if f is monic then $c_n = 1$...)

Thus we can conclude that $\frac{F[x]}{\langle f \rangle} \cong V_{c_f}$. Also, this leads to the following theorem.

Theorem (Rational Canonical Form): If $a \in M_n(F)$ and $f_1, \ldots, f_m$ are the invariant factors of a then a is similar to $\operatorname{diag}(c_{f_1}, \ldots, c_{f_m})$.

(This is the beginning of the final lecture I need to get through to fully catch up to the class. Unfortunately I didn't have time to do the third problem set.)

Firstly, we prove a uniqueness result related to the prior theorem. Note that given a matrix of the form:

$$a = \begin{bmatrix} 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -c_{n-1} \end{bmatrix} \in M_n(F),$$

we can find a unique monic polynomial $f \in F[x]$ such that $a = c_f$.

Theorem (Uniqueness of the Rational Canonical Form): Suppose $f_1, \ldots, f_m \in F[x]$ are monic polynomials satisfing that $f_1 \mid f_2 \mid \cdots \mid f_m$ and the matrix a is similar to $\mathrm{diag}(c_{f_1}, \ldots, c_{f_m})$. Then $f_1, \ldots, f_m$ are the invariant factors of the matrix a.

Proof:
$V_a \cong V_{\mathrm{diag}}(c_{f_1}, \ldots, c_{f_m}) \cong \bigoplus_{i=1}^m V_{c_{f_i}} \cong \bigoplus_{i=1}^m \frac{F[x]}{\langle f_i \rangle}$. Then as $f_1 \mid f_2 \mid \cdots \mid f_m$, we can conclude by the uniqueness part of the fundamental theorem of finitely generated modules over a P.I.D. that $f_1, \ldots, f_m$ are the invariant polynomials. ■

On *page 518* I wrote what it means for a unital commutative ring to be an F-algebra. However, for this next section it will be necessary to define what it means for a non-commutative unit ring to be an F-algebra. Also in general I want to prove some things I didn't think of proving back on page 518.

If F is a field and A is a noncommutative unital ring, we say that A is an F-algebra if there exists a ring homomorphism $f : F \to A$ such that $f(1_f) = f(1_A)$ and $f(c)a = af(c)$ for all $a \in A$ and $c \in F$.

- Note that if $A \neq \{0\}$ then f is necessarily injective. This is because $\ker(f)$ is an ideal of F. That said, the only ideals of F are F and $\{0\}$, and we know that $\ker(f) \neq F$. After all, $1_F \notin \ker(f)$ since $f(1_F) = 1_A \neq 0_A$. It follows that $\ker(f) = \{0\}$ and that proves that f is injective.

  The significance of the prior result is that if A is an F-algebra then we can view F as being a field embedded into A.

- If $c \in F$ and $a \in A$ satisfy that $ca = 0$, then we must have that either $c = 0$ or $a = 0$. After all, suppose $c \neq 0$ but $ca = 0$. Then $a = c^{-1}ca = c^{-1}0 = 0$.

- I mentioned on page 518 that A is an F-vector space when we consider defining $c \cdot a = f(c)a$ for all $c \in F$ and $a \in A$. This is easy to show. That said, we do need to explicity assume that $f(c)a = af(c)$ for all $a \in A$ and $c \in F$ in order for us to have that $c \cdot (a_1 a_2) = (c \cdot a_1)a_2 = a_1(c \cdot a_2)$.

- For an example of a non-commutative F-algebra that we'll be caring about, consider the set of matrices $M_n(F)$ and define $f : F \to M_n(F)$ by $c \mapsto cI$.

Lemma: Suppose A is an F-algebra (and $A \neq \{0\}$). If $\dim_F(A) < \infty$ then for all $a \in A$ there exists a unique nonconstant monic polynomial $m_{a;F}(x) \in F[x]$ such that for all $g \in F[x]$ we have that $g(a) = 0 \Longleftrightarrow m_{a;F}(x) \mid g(x)$ in $F[x]$.

Note that $m_{a;F}$ is called the minimal polynomial of a over F.

Proof:
Let $e_a : F[x] \to A$ be the evaluation homomorphism given by $e_a(f(x)) := f(a)$. Then e_a is an F-algebra homomorphism (meaning it is a ring homomorphism from $F[x]$ into A and an F-linear map).

Note that e_a is a ring homomorphism specifically because $ca = ac$ for all $c \in F$ and $a \in A$. I won't prove that e_a is a homomorphism though.

We thus have that $\ker(e_a) \lhd F[x]$. Also, by the first isomorphism theorems for F-modules and rings, we have that $\frac{F[x]}{\ker(e_a)} \cong \mathrm{im}(e_a) \subseteq A$ as an F-algebra. Since $\dim_F(A) < \infty$ and $\dim_F(F[x]) = \infty$, we must have that $\ker(e_a) \neq \{0\}$. Hence as $F[x]$ is a P.I.D. and F is a field, there must exist a unique monic polynomial $m_{a;F}(x) \in F[x]$ such that:

$$\ker(e_a) = \langle m_{a;F} \rangle.$$

Also note that $\ker(e_a) \neq F[x]$ as the only constant polynomial contained in $\ker(e_a)$ is the zero polynomial. Therefore, we can conclude that $m_{a;F} \neq 1$. And finally, $g \in \ker(e_a)$ iff $m_{a;F} \mid g$ in $F[x]$. ■

For an application of the prior lemma, note that if $a \in M_n(F)$ then $\mathrm{Ann}(V_a) = \langle m_{a;F}(x) \rangle$.

Why?

$g(x) \cdot V_a = 0$ iff $\forall v \in F^n$, $g(a)|v\rangle = 0$. But the latter happens iff $g(a) = 0$, and that happens iff $m_{a;F}(x) \mid g(x)$.

Given a field F and $f \in F[x]$, we have that $\mathrm{Ann}(\frac{F[x]}{\langle f \rangle}) = \langle f \rangle$ (where we are viewing $\frac{F[x]}{\langle f \rangle}$ as an $F[x]$-module).

Why?

If $g \in \langle f \rangle$ then $g(x) = f(x)h(x)$. In turn, for any $s(x) + \langle f \rangle$ in $F[x]/\langle f \rangle$ we have that $g(x)(s(x) + \langle f \rangle) = f(x)h(x)s(x) + \langle f \rangle \equiv 0 \pmod{f(x)}$. And this proves that $\langle f \rangle \subseteq \mathrm{Ann}(\frac{F[x]}{\langle f \rangle})$. To show the other inclusion, suppose $g \in \mathrm{Ann}(\frac{F[x]}{\langle f \rangle})$. Then $g(x)\,(1 + \langle f \rangle) = g(x) + \langle f \rangle = 0 + \langle f \rangle$. The only way this is possible is if $g \in \langle f \rangle$.

More generally, the above reasoning shows that if A is a ring and $\mathfrak{a} \lhd A$, then $\mathrm{Ann}(\frac{A}{\mathfrak{a}}) = \mathfrak{a}$ (where we are viewing $A/\mathfrak{a}$ as an A-module).

Suppose $\{M_i\}_{i \in I}$ is a family of A-modules. Then $\mathrm{Ann}(\bigoplus_{i \in I} M_i) = \bigcap_{i \in I} \mathrm{Ann}(M_i)$.

Why?

If $a \in \mathrm{Ann}(\bigoplus_{i' \in I} M_{i'})$ and we consider the projection $P_i : \bigoplus_{i' \in I} M_{i'} \to M_i$, then we know that $a \cdot m_i = P_i(a \cdot (m_{i'})_{i' \in I}) = P_i((0)_{i' \in I}) = 0$. In particular, this shows that $a \cdot m_i = 0$ for all $i \in I$ and $m_i \in M_i$. Hence $a \in \bigcap_{i \in I} \mathrm{Ann}(M_i)$.

Conversely, if $a \in \bigcap_{i \in I} \mathrm{Ann}(M_i)$ then we know that $a \cdot m_i = 0$ for any $m_i \in M_i$ and $i \in I$. In turn, $a \cdot (m_{i'})_{i' \in I} = (a \cdot m_{i'})_{i' \in I} = 0$.

As a corollary, if F is a field and $f_1, \ldots, f_m \in F[x]$, then:

$$\mathrm{Ann}(\bigoplus_{i=1}^m \frac{F[x]}{\langle f_i \rangle}) = \bigcap_{i=1}^m \langle f_i \rangle = \langle \mathrm{lcm}(f_1, \ldots, f_m) \rangle.$$

To see what I specifically mean by a least common multiple, note that if D is a P.I.D. and $\langle d_1 \rangle, \ldots, \langle d_m \rangle$ are ideals in D, then there must exist some element $\ell \in D$ with $\langle \ell \rangle = \bigcap_{i=1}^m \langle d_i \rangle$. In turn, ℓ satisfies the property that $d_i \mid \ell$ for all i. Furthermore, if x satisfies that $d_i \mid x$ for all i, then $\ell \mid x$.

Proposition: Suppose F is a field. Then for all $a \in M_n(F)$ we have that $m_{a;F}(x) \in F[x]$ is the largest invariant factor of a.

Proof:

Suppose $f_1 \mid f_2 \mid \cdots \mid f_m$ are the invariant factors of a. Then, we know that:

$$\langle m_{a;F}(x) \rangle = \mathrm{Ann}(V_a) = \mathrm{Ann}(\tfrac{F[x]}{\langle f_1 \rangle} \oplus \cdots \oplus \tfrac{F[x]}{\langle f_m \rangle}) = \bigcap_{i=1}^m \langle f_i \rangle = \langle f_m \rangle$$

As a side note, if $M_1 \cong M_2$ as A-modules, then both $\mathrm{Ann}(M_1)$ and $\mathrm{Ann}(M_2)$ are equal subsets of A. In other words, it would be incorrect (or at least incredibly misleading) to write $\mathrm{Ann}(M_1) \cong \mathrm{Ann}(M_2)$ as opposed to $\mathrm{Ann}(M_1) = \mathrm{Ann}(M_2)$.

It follows that $m_{a;F}(x)$ and $f_m(x)$ are companion elements of $F[x]$. Yet as both are monic, the only way this is possible is if $m_{a;F}(x) = f_m(x)$. ■

The characteristic polynomial of a matrix $a \in M_n(A)$ (where A is commutative unital ring) is defined as the polynomial $\chi_a(t) := \det(tI - a)$ (where $tI - a \in M_n(A[t])$).

Lemma: If $a_1, \ldots, a_m$ are square matrices over some ring A, then:

$$\det(\mathrm{diag}(a_1, \ldots, a_m)) = \prod_{i=1}^m \det(a_i).$$

Proof:

By noting that $\mathrm{diag}(a_1, \ldots, a_{m-1}, a_m) = \mathrm{diag}(\mathrm{diag}(a_1, \ldots, a_{m-1}), a_m)$, it suffices to prove that $\det(\mathrm{diag}(a, b)) = \det(a)\det(b)$ where $a \in M_{n_1}(A)$ and $b \in M_{n_2}(A)$ for some integers n_1 and n_2.

Next, we proceed by induction on n_1. For our base case, note that if $n_1 = 1$ (so that a is just a scalar in A), then we can take the Laplace expansion of the determinant formula to get that:

$$\det(\mathrm{diag}(a, b)) = (-1)^{1+1} a \det(b) + \sum_{i=1}^{n_2} 0 = \det(a)\det(b).$$

As for the induction step, note again by the Laplace expansion formula that:

$$\begin{aligned}\det(\mathrm{diag}(a, b)) &= \textstyle\sum_{i=1}^{n_1} (-1)^{1+i} a_{1,i} \det(\mathrm{diag}(a(1,i), b)) \\ &= \textstyle\sum_{i=1}^{n_1} (-1)^{1+i} a_{1,i} \det(a(1,i)) \det(b) \\ &= (\textstyle\sum_{i=1}^{n_1} (-1)^{1+i} a_{1,i} \det(a(1,i))) \cdot \det(b) = \det(a)\det(b). \ ■\end{aligned}$$

Lemma: If $b = gag^{-1}$ where $a, b, g \in M_n(A)$, then $\chi_a = \chi_b$.

Proof:

$$\begin{aligned}\chi_b = \det(tI - b) = \det(bI - gag^{-1}) &= \det(g(tI - a)g^{-1}) \\ &= \det(g)\det(tI - a)\det(g^{-1}) = \det(tI - a) = \chi_a.\end{aligned}$$

Lemma: Suppose f is a monic nonconstant polynomial in $F[t]$ (where F is a field). If c_f is the companion matrix then $\chi_{c_f} = f$.

Proof:

If $f(t) = t^n + c_{n-1}t^{n-1} + \ldots + c_1 t + c_0$, then:

$$tI - c_f = \begin{bmatrix} t & 0 & \cdots & 0 & c_0 \\ -1 & t & \cdots & 0 & c_1 \\ 0 & -1 & \cdots & 0 & c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & t & c_{n-2} \\ 0 & 0 & \cdots & -1 & t + c_{n-1} \end{bmatrix}$$

By the laplace expansion formula, we get that:

$$\chi_{c_f}(t) = t\det\left(\begin{bmatrix} t & \cdots & 0 & c_1 \\ -1 & \cdots & 0 & c_2 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & t & c_{n-2} \\ 0 & \cdots & -1 & t + c_{n-1} \end{bmatrix}\right) + (-1)^{1+n}c_0\det\left(\begin{bmatrix} -1 & t & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t \\ 0 & 0 & \cdots & -1 \end{bmatrix}\right)$$

$$= t\chi_{c_{\left(\frac{f-c_0}{t}\right)}} + (-1)^{1+n}c_0(-1)^{n-1} = t\chi_{c_{\left(\frac{f-c_0}{t}\right)}} + c_0$$

And now it follows by doing induction on the degree of f that $\chi_{c_f} = f$.

Technically I still need to show a base case. Suppose $f(t) = t + c_0$. Then $tI - c_f = [t + c_0]$ and we trivially have that $\chi_{c_f}(t) = t + c_0 = f(t)$. ■

Corollary: If $f_1 \mid f_2 \mid \cdots \mid f_m$ are the invariant factors of a matrix $a \in M_n(F)$ then:
$m_{a;F}(x) = f_m(x)$ and $\chi_a(x) = \prod_{i=1}^m f_i(x)$.

Proof:
We already proved that $m_{a;F}(x) = f_m(x)$. To show the other equality, note that as a is similar to $\mathrm{diag}(c_{f_1}, \ldots, c_{f_m})$ we have that:

$$\chi_a = \chi_{\mathrm{diag}(c_{f_1},\ldots,c_{f_m})} = \prod_{i=1}^m \chi_{c_{f_i}} = \prod_{i=1}^m f_i. \blacksquare$$

(Note that $\chi_{\mathrm{diag}(c_{f_1},\ldots,c_{f_m})} = \prod_{i=1}^m \chi_{c_{f_i}}$ because of the first of our three prior lemmas plus the fact that $tI - \mathrm{diag}(c_{f_1}, \ldots, c_{f_m}) = \mathrm{diag}(tI - c_{f_1}, \ldots, tI - c_{f_m})$.)

Consequently, we get the following theorem:

Cayley Hamilton Theorem: Suppose F is a field and $a \in M_n(F)$. Then:

- $m_{a;F}(x) \mid \chi_a(x)$,
- If $p(x)$ is an irreducible factor of $\chi_a(x)$ then $p(x) \mid m_{a;F}(x)$.

Bibliography:

Everything is cited in the order it shows up in the journal with the exception that the first citation is for the book that got me to actually sit down and make a bibliography. Also, I decided to write my citations according to APA 7.

1. Hewitt, E., & Stromberg, K. (1975). *Real and Abstract Analysis*. Springer.
2. Rudin, W. (1976). *Principles of Mathematical Analysis* (3rd ed.). McGraw Hill.
3. Munkres, J. R. (2014). *Topology*. Pearson.
4. Artin, M. (2010). *Algebra* (2nd ed.). Pearson.
5. Folland, G. B. (1999). *Real Analysis : Modern Techniques and Their Applications* (2nd ed.). John Wiley And Sons.
6. Munkres, J. R. (1994). *Analysis on manifolds*. Addison-Wesley.
7. Guillemin, V., & Haine, P. (2019). *Differential Forms*. World Scientific Publishing Company.
8. Spencer, J. (1981). Balancing unit vectors. *Journal of Combinatorial Theory, Series A*, 30(3), 349–350. https://doi.org/10.1016/0097-3165(81)90033-9
9. Lee, J. M. (2013). *Introduction to smooth manifolds*. Springer.
10. Big O notation. (2025, September 9). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Big_O_notation&oldid=1310464597#Little-o_notation
11. Zimmer, R. J. (1990). *Essential Results of Functional Analysis*. University of Chicago Press.
12. Garrett, P. (2014, April 23). *Seminorms and locally convex spaces*. Retrieved September 26, 2025, from https://www-users.cse.umn.edu/~garrett/m/fun/notes_2012-13/07b_seminorms.pdf
13. Conway, J. B. (1978). *Functions of One Complex Variable* (2nd ed.). Springer.
14. Einsiedler, M., & Ward, T. (2010). *Ergodic Theory*. Springer.
15. Furstenberg, H. (1981). *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton University Press. http://www.jstor.org/stable/j.ctt7zv9zv
16. Dummit, D., & Foote, R. (2003). *Abstract Algebra* (3rd ed.). John Wiley And Sons.
17. Katz, J. (2017, December 16). *Brouwer's Fixed Point Theorem*. Retrieved October 17, 2025, from https://math.uchicago.edu/~may/REU2017/REUPapers/Katz.pdf
18. Howard, R. (2004, December). *The Milnor-Rogers Proof of the Brouwer Fixed Point Theorem*. Retrieved October 19, 2025, from https://people.math.sc.edu/howard/notes/brouwer.pdf
19. Rudin, W. (1991). *Functional Analysis* (2nd ed.). McGraw Hill.
20. (https://math.stackexchange.com/users/465145/david-gao), D. G. (n.d.). Proof that convergence in probability is not locally convex. Retrieved from https://math.stackexchange.com/q/4949800