18.701 Notes

Lecturer: Henry Cohn

Andrew Liu

Fall 2022

My notes for 18.701, "Algebra I". The instructor for this course was Henry Cohn (`https://cohn.mit.edu/`).

Last updated on Saturday 27th May, 2023.

Contents

1 September 7, 2022

1.1 Class Policy

Lecturer: Professor Henry Cohn, `cohn@mit.edu`. Ten problem sets, six short quizzes, and two exams. Lowest two problem set scores and lowest quiz score will be dropped. Grading breakdown:

- 40% problem sets, 8 at 5% each
- 20% quizzes, 5 at 4% each
- 40% exams, 2 at 20% each

The textbook for this class is *Algebra, 2nd Ed.* by Michael Artin. Throughout the course of the semester, we'll be covering topics in both **linear and abstract algebras**.

1.2 Basic Definitions

Definition 1.1
We'll start by introducing some notation for matrices and vectors.

- Given an $m \times n$ matrix A, we say that $A \in \mathbb{R}^{m\times n}$, and we notate A in a few different ways:

$$(A_{ij})_{1\le i\le n, 1\le j\le m} = \begin{array}{cc} & n \\ m & \begin{pmatrix} \quad & & \\ & A_{ij} & \\ & & \quad \end{pmatrix} \end{array} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{pmatrix}$$

- Similarly, for any n-dimensional vector x, we say that $x \in \mathbb{R}^n = \mathbb{R}^{n\times 1}$, and notate $x = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}^t$.
- Tensors are similar to matrices and vectors, but they're higher-dimensional equivalent. According to Prof. Cohn, they are "wildly unclassified".

Here is a list of things that matrices are useful for:

- Linear Transformations, i.e., mappings of the form $x \mapsto Ax$ over some vector space. Affine transformations are an extension of linear transformations by allowing a constant term, i.e., maps of the form $x \mapsto Ax + b$. We'll cover both in later lectures
- Bilinear Forms, which we'll cover in even later lectures
- An infinite number of other applications. We won't have time to cover everything during this semester

Per the first example, we say that matrices represent linear transformations because any $A \in \mathbb{R}^{m\times n}$ defines a transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$:

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \implies Ax = \begin{pmatrix} A_{11}x_1 + \ldots + A_{1n}x_n \\ \vdots \\ A_{m1}x_1 + \ldots + A_{mn}x_n \end{pmatrix} \in \mathbb{R}^m$$

The idea of an affine transformations is to additionally allow constant term translations after applying A, an idea that we'll cover in more detail later in 18.701.

Proposition 1.2

Let $A \in \mathbb{R}^{m\times n}$, and $B \in \mathbb{R}^{n\times p}$. Here are some key facts to remember about matrix multiplication.

- $AB \in \mathbb{R}^{m\times p}$
- $(AB)_{ik} = \sum_j A_{ij}B_{jk}$
- AB is a composition of linear transformations. For any $x \in \mathbb{R}^p$, $Bx \in \mathbb{R}^n$, and $A(Bx) \in \mathbb{R}^m$, but also $(AB)x \in \mathbb{R}^m$. It can be checked that composing $A \circ B$ is the same linear transformation as AB.
- Given another compatible matrix C, it is always true that $(AB)C = A(BC)$, which generalizes the idea in the last bullet point. In other words, matrix multiplication is associative.
- On the other hand, it is not usually true that matrix multiplication is commutative, i.e., that $AB = BA$.

The following is not a theorem, but is an important fact, so I'll put it in a theorem box for emphasis.

Theorem 1.3
Studying linear transformations and matrices in general should be viewed as a one-to-one relationship. In general, the intuition from one will almost always lead to intuition in the other, and studying one without the other lends itself to a poorer overall understanding of both. Linear transformations aren't the only application of matrices, but it is a very important one, and this perspective will more or less drive the way that we study matrices in this class.

1.3 Groups

Definition 1.4
A **group** G is a set G with binary operator $*$ satisfying three conditions:

- Associativity: $(fg)h = f(gh)$ for all $f, g, h \in G$.
- Identity: $\exists I \in G$ s.t. $gI = Ig = g \quad \forall g \in G$.
- Inverse: $\forall g \in G, \exists h \in G$ s.t. $gh = hg = 1$, $h = g^{-1}$.

The way we write the binary operator does not matter (e.g., $*$, $\cdot$, $\circ$, etc.) so long as we don't mix binary operators between different groups.

The point of groups is that they "act" on stuff. More on this in later lectures.

Example 1.5
Let's look at some examples and non-examples of groups.

- $(\mathbb{R}, +)$ is a group. On the other hand, $(\mathbb{R}, \times)$ is not a group, since 0 has no inverse.
- $\mathrm{GL}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n}, A \text{ invertible}\}$, with the operator of normal matrix multiplication, is a group. This group of matrices is called the **general linear group**.

- $\mathrm{SL}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n}, \det A = 1\}$ is called the **special linear group**. This group of matrices preserves volume and orientation (fixed determinant).
- S_n is the group of permutations of order n, with the operator of composition. This group is called the **symmetric group**. For instance, an element of S_6 might swap 1 and 3, fix 4, and cycle $2, 5, 6$, which we would notate $(13)(256)$.

2 September 9, 2022

2.1 Groups

Quick review of group definitions: a group G with binary operator satisfies:

- Associative: $(fg)h = f(gh) \quad \forall f, g, h \in G$
- Identities: $\exists 1 \in G$ s.t. $1g = g1 = g \quad \forall g \in G$
- Inverses: every element has an inverse.

Proposition 2.1

The group identity is unique. So are inverses.

Proof. If 1, $1'$ are both identities, then $1 = 11' = 1'$. Since the identity is unique, inverses are also necessarily unique. □

Definition 2.2

There is one group with one element (only the identity). We call this group the **trivial group**.

Definition 2.3

$C_n = \{1, g, \dots, g^{n-1}\}$, with $g^n = 1$, is the **cyclic group** of order n.

In general, there are lots of ways to represent different groups. For example, here are two multiplication tables for C_3:

$\cdot$	1	g	g^2
1	1	g	g^2
g	g	g^2	g
g^2	g^2	1	g

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	1	0

These different representations are **isomorphic**. Also, note the symmetry along the diagonal of each multiplication table. Each element in C_3 commutes, so we say that it is **abelian** (commutative). This property cannot be assumed to hold for general groups.

2.2 Classifying small groups

Let G be a group with $|G| = n$, that is, the **order** of G is n. Let's classify groups when n is small.

Example 2.4

$n = 1$.

There is only one possibility here, the trivial group.

Example 2.5

$n = 2$.

Let's try filling in a multiplication table and see what possibilities there are.

$\cdot$	1	g
1	1	g
g	g	?

The last element must be 1. If not, then g does not have an inverse, which violates the group laws. So, there is only one group with order 2, which is C_2.

Example 2.6

$n = 3$.

Let's use the same approach.

$\cdot$	1	g	h
1	1	g	h
g	g		
h	h		

Those are our freebies. We can use a bit of logic to deduce the rest of the table. We know $g^2 \in \{1, g, h\}$. We can't have $g^2 = 1$, otherwise $gh = h \implies g = 1$. We also can't have $g^2 = g \implies g = 1$. So $g^2 = h$, and the rest of the table falls through. In this case, there is also only one group of order 3, which is C_3.

Example 2.7

$n = 4$.

This time, it turns out there are two groups. C_4 works as usual, but we also get another group, $C_2 \times C_2$.

Definition 2.8

Given two groups $(G, \cdot)$ and $(H, \star)$, their **group product** is defined as the set

$$G \times H = \{(g, h) : g \in G, h \in H\},$$

with the operation $(g, h)(g', h') = (g \cdot g', h \star h')$.

Since all multiplications are distinct, its obvious why group products of two groups must itself also be a group.

Instead of the traditional "multiplication", sometimes we'll write abelian groups additively; for example, $C_n = \{0, 1, \ldots, n-1\}$ with addition modulo n. In this notation, here's the multiplication table for $C_2 \times C_2$:

$+$	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,0)$	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,1)$	$(0,1)$	$(0,0)$	$(1,1)$	$(1,0)$
$(1,0)$	$(1,0)$	$(1,1)$	$(0,0)$	$(0,1)$
$(1,1)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$

We know that $C_2 \times C_2$ and C_4 aren't isomorphic because all elements in $C_2 \times$

C_2 square to the identity, which naturally follows from the definition of group product.

Example 2.9

$n = 5$.

As before, the only possible group is C_5. In general, for any prime p, there is only one group with order p, which is C_p.

Example 2.10

$n = 6$.

There are also two groups for $n = 6$, C_6 and S_3. This is the first time that we have a non-abelian group.

Recall that S_n is the symmetric group of permutations. A permutation is a one-to-one correspondence from $\{1,\ldots,n\}$ to $\{1,\ldots,n\}$. The group operation for S_n is composition. Here's a concrete example for S_5:

i	1	2	3	4	5
$\pi(i)$	3	5	1	2	4
$\pi'(i)$	1	2	5	3	4
$\pi\pi'(i)$	3	5	4	1	2

2.3 Infinite Groups

Some groups have infinite order. For example, we introduced last lecture the general linear group:

$$\mathrm{GL}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n} : A \text{ invertible}\}$$

$\mathrm{GL}_n(\mathbb{R})$ is closed under matrix multiplication. Given any $A, B \in \mathrm{GL}_n(\mathbb{R})$, $(AB)^{-1} = B^{-1}A^{-1}$, since $(AB)(B^{-1}A^{-1}) = AA^{-1} = I_n$, so AB is also invertible and in the group. You can also argue this using the fact that a matrix is invertible if and only if its determinant is non-zero.

Definition 2.11

H is a **subgroup** of G if $H \subseteq G$ and H forms a group itself under G's operation with the same identity, inverses, etc.

Example 2.12
Here are some examples of common subgroups.

- The trivial group is a subgroup of any group.
- $(\mathbb{Z}, +)$ is a subgroup of $(\mathbb{R}, +)$
- $\mathrm{SL}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n} : \det A = 1\}$ is a subgroup of $\mathrm{GL}_n(\mathbb{R})$.
- The **special orthogonal group** of order 2 is defined as the set of two-dimensional rotation matrices:

$$SO_2(\mathbb{R}) = \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R} \right\}$$

  Since these matrices always have determinant 1, we have the chain of subgroups

$$SO_2(\mathbb{R}) \subseteq SL_2(\mathbb{R}) \subseteq GL_2(\mathbb{R}).$$

2.4 Symmetry Groups

Definition 2.13
Let $S \subseteq \mathbb{R}^n$. Then, the **symmetry group** of S is equal to the set of all rigid motions T of $\mathbb{R}^n$ such that $\{Tx : x \in S\} = S$.

For instance, an equilateral triangle has symmetries isomorphic to S_3. The isomorphism arises when you label the vertices of the triangle $1, 2, 3$; then, each rigid motions directly maps to a different permutation. It turns out that this correspondence is only true for $n = 3$, and not true in general.

Definition 2.14
An **isomorphism** $f : G \to H$ between the groups $(G, \cdot)$ and $(H, \star)$ is a one-to-one correspondence between the elements of G and H such that $f(g_1 \cdot g_2) = f(g_1) \star f(g_2)$ for all $g_1, g_2 \in G$.

When G and H are isomorphic, we write $G \cong H$.

3 September 12, 2022

3.1 Subgroups

Let G be a group, and $S \subseteq G$ any subset of G.

Definition 3.1
The subgroup of G generated by S is equal to the intersection of all subgroups of G containing S.

Example 3.2
For any $g \in G$, when $S = \{g\}$, this subgroup is $\langle g \rangle = \{g^n : n \in \mathbb{Z}\}$.

If $|\langle g \rangle|$ is infinite, then we say that the **order** of g is infinite. In this case, the subgroup is isomorphic to $\mathbb{Z}$ (the infinite cyclic group). Otherwise, the order of g is some finite k, meaning that k is the minimum power of g that makes it equal to the identity, and $\langle g \rangle \cong C_k$.

3.2 Subgroups of $\mathbb{Z}$

For each $d \in \mathbb{Z}$, $\langle d \rangle = d\mathbb{Z}$ is a subgroup of $\mathbb{Z}$. This turns out to be the only possible type of subgroup.

$$
\begin{tikzcd}[row sep=tiny, column sep=tiny]
& 4\mathbb{Z} \arrow[dr, phantom, sloped, "\subseteq"] & & \\
& & 2\mathbb{Z} \arrow[dr, phantom, sloped, "\subseteq"] & \\
\cdots & 6\mathbb{Z} \arrow[ur, phantom, sloped, "\subseteq"] \arrow[dr, phantom, sloped, "\subseteq"] & & \mathbb{Z} \\
& & 3\mathbb{Z} \arrow[ur, phantom, sloped, "\subseteq"] & \\
& 9\mathbb{Z} \arrow[ur, phantom, sloped, "\subseteq"] & &
\end{tikzcd}
$$

Theorem 3.3
Every subgroup of $\mathbb{Z}$ has a single generator.

Proof. Let H be a subgroup of $\mathbb{Z}$. If $H = \{0\}$, then $H = 0\mathbb{Z}$. Otherwise, H contains some positive integer.

Let d be the smallest positive integer in H, which exists by the well-ordering principle. For any $n \in H$, the remainder theorem implies that there exists $q, r \in \mathbb{Z}$ such that $n = dq + r$, with $0 \le r < d$. Since $d \in H$, and $n \in H$, $r = n - dq \in H$. But d is the smallest positive integer in H, so $r = 0 \implies n \in d\mathbb{Z} \implies H \subseteq d\mathbb{Z}$. We also know $d\mathbb{Z} \subseteq H$, since $d \in H$, thus $H = d\mathbb{Z}$ has a single generator, as desired. $\square$

3.3 Homomorphisms

Definition 3.4

Given two groups $(G, \cdot)$ and $(H, \star)$, a **homomorphism** from G to H is a map $f : G \to H$ satisfying $f(g \cdot g') = f(g) \star f(g')$ for all $g, g' \in G$.

This definition matches the definition for an isomorphism, with the exception that our map does not need to be one-to-one. In other words, an isomorphism is a one-to-one homomorphism.

Proposition 3.5

$f(1_G) = 1_H$.

Proof. $f(1_G) \star f(1_G) = f(1_G) \implies f(1_G) = 1_H$ by multiplying by the inverse of $f(1_G)$ on both sides. $\square$

Proposition 3.6

$f(g^{-1}) = f(g)^{-1}$ in H.

Proof. Since the identity is preserved under our map, $f(g)f(g^{-1}) = f(gg^{-1}) = 1_H \implies f(g^{-1}) = f(g)^{-1}$. $\square$

Now, let's look at some examples of homomorphisms.

Example 3.7

Consider any subgroup $H \subseteq G$.

The identity map $f : H \to G$ is a homomorphism, so homomorphisms generalize subgroups. f is an "inclusion map".

Example 3.8

$\mathbb{Z}$ and C_n.

There is a homomorphism $\mathbb{Z} \to C_n$ taking $k \mapsto k \pmod{n}$, also called reduction mod n. In this lens, homomorphisms also generalize modulo arithmetic.

Example 3.9

S_n and $\mathrm{GL}_n(\mathbb{R})$.

Verify that the map $S_n \to \mathrm{GL}_n(\mathbb{R})$ with $\pi \mapsto$ (permutation matrix of π) is a homomorphism.

Example 3.10

The determinant is a homomorphism.

$\det : \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^x$ with $A \mapsto \det A$ is a homomorphism. (Recall that $\mathbb{R}^x$ is the set of non-zero real numbers, which is a group).

Example 3.11

The sign homomorphism.

The sign homomorphism is defined as the map $\mathrm{sgn} : S_n \to \{\pm 1\}$ with $\mathrm{sgn}(\pi) = \det(\text{perm. matrix})$. This is a homomorphism, since

$$\det(A_\pi)\det(A_{\pi^{-1}}) = 1,$$

so all determinants are ± 1.

Example 3.12

Exponentiation is a homomorphism.

The map $\exp : \mathbb{C} \to \mathbb{C}^x$ with $\exp z = e^z$ is a homomorphism (here we take $\mathbb{C}$ over addition and $\mathbb{C}^x$ over multiplication). exp is not isomorphic, since $e^{2\pi i} = e^0 = 1$ (the map is not injective).

4 September 14, 2022

4.1 Kernels and images

Every homomorphism $f : G \to H$ has two subgroups associated with it:

Definition 4.1
The **kernel** of f is
$$\ker(f) = \{g \in G : f(g) = 1\}.$$

Definition 4.2
The **image** of f is
$$\mathrm{im}(f) = \{h \in H : \exists g \in G \text{ s.t. } f(g) = h\}$$

Unlike isomorphisms, homomorphisms lose information. The kernel can roughly be thought of as what information we lost under the mapping. The image is how much of H we're actually hitting.

Proposition 4.3
f is injective if and only if $\ker(f) = \{1\}$.

Proof.
$$f(g) = f(g') \iff f(gg'^{-1}) = 1 \iff gg'^{-1} \in \ker(f).$$

If f is injective, then $g = g'$, so $\ker(f) = \{1\}$. On the other hand, if the kernel was trivial, then $g = g'$, so f is injective. $\square$

The above proposition shows that f is an isomorphism if and only if $\ker(f) = \{1\}$ and $\mathrm{im}(f) = H$. Let's look at some examples of kernels and images for specific groups.

Example 4.4
$$\exp : \mathbb{C} \to \mathbb{C}^x.$$

$\mathrm{im}(\exp) = \mathbb{C}^x$, and $\ker(\exp) = 2\pi i\mathbb{Z}$.

Example 4.5

$$\det : \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^x.$$

$\mathrm{im}(\det) = \mathbb{R}^x$ (recall the general linear group only contains invertible matrices, which is what allows it to be a group in the first place). $\ker(\det) = \mathrm{SL}_n(\mathbb{R})$.

Example 4.6

$$\mathrm{sgn} : S_n \to \pm 1.$$

$\mathrm{im}(\mathrm{sgn}) = \{\pm 1\}$, and $\ker(\mathrm{sgn}) = A_n$. This is called the **alternating group**, and contains all permutations with an even number of swaps (transpositions). We'll deal more with this group later in the semester.

Example 4.7

Given any group G and $g \in G$, define the mapping $f : \mathbb{Z} \to G$ by $f(n) = g^n$.

It is easy to show that f is a homomorphism. $\mathrm{im}(f) = \langle g \rangle$, or the subgroup generated by g. $\ker(f) = k\mathbb{Z}$ if g has finite order k. Otherwise, $\ker(f) = \{0\}$.

4.2 Cosets

Consider $G = \mathbb{Z}$. For any $k\mathbb{Z} \subseteq \mathbb{Z}$, there is a nice way to partition G with different translations $k\mathbb{Z}$:

$$\begin{aligned}\mathbb{Z} &= (2\mathbb{Z}) \cup (1 + 2\mathbb{Z}),\\ \mathbb{Z} &= (3\mathbb{Z}) \cup (1 + 3\mathbb{Z}) \cup (2 + 3\mathbb{Z}),\\ &\text{etc.}\end{aligned}$$

This generalizes for non-abelian groups.

Definition 4.8

For any group G and subgroup H, a **left coset** of H in G is a set

$$gH = \{gh : h \in H\}$$

with $g \in G$.

Right cosets are defined analogously. **In general, cosets are not subgroups**. This is because gH does not contain the identity unless $g \in H$ (in which case $g^{-1} \in H$, so $gg^{-1} = 1$).

We'll work with left cosets, but all of the following results also hold for right cosets. It turns out the cosets partition groups in exactly the way that we were looking for earlier.

Proposition 4.9

All cosets have the same size.

Proof. There is a one-to-one correspondence between H and gH; namely, multiply all elements in H by g to get gH, and multiply all elements in gH by g^{-1} to get H, so every coset has the same size as H itself. □

Proposition 4.10

Every element of G is in some coset.

Proof. $g \in gH$, since $1 \in H$. □

Proposition 4.11

If $gH \cap g'H \neq \emptyset$, then $gH = g'H$.

Proof. If $gh = g'h'$ for some $h, h' \in H$, then $g = g'h'h^{-1} \implies gH = g'h'h^{-1}H = g'H$. □

Together, the above three propositions show that we can partition G into cosets of equal size.

4.3 Lagrange's Theorem

Definition 4.12

For any group G and subgroup H, we define the **index** of H in G as

$$[G : H] = \text{left cosets of } H \text{ in } G.$$

The index can be infinite. For example $[2\mathbb{Z} : 6\mathbb{Z}] = 3$, while $[\mathbb{Z} : \{0\}] = \infty$.

Theorem 4.13 (Lagrange's Theorem)

$$|G| = |H| \cdot [G : H].$$

This encapsulates the partition idea that we uncovered with respect to the cosets of H in G. An important thing to note here is that whenever H is a subgroup of G, $|H|$ divides $|G|$, which allows us to confirm one of the observations that we made in Lecture 2:

Corollary 4.14
If $|G| = p$ with p prime, then $G \cong C_p$.

Proof. Let $g \in G$, with $g \neq 1$. Then $|\langle g\rangle| > 1$, so $|\langle g\rangle| = p$ by Lagrange. This shows that G has one generator, so it must be C_p. □

Back to the idea of the kernel representing what information is lost under any homomorphism. Why? Consider a homomorphism $f : G \to H$, with $\ker(f) = K$. Then,

$$f(g) = f(g') \iff f(g^{-1}g') = 1 \iff g^{-1}g' \in K \iff g^{-1}g'K = K \iff g'K = gK,$$

so elements in the same coset of K in G map to the same thing. When K is large, lots of elements map to the same thing, so information is lost, and vice versa. In fact, when $|K| = 1$, the mapping is injective (Proposition 4.3), so no information is lost.

add diagram

5 September 16, 2022

5.1 Normal Subgroups

Recall from last lecture:

Definition 5.1
Given a group G and a subgroup H, a left coset of H in G is any

$$gH = \{gh : h \in H\}$$

with $g \in G$.

These cosets partition G into subsets of the same size. By Lagrange's Theorem:

$$|G| = |H|[G : H].$$

Right cosets $Hg = \{hg : h \in H\}$ behave the same way. An easy way to show that this is true, i.e., that every right and left coset has the same number of elements, is to see that $(gH)^{-1} = H^{-1}g^{-1} = Hg^{-1}$, which works since H is closed under inverses.

How do the partitions differ between right and left cosets?

Example 5.2
Let $G = S_3$, and $H = \{1, (123), (132)\}$.

In this example,

$$(12)H = \{(12), (23), (13)\} = H(12),$$

so the right and left cosets are the same. It can be shown that $gH = Hg$ for all $g \in G$, so the right and left cosets make the same partition. This is not typical. It is also worth mentioning that H is the kernel of the sign homomorphism $\operatorname{sgn} : G \to \{\pm 1\}$.

Example 5.3
Let $G = S_3$, and $H = \{1, (12)\}$.

In this example,

$$(13)H = \{(13), (123)\} \qquad H(13) = \{(13), (132)\}$$
$$(23)H = \{(23), (132)\} \qquad H(23) = \{(23), (123)\},$$

so the partitions aren't the same. This is typical.

Suppose we have a homomorphism $f : G \to H$. We can prove that the observation we made in Example 5.2 is not a coincidence. Let $K = \ker(f)$. Then, for all

$g \in G$, $gK = \{gk : k \in K\} = \{g' \in G : f(g') = f(g)\}$ (we proved the second equality last lecture). Similarly, $Kg = \{kg : k \in K\} = \{g' \in G : f(g') = f(g)\}$, so $gK = Kg$ always.

Definition 5.4

H is a **normal subgroup** of G if H is a subgroup of G and $gH = Hg$ for all $g \in G$.

A few things to note:

- $gH = Hg \iff gHg^{-1} = H$. Subgroups of this form are called **conjugate subgroups**, and these are always subgroups.
- $gH = Hg \not\iff gh = hg \forall h \in H$. Rather, $\{gh : h \in H\} = \{hg : h \in H\}$.
- "normal" does not mean typical. In fact, normal subgroups are not typical at all. The vast majority of subgroups won't be normal (the naming is counter-intuitive).

Now we can prove what we proved above more succinctly:

Proposition 5.5

Let $f : G \to H$ be a homomorphism with kernel K. Then K is a normal subgroup in G.

Proof. For all $k \in K$, $f(g^{-1}kg) = f(g^{-1})f(k)f(g) = 1$, so $g^{-1}kg \in K \iff g^{-1}Kg \subseteq K$. Substituting g^{-1} for g further gives $gKg^{-1} \subseteq K$, so $gKg^{-1} = K$ and the result follows. □

The converse is also true!

5.2 Correspondence Theorem

Theorem 5.6 (Correspondence Theorem)
Let $f : G \to G'$ be a surjective homomorphism, i.e., $\mathrm{im}(f) = G'$. Then, there exists a bijection between the set of all subgroups H' of G', and the set of all subgroups H of G with $K \subseteq H$.
The bijection is given by

$$\begin{aligned}\text{forward direction: } & H \mapsto f(H) = \{f(h) : h \in H\}\\ \text{reverse direction: } & H' \mapsto f^{-1}(H') = \{g \in G : f(g) \in H'\}.\end{aligned}$$

Let's look at some examples.

Example 5.7
$\det : GL_n(\mathbb{R}) \to \mathbb{R}^\times$.

$\ker(\det) = \mathrm{SL}_n(\mathbb{R})$. By the correspondence theorem, subgroups of $\mathrm{GL}_n(\mathbb{R})$ that contains $\mathrm{SL}_n(\mathbb{R})$ correspond with subgroups of $\mathbb{R}^\times$. For example, $\{A \in \mathrm{GL}_n(\mathbb{R}) : \det A \in \mathbb{Q}\} \subseteq \mathrm{GL}_n(\mathbb{R})$ can be seen as corresponding to $\mathbb{Q}^\times \subseteq \mathbb{R}^\times$.

Example 5.8
$f : \mathbb{Z} \to C_n$ with $1 \mapsto$ generator.

$\ker(f) = n\mathbb{Z}$, so the correspondence theorem gives us

$$\begin{aligned}\text{subgroups of } C_n &\iff \text{subgroups } d\mathbb{Z} \text{ of } \mathbb{Z} \text{ that contain } n\mathbb{Z}\\ &\iff \text{positive integers } d \text{ s.t. } d \mid n.\end{aligned}$$

This intuitively makes sense, since subgroups of C_n are also cyclic, and choosing a subgroup amounts to choosing the new order (which divides n) of the subgroup.

6 September 19, 2022

6.1 Quotient Groups

Recall from last lecture: given group G and subgroup $H \subseteq G$,

$$H \text{ normal} \iff gH = Hg \quad \forall g \iff gHg^{-1} = H \quad \forall g \iff ghg^{-1} \subseteq H \quad \forall g, h.$$

The last equivalence follows, because the last statement implies $gH \subseteq Hg$, and you can flip g and g^{-1} to get $gH \supseteq Hg$, so $gH = Hg$.

We showed last time that all kernels are normal. Most subgroups aren't normal. Given H normal, does there exist a homomorphism $f : G \to G'$ such that $\ker(f) = H$? The answer is yes, which we'll show today.

Claim 6.1
If f is surjective, then G' is uniquely determined by G, H up to isomorphism.

The idea is that G' is determined by left cosets of H in G. If H is the kernel, then gH is the preimage of $f(g)$ in G. In other words, each coset corresponds to a unique element of $\text{im}(f)$.

Here's how we can turn cosets into a group.

Definition 6.2
Given cosets C_1 and C_2, let their product be

$$C_1C_2 = \{c_1c_2 : c_1 \in C_1, c_2 \in C_2\}.$$

This way of multiplying cosets would be terrible if H was not normal. Given that H is normal, then

$$g_1H \cdot g_2H = g_1(Hg_2)H = g_1g_2H,$$

which is also a coset. Note that the second equality only works when H is normal. Since the cosets are closed under this multiplication, we can treat them like a group.

Definition 6.3

If G is a group with normal subgroup H, the **quotient group**

$$G/H = \{gH : g \in G\},$$

with coset multiplication.

This is a group:

- Closure, since H is normal
- Associative, since $(g_1 H g_2 H) g_3 H = (g_1 g_2 g_3) H = g_1 H (g_2 H g_3 H)$.
- Identity H
- Inverse, $(gH)^{-1} = g^{-1} H$.

Definition 6.4

The **canonical map** $\pi : G \to G/H$ given by $g \mapsto gH$.

Theorem 6.5

The canonical map π is a surjective homomorphism with $\ker(\pi) = H$.

Proof. By definition, π is surjective. π is also a homomorphism:

$$\pi(g_1 g_2) = g_1 g_2 (H) = g_1 H g_2 H = \pi(g_1)\pi(g_2).$$

Now $\pi(g) = H \iff gH = H \iff g \in H$, so $\ker(\pi) = H$. $\square$

6.2 First Isomorphism Theorem

Theorem 6.6 (First Isomorphism Theorem)

Given a surjective homomorphism $f : G \to G'$ with $\ker(f) = H$, there exists a unique isomorphism $\overline{f} : G/H \to G'$ s.t. $f = \overline{f} \circ \pi$.

In other words, this diagram:

$$\begin{tikzcd} G \arrow[rr, "f"] \arrow[dr, "\pi"'] & & G' \\ & G/H \arrow[ur, "\overline{f}"'] & \end{tikzcd}$$

is a commutative diagram (different paths in the diagram leads to the same mapping).

Proof. $\overline{f}$ is uniquely determined by $\overline{f}(gH) = f(g)$. This implies that $\overline{f}$ is well-defined and injective, since $f(g_1H) = f(g_2H) \iff f(g_1) = f(g_2) \iff g_1H = g_2H$ ($H = \ker(f)$). $\overline{f}$ is surjective since f is surjective.

Also, $\overline{f}$ is a homomorphism since f is a homomorphism:

$$\overline{f}(g_1Hg_2H) = f(g_1g_2) = f(g_1)f(g_2) = \overline{f}(g_1H)\overline{f}(g_2H),$$

so $\overline{f}$ is an isomorphism and we're done. □

This theorem gives us a nice way to break down any homomorphism $f : G \to H$ that exists between arbitrary groups G, H. Specifically, f is surjective when you restrict the mapping to $\mathrm{im}(f)$, so f is the composition of three mappings:

- quotient $\pi : G \to G/\ker(f)$
- isomorphism $\overline{f} : G/\ker(f) \to \mathrm{im}(f)$
- inclusion $\mathrm{im}(f) \to H$,

where the inclusion just maps everything to itself and takes care of the fact that the codomain of the original mapping was to all of H.

$$\begin{tikzcd} & G \arrow[rr, "f"] \arrow[dl, "\pi"'] \arrow[dr, "f"'] & & H \\ G/H \arrow[rr, "\overline{f}"'] & & \mathrm{im}(f) \arrow[ur, "\mathrm{inc}"'] & \end{tikzcd}$$

Here are some examples of the first isomorphism theorem giving useful equivalencies for common groups.

Example 6.7
$\mathbb{Z}/n\mathbb{Z} \cong C_n$.

Recall from last lecture the homomorphism $f : \mathbb{Z} \to C_n$ mapping 1 to any generator. $\ker(f) = n\mathbb{Z}$, hence the correspondence between the cosets and C_n.

Example 6.8
$GL_n(\mathbb{R})/SL_n(\mathbb{R}) \cong \mathbb{R}^x$.

The homomorphism $\det : GL_n(\mathbb{R}) \to \mathbb{R}^x$ has kernel $\mathrm{SL}_n(\mathbb{R})$.

Example 6.9
$S_n/A_n \cong \{\pm 1\} \cong C_2$.

7 September 21, 2022

7.1 Conjugation in S_n

Given $\pi \in S_n$, the conjugate of any elt wrt π is given by

$$\pi(a_1 a_2 \dots a_k)\pi^{-1} = (\pi(a_1)\pi(a_2)\dots\pi(a_k)).$$

This can be proved by considering the movement of any element. Suppose for some a_i that $\pi^{-1}(a_i) = a_j$. Then the left hand side maps $a_i \mapsto a_j \mapsto a_{j+1} \mapsto \pi(a_{j+1})$. On the other hand, since $\pi(a_j) = a_i$, the right hand side maps $a_i \mapsto \pi(a_{j+1})$, so they map to the same thing.

7.2 Fields and Vector Spaces

We started with group theory, now we'll do some linear algebra before combining them again later.

Definition 7.1
A **field** is a set F with binary operators $+, \cdot$ such that

1. $(F, +)$ is an abelian group
2. $(F \backslash \{0\}, \cdot)$ is an abelian group
3. $x(y+z) = xy + xz$

We need the third condition (distributive law) in order to define how our two binary operators interact.

Example 7.2
Here are some examples of fields.

$\mathbb{R}$, $\mathbb{Q}$, and $\mathbb{C}$ are all fields. Note the exclusion of 0 when we're dealing with multiplication, which makes them all valid abelian groups. $\mathbb{Z}/p\mathbb{Z} = \{0, 1, \ldots, p-1\}$ is also a field, where we define $+, \cdot$ modulo p.

Example 7.3
Here are some non-examples of fields.

$\mathbb{Z}$ is not a field, since we can't do division (i.e., inverses) in $\mathbb{Z}$. In general, if you can't do division, the set that you're working with can't be a field. Unlike the previous example with prime order, $\mathbb{Z}/6\mathbb{Z}$ is not a field, since 2 and 3 do not have multiplicative inverses.

Lemma 7.4
Multiplicative inverses for elements in $\mathbb{Z}/p\mathbb{Z}^x$ exist.

Proof. Given $a \not\equiv 0 \pmod{p}$, we want to show the existence of some x such that $ax \equiv 1 \pmod{p}$. Consider the subgroup of $(\mathbb{Z}, +)$ generated by a, p:

$$a\mathbb{Z} + p\mathbb{Z} = \{ax + py : x, y \in \mathbb{Z}\} = d\mathbb{Z},$$

since every subgroup in $\mathbb{Z}$ has a single generator. a, p must be multiples of d, but $\gcd(a, p) = 1 \implies d = \pm 1$. Taking $d = 1$, we conclude that $ax + py = 1$ for some $x, y \in \mathbb{Z}$, so $ax \equiv 1 \pmod{p}$ has a solution. $\square$

Definition 7.5

A **vector space** V over field F is defined as a set V, a binary operator $+$, and a scalar multiplication $F \times V \to V$ with $(\lambda, v) \mapsto \lambda v$. These operators must satisfy the following:

- $(V, +)$ is an abelian group
- $1v = v \quad \forall v \in V$
- $\lambda(\mu v) = \mu(\lambda v) \quad \forall \mu, \lambda \in F$
- $\lambda(v + w) = \lambda v + \lambda w$
  $(\lambda + \mu)v = \lambda v + \mu v \quad \forall v, w \in V, \mu, \lambda \in F$

Here are some common examples of vector spaces.

Example 7.6

Examples of vector spaces.

- F^n, the set of n-dimensional column vectors with elements in field F, is an F-vector space.
- Given $A \in F^{m \times n}$, $\{x \in F^n : Ax = 0\}$, or the set of solutions to homogenous linear equations, is an F-vector space. This is also a **subspace** of F^n.
- $\mathbb{C}$ is an $\mathbb{R}$-vector space (and a $\mathbb{C}$-vector space)
- $\mathbb{R}$ is a $\mathbb{Q}$-vector space which is infinite dimensional
- {cont. functions from $\mathbb{R}$ to $\mathbb{R}$} is an $\mathbb{R}$-vector space which is also infinite dimensional
- {solns to $y'' = -y$} is an $\mathbb{R}$-vector space which is 2-dimensional, since all solutions are a linear combination of $\cos t$ and $\sin t$ (equivalently, e^{it} and e^{-it}). This vector space is also a subspace of the previous example.
- polynomials with real coefficients of degree $< n$ is an n-dimensional vector space over $\mathbb{R}$

8 September 26, 2022

8.1 Vector spaces

Let V be a vector space over field F.

Definition 8.1
A **linear combination** of $v_1, v_2, \ldots, v_n \in V$ is any element $\lambda_1 v_1 + \ldots + \lambda_n v_n$ with $\lambda_i \in F$.

Definition 8.2
Let $S \subseteq V$ be a subset of V. Then $\operatorname{span}(s) = \{\text{all lin. combs of } S\}$.

This is the smallest subspace of V containing S.

Definition 8.3
A subset $S \subseteq V$ is **linearly dependent** if there exists coefficients $\lambda_i \in F$ for $s \in S$ such that $\sum_{s \in S} \lambda_s s = 0$ and the coefficients λ_s are not all 0 (and there are only finitely many non-zero coefficients).

We say that a subset $S \subseteq V$ is linearly independent when this quality does not hold. That is, $\sum_{s \in S} \lambda_s s = 0 \iff \lambda_s = 0 \quad \forall s \in S$.

Proposition 8.4
If $v_1, \ldots, v_n$ are linearly independent, then $\lambda v_1 + \ldots + \lambda_n v_n = \lambda_1' v_1 + \ldots + \lambda_n' v_n \iff \lambda_i = \lambda_i' \quad \forall i$.

Proof.

$$\sum_i v_i(\lambda_i - \lambda_i') = 0 \iff \lambda_i = \lambda_i',$$

by the definition of linear independence. □

Definition 8.5
A **basis** of V is a subset S that spans V and is linearly independent.

The standard basis for $\mathbb{R}^3$:

$$\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix}\right\}.$$

Let $V = \{f : \mathbb{R} \to \mathbb{C} \text{ s.t. } f'' = -f\}$. A basis that works is $\{\sin(x), \cos(x)\}$. So does $\{e^{ix}, e^{-ix}\}$.

Definition 8.6

V is **finite-dimensional** if it has a finite basis.

Lemma 8.7

If $v_1, \ldots, v_r$ spans V and $w_1, \ldots, w_s$ are linearly independent, then $r \geq s$.

Proof. Write $w_j = \sum_{i=1}^r A_{ij} v_i$ for $A \in F^{r\times s}$. If $r < s$, then there exists nonzero $x \in F^s$ such that $Ax = 0$. But then

$$\sum_{j=1}^s x_j w_j = \sum_{i=1}^r \left(\sum_{j=1}^s A_{ij} x_j\right) v_i = 0,$$

which is a contradiction. □

Definition 8.8

If V, W are are vector spaces over F, then a **linear transformation** from V to W is a homomorphism $T : V \to W$ such that $T(v + v') = T(v) + T(v')$ and $T(\lambda v) = \lambda T(v)$ for all $v \in V, \lambda \in F$.

We call T a **linear operator** if $V = W$. Moreover, T is an isomorphism if it's a bijection, in which case T^{-1} is also a linear transformation.

9 September 28, 2022

More definitions + linear algebra. Bleh

10 September 30, 2022

11 October 3, 2022

12 October 5, 2022

12.1 Cauchy-Schwarz

Definition 12.1

The **inner product** or **dot product** of any two vectors $x, y \in \mathbb{R}^n$ is given by

$$\langle x, y\rangle = x \cdot y = x^t y = x_1 y_1 + \ldots + x_n y_n$$

Using our definition of inner product (does not necessarily have to be a dot product), we can define a few more quantities.

Definition 12.2

The **length** of a vector x is given by $|x| = \sqrt{\langle x, x\rangle}$.

Definition 12.3

The **distance** between two vectors x, y is given by $|x - y|$.

Like the usual dot product, we would like if $\langle x, y\rangle = |x||y|\cos\theta$, for some meaning of θ in our vector space. What happens if we view this definition as an angular measurement in and of itself?

Well, this should work as long as we don't have any domain errors. If $x, y \neq 0$, is it true that

$$\left|\frac{\langle x, y\rangle}{|x||y|}\right| \leq 1?$$

The answer is yes!

Theorem 12.4 (Cauchy-Schwarz Inequality)

$$|\langle x, y\rangle| \leq |x||y| \quad \forall x, y \in \mathbb{R}^n$$

Proof. We start with

$$|x - \lambda y|^2 \geq 0 \quad \forall \lambda \in \mathbb{R}.$$

Expanding,

$$\begin{aligned} |x - \lambda y|^2 &= \langle x - \lambda y, x - \lambda y \rangle \\ &= |x|^2 - 2\lambda \langle x, y \rangle + \lambda^2 |y|^2 \geq 0. \end{aligned}$$

Taking the discriminant implies the result. □

12.2 Orthogonal Matrices

Definition 12.5

A basis $x_1, x_2, \ldots, x_n \in \mathbb{R}^n$ is **orthogonal** if $\langle x_i, y_j \rangle = 0$ for $i \neq j$, and **orthonormal** if it additionally satisfies $|x_i| = 1 \quad \forall i$.

Definition 12.6

A matrix $A \in \mathbb{R}^{n \times n}$ is **orthogonal** if

$$\langle Ax, Ay \rangle = \langle x, y \rangle \quad \forall x, y \in \mathbb{R}^n$$

Theorem 12.7

For $A \in \mathbb{R}^{n \times n}$, the following are equivalent:

(1) A is orthogonal

(2) $|Ax| = |x| \quad \forall x \in \mathbb{R}^n$

(3) $A^t A = I_n$

(4) The columns of A are orthonormal.

(5) The rows of A are orthonormal.

Fun fact: IOAA 2022 DA Q2 was basically just (3), Andrew didn't know linear algebra so he mindlessly bashed out inverses and wasted all his time, sad.

Proof. (1) $\implies$ (2): When we set $x = y$ in the definition of orthogonality, $\langle Ax, Ax \rangle =$

$\langle x, x\rangle \iff |Ax|^2 = |x|^2$.

(2) $\implies$ (1): We can express the inner product in terms of length (this technique is called **polarization**), by

$$\langle x, y\rangle = \frac{|x+y|^2 - |x|^2 - |y|^2}{2}.$$

Therefore, if lengths are preserved, then so are inner products.

(3) $\implies$ (1): We have $\langle x, y\rangle = x^t y$. On the other hand, $\langle Ax, Ay\rangle = (Ax)^t Ay = x^t A^t Ay = x^t y \iff A^t A = I_n$.

(1) $\implies$ (3): Let $e_1, \dots, e_n$ be the standard basis. Then

$$\langle e_i, e_j\rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

(δ_{ij} is called the **kronecker delta**.) So we have

$$\delta_{ij} = \langle Ae_i, Ae_j\rangle = e_i^t(A^t A)e_j = (A^t A)_{ij},$$

thus $A^t A$ is the identity.

(3) $\iff$ (4):

$$A^t = \begin{pmatrix} — & a_1 & — \\ & \vdots & \\ — & a_2 & — \end{pmatrix}, \quad A = \begin{pmatrix} | & & | \\ a_1 & \dots & a_n \\ | & & | \end{pmatrix}$$

$(A^t A)_{ij} = a_i^t a_j$, so $A^t A = I_n \iff a_i^t a_j = \delta_{ij} \iff a_1, a_2, \dots, a_n$ are orthonormal.

The proof for (3) $\iff$ (5) is analogous, so we are done. □

Definition 12.8

The **orthogonal group**

$$O_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n} : A^t A = I_n\} \subset GL_n(\mathbb{R})$$

Definition 12.9

The **special orthogonal group**

$$SO_n(\mathbb{R}) = \{A \in O_n(\mathbb{R}) : \det A = 1\}$$

Note: $\det(A^t)\det(A) = 1 \implies \det A = \pm 1$, so $SO_n(\mathbb{R})$ has index 2 in $O_n(\mathbb{R})$.

13 October 7, 2022

13.1 Characterizing $O_n(\mathbb{R})$

Last lecture, we defined the orthogonal group of matrices over $\mathbb{R}$ as the set of matrices that preserves length and inner product. Today, we're going to look at some examples of $O_n(\mathbb{R})$, and classify the isometries of $\mathbb{R}^n$.

Example 13.1

$n = 1$.

In this case, $O_1(\mathbb{R}) = \{\pm I_1\}$.

Example 13.2

$n = 2$.

We have

$$O_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : \begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \text{ orthonormal} \right\}.$$

Since $a^2 + c^2 = 1$, we may paramaterize $a = \cos\theta$ and $c = \sin\theta$. By the orthogonal property, this implies $b = \pm\sin\theta$ and $d = \mp\cos\theta$.

This gives us

$$SO_2(\mathbb{R}) = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R} \right\},$$

which is the set of rotations. The other coset is

$$\left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} : \theta \in \mathbb{R} \right\}.$$

We claim these are the set of reflections. The characteristic equation here is $\lambda^2 = 1 \implies \lambda = \pm 1$. So, there exists $v, w \in \mathbb{R}^2$ with $Av = v$ and $Aw = -w$, in which case

$$\langle v, w \rangle = \langle Av, Aw \rangle = \langle v, -w \rangle \implies \langle v, w \rangle = 0.$$

This implies that A is a reflection through the line through v. We say that matrices in $O_2(\mathbb{R})$ are **orientation-preserving** if their determinant is 1, and **orientation-reversing** if their determinant is -1.

A key fact is that $SO_2(\mathbb{R})$ consists only of rotations. When we try to multiply two elements in the other coset, we get a composition of two reflections. Since reflections preserve distance to the origin, this is equivalent to a rotation, so this agrees with our analysis.

We'll upgrade our next example to a theorem, since its an important (and not necessarily intuitive) result.

Theorem 13.3
Every element of $SO_3(\mathbb{R})$ is a rotation about some axis in $\mathbb{R}^3$.

First, we prove a lemma.

Lemma 13.4
Every matrix $A \in SO_3(\mathbb{R})$ has an eigenvalue of 1.

Proof. We want to show that $\det(A - I_3) = 0$. We know $AA^t = I_3$, so $\det(A) = \det(A^t) = 1$. Then,

$$\begin{aligned}
\det(A - I_3) &= \det(A - AA^t) \\
&= \det(A)\det(I_3 - A^t) \\
&= \det((I_3 - A)^t) \\
&= \det(A - I_3) \cdot (-1)^3,
\end{aligned}$$

which is enough to imply our result. □

Now, we are ready to prove Theorem 13.3.

Proof. By our lemma, there exists $x \neq 0$ such that $Ax = x$; in other words, A fixes some pole in $\mathbb{R}^3$. Without loss of generality, let $|x| = 1$. Let B be the basis matrix formed when we extend x to an orthonormal basis x, y, z of $\mathbb{R}^3$. Then, under our new basis,

$$A \mapsto B^{-1}AB = \begin{pmatrix} 1 & 0 & 0 \\ 0 & a & b \\ 0 & c & d \end{pmatrix}.$$

Note that changing our basis does not disturb the determinant, since $\det B \det B^{-1} = 1$. To preserve orthonormality, $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SO_2(\mathbb{R})$ is a rotation about our fixed pole, so we're done. $\square$

Interesting things to note: Lemma 13.4 holds for any n odd, since all logic holds as long as we are flipping the sign for an odd number of rows in the last equality. Also, nicer characterizations of $SO_n(\mathbb{R})$ when $n \geq 4$ are more difficult. For example, we cannot necessarily guarantee single rotations, because

$$\begin{pmatrix} \cos\alpha & -\sin\alpha & 0 & 0 \\ \sin\alpha & \cos\alpha & 0 & 0 \\ 0 & 0 & \cos\beta & -\sin\beta \\ 0 & 0 & \sin\beta & \cos\beta \end{pmatrix} \in SO_4(\mathbb{R}),$$

which performs two rotations simultaneously in orthogonal subspaces at the same time.

13.2 Isometries of $\mathbb{R}^n$

Definition 13.5

$f : \mathbb{R}^n \to \mathbb{R}^n$ is an **isometry** if f preserves distances. In other words,

$$|f(x) - f(y)| = |x - y| \quad \forall x, y.$$

For example, $f(x) = Ax$ for $A \in O_n(\mathbb{R})$ is an isometry. Also, $f(x) = x + b$ for $b \in \mathbb{R}^n$ is an isometry. It turns out that the composition of these two generates all possible isometries.

Theorem 13.6

Every isometry of $\mathbb{R}^n$ is of the form $x \mapsto Ax + b$ where $A \in O_n(\mathbb{R})$, $b \in \mathbb{R}^n$.

Proof. We may assume $f(0) = 0$, since $x \mapsto x - f(0)$ is an isometry.

First, we show that f necessarily preserves inner products. We can do this by using the same polarization technique that we used last lecture:

$$\begin{aligned}\langle x, y\rangle &= \frac{|x-0|^2+|y-0|^2-|x-y|^2}{2} \\ &= \frac{|f(x)-f(0)|^2+|f(y)-f(0)|^2+|f(x)-f(y)|^2}{2} = \langle f(x), f(y)\rangle.\end{aligned}$$

Next, we show that f is linear, i.e., $f(x+y) = f(x)+f(y) \quad \forall x, y \in \mathbb{R}^n$.

$$\begin{aligned} & z = x+y \\ \iff & |z-x-y|^2 = 0 \\ \iff & \langle z,z\rangle + \langle x,x\rangle + \langle y,y\rangle - 2\langle x,z\rangle - 2\langle y,z\rangle + 2\langle x,y\rangle = 0 \\ \iff & |f(z)-f(x)-f(y)|^2 = 0 \\ \iff & f(x+y) = f(x)+f(y),\end{aligned}$$

where the transition from the third to fourth line is justified by the fact that f preserves inner products (so we can replace $a \mapsto f(a)$ everywhere).

Finally, we show that f preserves scalar multiplication, i.e., $f(\lambda x) = \lambda f(x)$. The proof here is exactly the same as the proof for showing linearity.

$$\begin{aligned} & y = \lambda x \\ \iff & |y-\lambda x|^2 = 0 \\ \iff & \langle y,y\rangle - 2\lambda\langle y,x\rangle + \langle y,y\rangle = 0 \\ \iff & |f(y)-\lambda f(x)|^2 = 0 \\ \iff & f(y) = \lambda f(x).\end{aligned}$$

Since f is linear and preserves scalar multiplication, f is a linear operator. Moreover, f preserves inner products, so we must have $f \in O_n(\mathbb{R})$, and we're done. □

Definition 13.7

$\mathbf{M_n}$ is the group of isometries of $\mathbb{R}^n$.

As a caveat, Prof. Cohn notes that this notation technically is not standardized, its just what Artin uses.

14 October 12, 2022

14.1 Affine Transformations

Review from last lecture: M_n is the set of all isometries (rigid motions) of $\mathbb{R}^n$, given by

$$M_n = \{x \mapsto Ax + b : A \in O_n(\mathbb{R}), b \in \mathbb{R}^n\}.$$

Example 14.1

What happens when we compose isometries?

Composing $x \mapsto A'x + b'$ and $x \mapsto Ax + b$ gives

$$A'(Ax + b) + b' = A'Ax + A'b + b'.$$

In particular, $\pi : M_n \to O_n(\mathbb{R})$ given by $(x \mapsto Ax + b) \mapsto A$ is a homomorphism, where $\ker(\pi)$ is given by the set of translations. It is also worth noting that $M_n \ncong O_n(\mathbb{R}) \times \mathbb{R}^n$, since the translation is not completely symmetric. If the translation term was $b + b'$ instead of $A'b + b'$, then it would be a direct product. Instead, this is an example of a **semi-direct product**.

Definition 14.2

An **affine transformation** is a mapping of a plane given by $x \mapsto Ax + b$. This generalizes the linear transformation, which requires $b = 0$.

Linear transformations preserve linearity. We can make an analogous statement for affine transformations.

Proposition 14.3

Affine transformations preserve weighted averages.

Proof. Let $T(x) = Ax + b$ be an affine transformation.

Then,

$$\begin{aligned} T(\lambda_1 x_1 + \ldots + \lambda_n x_n) &= \lambda_1 A x_1 + \ldots + \lambda_n A x_n + b \\ &= \lambda_1 T(x_1) + \ldots + \lambda_n T(x_n) + (1 - \lambda_1 - \ldots - \lambda_n) b \\ &= \lambda_1 T(x_1) + \ldots + \lambda_n T(x_n) \iff \sum \lambda_i = 1. \end{aligned}$$

□

14.2 Symmetry Groups

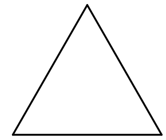
Definition 14.4
Given any subset $S \subseteq R^n$, its **symmetry group** is given by the subset of M_n

$$\{T \in M_n : TS = S\}.$$

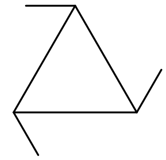
Example 14.5
Here are some general examples of symmetry groups.

- The symmetry group of $\mathbb{R}^n$ is M_n itself.
- The symmetry group of $\{0\}$ is $O_n(\mathbb{R})$
- The symmetry group of any sphere centered at 0 is $O_n(\mathbb{R})$.

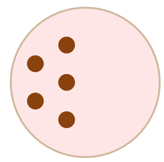
Example 14.6
Let's look at some more specific examples of symmetry groups.



This triangle has six symmetries, namely, D_3.



This triangle only has three symmetries, namely, C_3. This triangle cannot be equivalent to itself under any isometry which reverses the orientation of the plane (e.g., any isometry that includes a single reflection), so we say that it is **chiral**.



Prof. Cohn tells us a story about “none pizza with left beef". This pizza has nothing on it (including cheese or sauce), except for beef on the left half of the pizza. While it might initially seem like the pizza is chiral, it’s actually not, since a reflection from any axis θ from the vertical amounts to a rotation by $180-2\theta$. The symmetry group here is C_1.

14.3 Classifying rigid motions of $\mathbb{R}^2$

Rigid motions of $\mathbb{R}^2$ are transformations that preserve distance, so classifying the rigid motions amounts to classifying isometries. All isometries are of the form $x \mapsto Ax + b$ where $A \in O_2(\mathbb{R})$, i.e., $\det A = \pm 1$ and A is either orientation preserving or reversing.

First consider when A is orientation preserving.

- $A = I_2$. In this case $x \mapsto x + b$ is a **translation**.
- $A \in SO_2(\mathbb{R})$, $A \neq I_2$. In this case, A has eigenvalues $e^{\pm i\theta} \neq 1$, so $A - I_2$ is invertible. If we let $x_0 = (A - I_2)^{-1} b$, then

$$A(x + x_0) - x_0 = Ax + A(A - I_2)^{-1} b - (A - I_2)^{-1} b = Ax + b,$$

  so $x \mapsto Ax + b$ is conjugate to A under translation by x_0. In other words, this is a **rotation** about x_0.

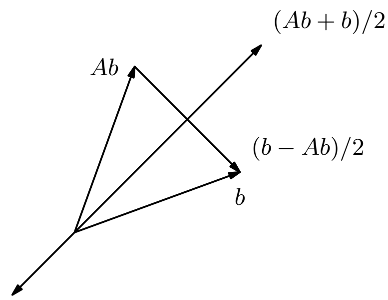
Now consider when A is orientation reversing.

- Let A be a reflection satisfying $Ab = -b$, i.e., b is perpendicular to the reflection line. Then

$$Ax + b = A\left(x - \frac{b}{2}\right) + \frac{b}{2},$$

  so $x \mapsto Ax + b$ is conjugate to the **reflection** A under translation by $-b/2$.

- Let A be a reflection with $Ab \neq -b$.



Then

$$Ax + b = \left(Ax + \frac{b - Ab}{2}\right) + \frac{b + Ab}{2}.$$

We know that $(b-Ab)/2$ is perpendicular to the reflection line, while $(b+Ab)/2$ is parallel to the reflection line (refer to the diagram). Therefore, the first term is conjugate to the reflection A by our previous case, while the second term is a translation parallel to our reflection line. Together, this gives us a **glide reflection**.

14.4 Finite Subgroups of $O_2(\mathbb{R})$

Now that we have classified all elements in M_2, lets look at some examples of subgroups, starting with the finite subgroups of $O_2(\mathbb{R})$.

Example 14.7

The cyclic group $C_n = \langle g \rangle$.

The cyclic group is the group of rotations by $2\pi k/n$. This group forms the rotational symmetries of an n-gon.

Example 14.8

The dihedral group $D_n = \langle g, h \rangle$.

The dihedral group is the group of rotations by $2\pi k/n$ with reflections; i.e., multiplication satisfies $g^n = 1$, $h^2 = 1$, and the reflection law $hgh = g^{-1}$. This group forms the full symmetry group of an n-gon. Sometimes, the dihedral group is labelled D_{2n}, since its order is $2n$.

15 October 14, 2022

16 October 17, 2022

16.1 Discrete Subgroups of M_2

Definition 16.1
$G \subseteq M_2$ is **discrete** if (1) $\exists \epsilon > 0$ such that no two distinct translation elements in G are within ϵ of each other and (2) $\exists \epsilon_\theta > 0$ such that no two distinct rotation elements in G are within ϵ_θ of each other.

Example 16.2
Discrete subgroups of $\mathbb{R}^2$.

- $\{0\}$. This is the trivial group, so it is discrete.
- $\mathbb{Z}\alpha$, where $\alpha \neq 0$.
- $\mathbb{Z}\alpha + \mathbb{Z}\beta$, where α, β linearly independent

Definition 16.3
Let G be a discrete subgroup of M_2. Define $\pi : G \to O_2(\mathbb{R})$ such that $\pi(Ax + b) = A$. Then, the **point group** $\overline{G} = \operatorname{im} \pi$, and the **lattice** $L = \ker \pi$.

L is a normal subgroup in G, so we also have $\overline{G} = G/L$. These definitions should feel somewhat intuitive. The kernel of π is the set of translations, since $A = I$, and this corresponds to our lattice. The image of π is the set of all possible rotations / reflections of a single point in the plane, which corresponds to the point group.

Theorem 16.4
$\forall A \in \overline{G}$, $b_0 \in L$, we have $Ab_0 \in L$. In other words, $\overline{G}$ preserves the lattice.

Intuitively, if this wasn't true, this would be pretty catastrophic. For example, if your lattice was a square grid, and your point group somehow did not preserve the symmetries of a square, you would generate points outside of your lattice, and therefore your lattice would not be a square grid.

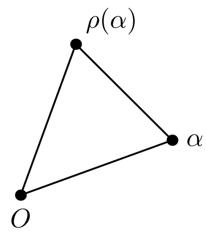
Proof. Since $A \in \overline{G}$, there exists some map $\varphi \in G$ taking $x \mapsto Ax + b$. Note that $\varphi^{-1} = A^{-1}x - A^{-1}b$. Conjugating the map $x \mapsto x + b_0$ (which is in G, since it is in L) by φ gives

$$x \mapsto A(A^{-1}x - A^{-1}b + b_0) + b = x + Ab_0,$$

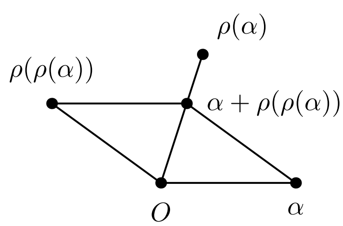
so $Ab_0 \in L$. □

Theorem 16.5 (Crystallographic Restriction)
If $L \neq \{0\}$, then $\overline{G} = C_n$ or D_n with $n \in \{1, 2, 3, 4, 6\}$.

Proof. Pick some $\alpha \in L - \{0\}$ with α minimal. Suppose $\rho \in G$ is a rotation by $2\pi/n$.



Whenever $n > 6$, like in the picture above, $|\rho(\alpha) - \alpha| < |\alpha|$. By our last theorem $\rho(\alpha) \in L \implies \rho(\alpha) - \alpha \in L$, so this contradicts the minimality of $|\alpha|$. Therefore, $n \leq 6$.



If ρ is a rotation by $2\pi/5$, $0 < |\alpha + \rho(\rho(\alpha))| < |\alpha|$, so $n \neq 5$. □

add something about frieze groups here

17 October 19, 2022

17.1 Group Actions

Definition 17.1

Given a group G and set S, the **action** of G on S is a function $G \times S \to S$ mapping $(g,s) \mapsto g\cdot$ satisfying

(i) $1 \cdot s = s \quad \forall s \in S$

(ii) $(gh) \cdot s = g \cdot (hs) \quad \forall g,h \in G, s \in S$

Intuitively, a group action can be thought of as a composition of "functions" (elements in g) on elements in S, such that elements of S are mapped to other elements of S.

Here are some examples:

- Given a field F, $GL_n(F)$ acts on F^n through the action $A \cdot x = Ax$.
- S_n acts on $\{1,2,\dots,n\}$ through the action $\pi \cdot i = \pi(i)$.
- M_n acts on $\mathbb{R}^n$ by $f \cdot x = f(x)$.

Proposition 17.2

An action of G on S implies a homomorphism $f : G \to \pi(S)$, where $\pi(S)$ denotes the set of permutations of S.

Proof. Let $f(g)(s) = g \cdot s$. First, we show that $f(g)$ is an element of $\pi(s)$, which amounts to showing that $f(g)$ is a bijective map. The co-domain of f is S by the definition of a group action.

First, $f(g^{-1})(f(g)(s)) = f(g^{-1})(g \cdot s) = s$. Thus $f(g^{-1})f(g)$ is the identity function, implying that $f(g)$ is injective; if not, then $f(g^{-1})f(g)$ maps two elements to the same element.

Second, $f(g)f(g^{-1})(s) = s$, so $f(g)f(g^{-1})$ is also the identity function. This implies that $f(g)$ is surjective, since $f(g)f(g^{-1})$ is surjective. Thus, $f(g)$ is a bijective mapping.

It remains to show that f is a homomorphism:

$$\begin{aligned} f(g_1g_2)(s) &= (g_1g_2)\cdot s \\ &= g_1\cdot(g_2\cdot s) \\ &= f(g_1)f(g_2)(s). \end{aligned}$$

□

Theorem 17.3 (Cayley's theorem)
If $|G| = n$, then G is isomorphic to a subgroup of S_n.

Proof. Consider the action of G on itself given by $g\cdot h = gh$. Using the above proposition, there is a homomorphism $f : G \to \pi(G)$. Then, since $\ker G = \{1\}$, f is injective, so G is isomorphic to some subgroup of $\pi(G)$, implying the result. □

17.2 Orbit-Stabilizer Theorem

Suppose G acts on S.

Definition 17.4
The **orbit** of $s \in S$ is the set

$$Gs = \{g\cdot s : g \in G\}.$$

In other words, the orbit of s as its image over all possible elements in G. Orbits may overlap; for example, when s_1 is in the orbit of s_2, then s_2 will also be in the orbit of s_1, and in particular the orbits of s_1 and s_2 will be the same.

Definition 17.5
The **stabilizer** of $s \in S$ is the set

$$\mathrm{stab}_G(s) = \{g \in G : g\cdot s = s\}.$$

In other words, the stabilizer of s is all elements in G that fix s.

Definition 17.6

The action of G on S is **transitive** if $Gs = S$ for some (all) $s \in S$.

Example 17.7

The action of D_4 on $\mathbb{R}^2$. D_4 is the dihedral group of order 4, and the possible set of actions are $90°$ rotations of the plane, or reflections of the plane.

Let O be the origin. Then,

- $D_4 O = \{O\}$.
- $\text{stab}_{D_4}(O) = D_4$.

Let $XYZW$ be a square with center at O. Then,

- $D_4 X = \{X, Y, Z, W\}$.
- $\text{stab}_{D_4}(X) = \{1, \text{reflection}(XZ)\}$.

Proposition 17.8

The set of orbits partition S.

Proof. Suppose $Gs_1 \cap Gs_2 \neq \emptyset$. Then, $g_1 s_1 = g_2 s_2$ for some $g_1, g_2 \in G$. But then $s_1 = g_1^{-1} g_2 s_2$, so $s_1 \in Gs_2$, and $s_2 = g_2^{-1} g_1 s_1$, so $s_2 \in Gs_1$, and therefore $Gs_1 = Gs_2$. In other words, if any two orbits overlap, they must be the same orbit. Since all elements of s are part of their own orbit, all elements of S are in some orbit, so the proposition follows. $\square$

Theorem 17.9

Let G act on S, $s \in S$, and $H = \text{stab}_G(s)$. Then, there exists a bijection $G/H \to Gs$ mapping $gH \mapsto gs$. (Here, let G/H denote the set of left cosets of H, not the normal quotient group).

Proof. Let f be our bijection. Then, f is well-defined, since $g_1 h_1 = g_2 h_2 \implies g_1 h_1 s = g_2 h_2 s$ is always true since $h_1 s = h_2 s = s$ (H is the stabilizer). Therefore, $g_1 H = g_2 H \implies g_1 s = g_2 s$.

f is surjective, since every element of Gs is $gs = f(gH)$. If $f(g_1H) = f(g_2H)$, then

$$\begin{aligned} g_1 s &= g_2 s \\ \implies g_2^{-1} g_1 s &= s \\ \implies g_2^{-1} g_1 &\in H \\ \implies g_1 &\in g_2 H. \end{aligned}$$

Applying this symmetrically implies $g_1H = g_2H$, so f is injective, and the result follows. □

Theorem 17.10 (Orbit-Stabilizer Theorem)
For all $s \in S$,

$$|G| = |Gs| \cdot |\mathrm{stab}_G(s)|.$$

Proof. Follows from the previous Theorem. □

Example 17.11
Consider the group G, the set of rotations of a cube, acting on S. We can find $|G|$ in three different ways by letting S equal the set of faces, edges, or vertices of a cube.

- If S is the faces of a cube, then $|G| = 6 \cdot 4$, because for any $s \in S$, there are 6 ways to map s to another face ($|Gs|$) and 4 rotations preserving that face ($|\mathrm{stab}_G(s)|$).
- If S is the vertices of a cube, then $|G| = 8 \cdot 3$, because for any $s \in S$, there are 8 ways to map s to another vertex and 3 rotations preserving that vertex.
- If S is the edges of a cube, then $|G| = 12 \cdot 2$, because for any $s \in S$, there are 12 ways to map s to another edge and 2 ways to rotate the cube to preserve that edge, by flipping its vertices.

17.3 Addendum: Burnside's Lemma

Addendum: This was not covered in lecture, but the Orbit-Stabilizer Theorem implies Burnside's Lemma.

Theorem 17.12 (Burnside's Lemma)
Let $s \in S$ be a fixed point for $g \in G$ if $g \cdot s = s$. Let k be the number of orbits Gs. Then, the average number of fixed points for any $g \in G$ is equal to k.

Proof. Let S_i denote the ith orbit.

$$\begin{aligned}
\sum_{g\in G} |\{s \in S : g \cdot s = s\}| &= \sum_{s \in S} |\{g \in G : g \cdot s = s\}| \\
&= \sum_{i=1}^{k} \sum_{s \in S_i} |\mathrm{stab}_G(s)| \\
&= \sum_{i=1}^{k} \sum_{s \in S_i} \frac{|G|}{|S_i|} \\
&= k|G|,
\end{aligned}$$

where the third equality follows from the orbit-stabilizer theorem. □

Example 17.13
Consider the group G, the set of rotations of a square, acting on S, the set of all possible colorings of the square with n colors. In order to count the number of distinct ways to color the square, where colorings that can be obtained via rotation are considered the same, we want to find the number of orbits.

$G = C_4$ consists of the identity, a rotation by $\pm 90°$, and a rotation by $180°$. The identity fixes n^4 elements, the rotations by $\pm 90°$ fixes n elements, and the rotation by $180°$ fixes n^2 elements. Thus, the number of orbits is $(n^4 + n^2 + 2n)/4$, which is also the number of colorings up to rotation.

18 October 24, 2022

18.1 Conjugate stabilizers

Let G be a group and S be a set. Say that G acts on S by some action. Recall that the set of orbits partition S, so

$$|S| = \sum_i |Gs_i|,$$

where s_i are representatives of each orbit.

Proposition 18.1

Stabilizers of points in the same orbit are conjugate.

Proof.

$$\begin{aligned}\operatorname{stab}_G(gs) &= \{g' \in G : g' \cdot gs = gs\} \\ &= \{g' \in G : g^{-1}g'g \cdot s = s\},\end{aligned}$$

which is true if and only if $g^{-1}g'g \in \operatorname{stab}_G(s)$, so $\operatorname{stab}_G(gs) = g\operatorname{stab}_G(s)g^{-1}$. □

Example 18.2

Let G be the rotations of a cube, and S the set of vertices of the cube.

There's only one orbit, so the stabilizer is always trivial, and therefore every stabilizer is conjugate.

Example 18.3

Let $G = \mathbb{R}^2$ act by translation on S, the set of horizontal lines.

The stabilizers are all $\mathbb{R} \times \{0\}$, which are normal subgroups, so they are conjugate.

18.2 Regular Polyhedra

Let's examine the possibilities for regular polyhedra. A regular polyhedron is defined as a three dimensional geometric solid with identical regular polygons as faces. Let's do casework on the polygon for each face:

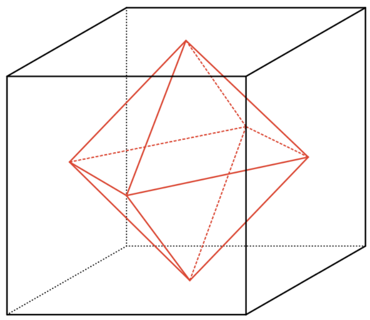
- Equilateral triangles. When three meet at a vertex, we get a tetrahedron. When four meet at a vertex, we get an octahedron. When five meet at a vertex, we get an icosahedron. We can't have ≥ 6 meet at a vertex, because six equilateral triangles becomes flat (a regular hexagon).
- Square. When three meet at a vertex, we get a cube. We can't have ≥ 4 for the same reason as before.
- Pentagon. When three meet at a vertex, we get a dodecahedron. We can't have ≥ 4 since $4 \cdot 108 > 360$.
- Hexagon. We can't have three meet at a vertex, since three hexagons would make a flat surface. Anything larger than a hexagon also won't work.

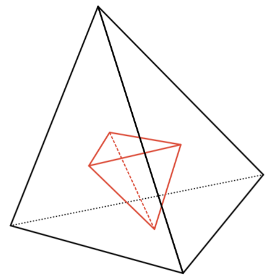
faces/# at vertex	3	4	5
triangle	tetrahedron	octahedron	icosahedron
square	cube	X	X
pentagon	dodecahedron	X	X
hexagon	X	X	X
$\vdots$			

Thus, there are **5** regular polyhedra in total. Here is a summary:

	vert.	edges	faces
tet.	4	6	4
cub.	8	12	6
oct.	6	12	8
icosa.	12	30	20
dodec.	20	30	12

Note that the pairs cubes and octahedra, icosahedra and dodecahedra, have the same number of edges, and a swapped number of vertices and faces. We say that these pairs of polyhedra are **dual**, because it is possible to interchange edges and vertices to obtain one from the other. A tetrahedron is said to be dual to itself. In general, dual shapes have the same symmetries.





18.3 Finite subgroups of $SO_3(\mathbb{R})$

Theorem 18.4
Every finite subgroup of $SO_3(\mathbb{R})$ is isomorphic to C_n, D_n, or the rotational symmetries of a platonic solid.

add proof

19 October 26, 2022

19.1 G acting on G

Let's consider the group actions of G on itself. One possible action is $(g,x) \mapsto gx$, but this is boring. A more exciting action is $(g,x) \mapsto gxg^{-1}$, otherwise known as the conjugation action.

Let's verify that is a valid action.

- $1 \times x = 1x1^{-1} = 1$
- $g \cdot (h \cdot x) = g \cdot (hxh^{-1}) = ghxh^{-1}g^{-1} = (gh)x(gh)^{-1} = gh \cdot x$

Definition 19.1

The **conjugacy class** $C(x)$ for each $x \in G$ is its orbit.

$$C(x) = \{gxg^{-1} : g \in G\}$$

Definition 19.2

The **centralizer** $Z(x)$ for each $x \in G$ is its stabilizer.

$$Z(x) = \{g \in G : gxg^{-1} = x\} = \{g \in G : gx = xg\}$$

This is also the set of elements in G that commute with x.

By the orbit-stabilizer theorem, $|G| = |C(x)||Z(x)|$.

Definition 19.3

The **center** Z is the set of all elements $g \in G$ that commutes with everything. In other words,

$$Z = \{g \in G : gx = xg \forall x \in G\} = \bigcap_{x \in G} Z(x)$$

In other words, $g \in Z \iff C(g) = \{g\} \iff Z(g) = G$. Also, since elements in Z commute with everything, Z is a normal subgroup in G.

We can provide an upper bound for the size of each conjugacy class as follows. For any $x \in G$, all powers of x commute with x, so $\langle x \rangle \subseteq Z(x)$. This implies $|\text{order}(x)| \mid |Z(x)| \implies |C(x)| \leq |G|/|\text{order}(x)|$.

19.2 Class Equations and p-groups

Definition 19.4

The **class equation** for a finite group G is given by

$$|G| = |C_1| + |C_2| + \ldots + |C_m|,$$

where each C_i are the conjugacy classes in G.

Let's look at a few examples of class equations for common groups.

Example 19.5

Consider $G = C_n$.

Its conjugacy class is given by

$$n = \underbrace{1 + 1 + \ldots + 1}_{n}.$$

The same is true for any abelian group with order n.

Example 19.6

Consider $G = S_3$.

By the property of conjugation in permutation groups, each conjugacy class is formed by the distinct cycle structures in G:

$$\{1\}, \{(12), (23), (13)\}, \{(123), (132)\}.$$

Therefore, the class equation is given by $6 = 1 + 2 + 3$.

Example 19.7

Consider $G = D_n$ (the dihedral group of order $2n$).

The conjugacy classes are formed by bab^{-1} for all $a, b \in G$. Let's do casework on a and b. Recall the multiplication rules for D_n: $x^n = 1$, $y^2 = 1$, and $yxy^{-1} = x^{-1}$.

a	x^i	x^i	$x^i y$	$x^i y$
b	x^j	$x^j y$	x^j	$x^j y$
bab^{-1}	$x^i x^j x^{-i}$	$x^j y x^i y x^{-j}$	$x^j x^i y x^{-j}$	$x^j y x^i y y x^{-j}$
	x^j	x^{-i}	$x^{i+2j} y$	$x^{-i+2j} y$

So x^i conjugates to $x^{\pm i}$, and $x^i y$ conjugates to $x^{\pm i + 2j} y$ for any j. If n is odd, then x^{2j} can be any power of x. If n is even, then x^{2j} and x^{1+2j} can only be the even and odd powers of x, respectively. This gives the following conjugacy classes for D_n:

When n is odd:

- $\{1\}$

- $\{x^i, x^{-i}\}$ for all $i \neq 0$
- $\{x^i y : \text{all } i\}$

When n is even:

- $\{1\}$
- $\{x^{n/2}\}$
- $\{x^i, x^{-i}\}$ for all $i \neq 0, n/2$
- $\{x^{2i} y : \text{all } i\}$
- $\{x^{2i+1} y : \text{all } i\}$

Definition 19.8
For p prime, G is a **p-group** if $|G| = p^k$.

Some examples of p-groups include C_p, C_{p^2}, or even

$$\left\{ \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{F}_p \right\} \in GL_3(\mathbb{F}_p).$$

Theorem 19.9
Every nontrivial p-group has a nontrivial center.

Proof. Let $|G| = p^k$. By the orbit-stabilizer theorem, all of its conjugacy classes have order dividing p^k. Therefore, its class equation can be written as

$$|G| = \sum_{i=0}^{k-1} c_i \cdot p^i,$$

where c_i is the number of conjugacy classes in G that have order p^i. This implies $p \mid c_0$; on the other hand, since the identity forms its own conjugacy class, $c_0 > 1$.

But we also know that $|C(x)| = 1 \iff x \in Z$, because it means that x commutes with everything. Thus, $Z \neq \{1\}$. $\square$

Corollary 19.10
If $|G| = p^2$ with p prime, then G is abelian.

Proof. By our theorem, we know that $|Z| = p$, or $|Z| = p^2$. If $|Z| = p^2$, then we are done. Otherwise, there is some $x \in Z(x)$ and $x \notin Z$. But, since $Z \subset Z(x) \subseteq G$, $|Z(x)| = p^2$, and therefore $Z(x) = G$. But this contradicts $x \notin Z$, so we're done. □

Corollary 19.11
If $|G| = p^2$ with p prime, then $G \cong C_{p^2}$ or $C_p \times C_p$.

Proof. If G has any element with order p^2, then $G \cong C_{p^2}$. Otherwise, all elements (with exception to the identity) have order p. Pick two elements x and y such that $y \in G - \langle x \rangle$. Then $\langle x \rangle \cap \langle y \rangle = \{1\}$ implies all $x^m y^n$ are distinct elements of G. Since there are p^2 elements of the form $x^m y^n$, this implies $G = \langle x, y \rangle$. Also, since G is abelian by the previous corollary, the map $\langle x, y \rangle \to C_p \times C_p$ given by $(x^m, y^n) \mapsto (m, n)$ is an isomorphism, so we are done. □

20 October 28, 2022

20.1 Simple Groups and Group Extensions

Definition 20.1
G is **simple** if $G \neq \{1\}$ and the only normal subgroups of G are $\{1\}$ and G.

Example 20.2
$G = C_p$ is simple for all prime p.

Definition 20.3
Let N be a normal subgroup in G. Then, G is an **extension** of G/N by N. This extension is **split** if G has a subgroup H isomorphic to G/N under the canonical map. Remember that the canonical map takes $g \mapsto gN$, so this says that there exists some subgroup of G that maps to all cosets of N.

Intuitively, extensions may seem equivalent to products (i.e., if G is an extension of Q by N, then G is isomorphic to $Q \times N$). But, they're not. Consider the following examples.

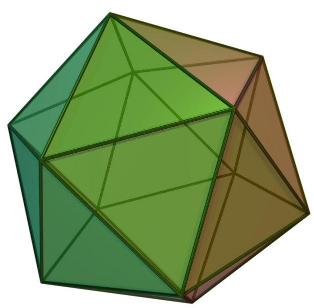
Example 20.4
Let $G = \mathbb{Z}$ and $N = 2\mathbb{Z}$.

Then, we can say that G is an extension of $G/N \cong C_2$ by N. This extension is not split, since there aren't any subgroups of G isomorphic to C_2. In this case, $G \ncong C_2 \times N$, since the former has a single generator, while the latter does not.

Example 20.5
Let $G = M_n$ and $N = O_n(\mathbb{R})$.

Then, we can say that G is a split extension of $G/N = \mathbb{R}^n$ by $O_n(\mathbb{R})$. This extension is split, because the set of translations is isomorphic to G/N under the canonical map. In this case, $G \ncong O_n(\mathbb{R}) \times \mathbb{R}^n$, since composition breaks the multiplication law. While they aren't a direct product, they are a **semidirect product**.

20.2 Rotations of an icosahedron



Let $G \subseteq SO_3(\mathbb{R})$ be the group of rotations of an icosahedron. Recall that an icosahedron has 20 faces, 30 edges, and 12 vertices. All faces are triangular and each vertex is the meeting of 5 faces.

To be explicit, when we say "the group of rotations", what we mean is that we choose any axis of rotation through the center of the icosahedron, and then rotate the figure through some angle such that its image after rotation is the same as the preimage (i.e., the alignment of the icosahedron is the same as before).

Now, we categorize the conjugacy classes of G. Recall that any conjugacy class is just a self-contained set of rotations that can be reached by some element in G. In other words, for any two rotations, if there is a rotation of the icosahedron that maps these two rotations together, they are in the same conjugacy class.

- The identity rotation. Conjugation of the identity always produces the identity again, so this is self contained.

- Any rotation that fixes a face. For these rotations, we choose our axis of rotation such that it goes through the center of our fixed face. In this case, this axis also goes through the center of the face opposite to our chosen face, so we have to consider pairs of opposite faces.

  There are 10 pairs of opposite faces. Moreover, each face borders 3 other faces, so there are two non-identity rotations per pair ($\pm 120°$), giving us 20 elements.

  This is self-contained, because regardless of our choice of perspective, the property that a rotation fixes a face is invariant. We must include both CW and CCW rotations, since for any pair of opposite faces (A, B), a CW rotation for A is a CCW rotation for B.

- Any rotation that fixes an edge. For these rotations, similar to the face rotations, we choose our axis of rotation such that it goes through the middle of our fixed edge. This axis will again go through the center of the edge opposite to our chosen edge, so we have to consider pairs of opposite edges.

  There are 15 pairs of opposite edges. For each of these pairs, there is one non-identity rotation ($180°$) which swaps the vertices of our chosen edge, giving us 15 elements.

- Any rotation of $\pm 72°$ that fixes a vertex. We choose our axis of rotation such that it goes through our chosen vertex. As before, this axis will also go through the vertex opposite our chosen vertex, so we need to consider pairs of vertices, of which there are 6. For each vertex, there are four non-identity rotations that fix that vertex. For any pair of vertices (A, B), a CW $72°$ rotation for A is the same as a CCW $72°$ for B, so these are contained in the same conjugacy class. This gives us $2 \cdot 6 = 12$ elements.

- Any rotation of $\pm 144^\circ$ that fixes a vertex. This gives us another $2 \cdot 6 = 12$ elements.

In sum, our class equation is given by

$$|G| = 1 + 20 + 15 + 12 + 12.$$

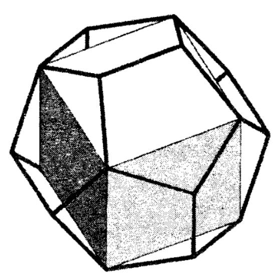
This is confirmed by the orbit-stabilizer theorem. If we let S be the set of faces, edges, or vertices of an icosahedron, the action of G on S gives $|G| = 20 \cdot 3 = 30 \cdot 2 = 12 \cdot 5$.

Theorem 20.6
G is simple.

Proof. If N is a normal subgroup of G, then $gN = Ng \implies N = gNg^{-1}$, so N must be a union of conjugacy classes. This implies that $|N|$ is equal to the sum of the numbers in some subset of $\{1, 12, 12, 15, 20\}$. But since we also must have $|N| \mid |G|$, this forces $|N| = 1$ or $|N| = 60$, so G is simple. $\square$

Theorem 20.7
$G \cong A_5$.

This is Theorem 7.4.4 in Artin.



Proof. The basic idea is to use the fact that there are five ways to inscribe a cube inside of an icosahedron. Because of duality, we can consider a dodecahedron instead (refer to the image above from Artin to help visualize). For any pentagonal face $ABCDE$, there are five ways to align an edge of the cube with diagonal vertices (i.e., AC, BD, CE, DA, EB). This completely determines the possible ways to inscribe a cube inside of the dodecahedron, since each edge of the cube (of which there are 12) lies on exactly one face of the dodecahedron (of which there are also 12).

Proposition 17.2 then gives a homomorphism φ from G to S_5 (the associated permutation representation). The kernel of this homomorphism is trivial, since all kernels are normal subgroups, and G is simple (the kernel can't be G, otherwise our homomorphism would do nothing).

This implies that φ is injective, so it determines an isomorphism from G to a subgroup of S_5. Now, restrict the sign homomorphism $S_5 \to \{\pm 1\}$ to G. If it was surjective, then the kernel would have order 30, which is not possible, since G is simple. So G must be a subgroup of A_5, which is the preimage of $+1$ under the sign homomorphism, but since they have the same size, they're the same group! $\square$

There turns out to be broad categorization of simple groups.

Theorem 20.8

There are four main categories of simple groups:

- C_p, with p prime
- A_n, for $n \geq 5$
- "groups of Lie type"
- 26 sporadic groups

Prof. Cohn says that even he doesn't fully know the proof for this Theorem, so we won't be going over it in class.

21 October 31, 2022

21.1 Jordan Hölder

Definition 21.1

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \ldots \supseteq G_r = \{1\}$$

is a **composition series** of length r when G_{i+1} is normal in G_i, and G_i/G_{i+1} is simple for all i.

Example 21.2

- $G = C_6 \overset{C_3}{\supseteq} \{1, g^3\} \overset{C_2}{\supseteq} \{1\}$
- $G = C_6 \overset{C_2}{\supseteq} \{1, g^2, g^4\} \overset{C_3}{\supseteq} \{1\}$
- $G = S_3 \overset{C_2}{\supseteq} A_3 \overset{C_3}{\supseteq} \{1\}$

In the first bullet point, we use $\{1, g, g^2\} \cong C_3$ and $\{1, g^3\} \cong C_2$. In the second bullet point, we use $\{1, g\} \cong C_2$ and $\{1, g^2, g^4\} \cong C_3$. In particular, the set of pairwise quotients are the same (up to isomorphism), which leads us to our main result today.

Theorem 21.3 (Jordan-Hölder)
r is uniquely defined by G, as is the set of all G_i/G_{i+1} up to permutations.

We'll start with a lemma.

Lemma 21.4
Let G be a group, and H, H' be normal subgroups of G. Then, if G/H, G/H' are simple and $H \neq H'$, $H/(H \cap H') \cong G/H'$ and $H'/(H \cap H') \cong G/H$.

insert diagram

Proof. First we prove that H and H' cannot be contained in the other. Suppose for the sake of contradiction that $H \subseteq H'$. Then H' corresponds to the subgroup H'/H

of G/H, and H'/H is normal in G/H because H' is closed by conjugation for any element in G.

G/H simple $\implies H'/H = \{1\}$ or $H'/H = G/H$. The former case implies $H' = H$, which is a contradiction. The latter case implies $H' = G$, which implies G/H' is not simple, contradiction.

So, $H \subsetneq H'$, and similarly $H' \subsetneq H$. Now consider $HH' = \{hh' : h \in H, h' \in H'\}$. This is a subgroup of G, because

$$h_1 h_1' h_2 h_2' = \underbrace{(h_1 h_2)}_{\in H} \underbrace{(h_2^{-1} h_1' h_2)}_{\in H'} h_2'$$

add explanation. This is a normal subgroup of G.

HH'/H is normal in G/H, so it must be $\{1\}$ or G/H. The former implies $H' \subseteq H$, which is impossible, so $HH'/H = G/H$.

Look at the map taking $H' \to HH' \to HH'/H$. $G/H = HH'/H \cong H'/(H' \cap H)$. Similarly, $G/H' \cong H/(H' \cap H)$. $\square$

Now we're ready to prove Jordan-Hölder.

Proof. We proceed with induction on r. The base case is when $r = 1$, in which case G itself is simple, so it holds.

For each r, we further induct on $|G|$. Assume true for a composition series of length smaller than r and all groups with size smaller than $|G|$.

Suppose we have two composition series:

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \ldots \supseteq G_r = \{1\}$$
$$G = H_0 \supseteq H_1 \supseteq H_2 \supseteq \ldots \supseteq H_s = \{1\},$$

for some $s \geq r$. If $G_1 = H_1$, then we're done, since $|G_1| < |G|$. So, suppose $G_1 \neq H_1$. Let $K_1 = G_1 \cap H_1$. Then $G_1/K_1 \cong G/H_1$ and $H_1/K_1 \cong G/G_1$ by our lemma.

Strategy: intersect everything with K_1. Let $K_i = G_i \cap K_1$. Then,

$$K_1 \supseteq K_2 \supseteq \ldots \supseteq K_r = \{1\}$$

is not a composition series but it is almost; it can contain duplicate elements which, once removed, becomes a composition series.

$K_i \to G_i \to G_i/G_{i+1}$. $K_i/(K_i \cap G_{i+1}) \cong$ normal subgroup of G_i/G_{i+1}. So K_i/K_{i+1} is either the trivial group or G_i/G_{i+1}.

Now consider

$$G_1 \supseteq G_2 \supseteq \ldots \supseteq G_r = \{1\},$$
$$G_1 \supseteq K_1 \supseteq K_2 \ldots \supseteq K_r = \{1\}.$$

These must have the same length by our induction hypothesis, since the first composition series has length $r-1$. This implies that the second series has a single duplicate element.

Now consider

$$H_1 \supseteq H_2 \supseteq \ldots \supseteq H_s = \{1\},$$
$$H_1 \supseteq K_1 \supseteq K_2 \ldots \supseteq K_r = \{1\}.$$

The bottom series has length $r-1$ when you remove the duplicate. This implies $s-1 = r-1$, so $s = r$.

Let's look again at all of our composition series (for the ones with K, remove the duplicate element):

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \ldots \supseteq G_r = \{1\}$$
$$G = G_0 \supseteq G_1 \supseteq K_1 \supseteq \ldots \supseteq K_r = \{1\}$$
$$G = H_0 \supseteq H_1 \supseteq K_1 \supseteq \ldots \supseteq K_r = \{1\}$$
$$G = H_0 \supseteq H_1 \supseteq H_2 \supseteq \ldots \supseteq H_s = \{1\}.$$

The first two series are the same by induction. The second and third are the same by our lemma. The last two are the same by induction. $\square$

22 November 2, 2022

22.1 Sylow's Theorem

According to Prof. Cohn, the correct way to prounounce Sylow is “See-low", at least for mathematicians.

Here is some motivation for his theorem(s). Whenever H is a subgroup of a finite group G, $|H| \mid |G|$. But there is not necessarily a subgroup with size every divisor of $|G|$. For example, $|A_5| = 60$, but there is no subgroup that has order 30 in A_5. What we can do instead is prove some results about the existence of subgroups with order dividing prime factors of $|G|$.

Definition 22.1

Let G be finite group and p a prime number such that $p^k \mid |G|$ but $p^{k+1} \nmid |G|$. Then, a ***Sylow p-subgroup*** of G is a subgroup of order p^k.

Note: many sources will split this theorem into multiple theorems, but Prof. Cohn will present it as one huge mega theorem.

Theorem 22.2 (Sylow's Theorem)

For any finite group G and prime p,

(1) G has a Sylow p-subgroup.

(2) All the Sylow p-subgroups in G are conjugate.

(3) Every p-subgroup (any subgroup whose order is a power of p) is contained in some Sylow p-subgroup.

(4) The number of Sylow p-subgroups is $1 \mod p$ and divides $|G|/p^k$.

22.2 Applications

We won't go over any proofs in class today. We proved (1) on Problem Set 7, and may present a different proof during lecture on Friday. Instead, we'll look through some examples to get a better sense of why this theorem is useful.

Example 22.3

Let's examine $G = S_4$.

$|G| = 24 = 2^3 \cdot 3$.

- There are 4 Sylow 3-subgroups in G, which are the 3-cycles: $\langle(123)\rangle$, $\langle(124)\rangle$, $\langle(134)\rangle$, $\langle(234)\rangle$. Since these generators all have the same cycle structure, all of these groups are conjugate to one another. Also, $4 \equiv 1 \pmod 3$ and 4 divides $|G|/3 = 8$.

- There are 3 Sylow 2-subgroups in G, which are the 4-cycles crossed with the 2-cycles (i.e., D_4): $\langle (1234), (13) \rangle$, etc. As before, conjugation works, and also, $3 \equiv 1 \pmod 2$ and 3 divides $|G|/8 = 3$.

Example 22.4
Let's examine $G = GL_2(\mathbb{F}_p)$.

$|G| = p(p-1)^2(p+1)$. One Sylow p–subgroup in G is $\left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbb{F}_p \right\}$. This can be generalized to $GL_n(\mathbb{F}_p)$ (i.e., the upper triangular matrices with ones along the diagonal).

Now, let's explore more arbitrary groups.

Example 22.5
Consider any G with $|G| = 15 = 3 \cdot 5$.

Let n_p be the number of Sylow p-subgroups. $n_3 \equiv 1 \pmod 3$ and $n_3 \mid 5 \implies n_3 = 1$. $n_5 \equiv 1 \pmod 5$ and $n_5 \mid 3 \implies n_5 = 1$. Therefore, there is a unique Sylow 3-subgroup, H, and a unique Sylow 5-subgroup, K in G.

H and K must be normal, since conjugating them with any element in G preserves their order, and they are both unique.

The normality of at least one of these groups is enough to imply that their set product $HK = \{hk : h \in H, k \in K\}$ is itself a subgroup of G. HK is closed, since

$$hkh'k' = \underbrace{hh'}_{\in H} \underbrace{(h')^{-1}kh'}_{\in K} k',$$

using the fact that K is normal. Also, every element has an inverse:

$$(hk)^{-1} = k^{-1}h^{-1} = h^{-1} \underbrace{hk^{-1}h^{-1}}_{\in K},$$

again using the fact that K is normal.

We further have that $3 \mid |HK|$, since H is a subgroup of HK, and $5 \mid |HK|$, since

K is a subgroup of HK. This implies $|HK| \mid 15$, which means $HK = G$.

Lemma 22.6

H,K are normal and $H \cap K = \{1\}$. Then $hkh^{-1}k^{-1} = 1$, for all $h \in H$, $k \in K$ (i.e., $hk = kh$, so multiplication is abelian).

Proof. $\underbrace{(hkh^{-1})}_{\in K}k^{-1} \in K$, since the first conjugate is in K. But also, $h\underbrace{(kh^{-1}k^{-1})}_{\in H} \in H$, since the second conjugate is in H. Therefore, this product is in $H \cap K$, so it must be 1. □

The intersection of H and K must be trivial, since they have different prime orders. Therefore, using the lemma, multiplication in G is abelian, so $G = HK$ implies $G \cong H \times K \cong C_3 \times C_5$.

Example 22.7

Let's examine any group G with $|G| = 2 \cdot 5$.

Since $n_2 \equiv 1 \pmod 2$ and $n_2 \mid 5$, $n_2 = 1$ or $n_2 = 5$. Since $n_5 \equiv 1 \pmod 5$ and $n_5 \mid 2$, $n_5 = 1$.

Let H, K be a Sylow 2-subgroup in G, and the Sylow 5-subgroup in G, respectively. K is normal because it is unique. HK is subgroup, and $G = HK$ using the same arguments as before. Let's denote $H = \langle x \rangle$ with $x^2 = 1$, and $K = \langle y \rangle$ with $y^5 = 1$. Then $xKx^{-1} = K \implies xyx^{-1} = y^n$ for some $n \in \{1, 2, 3, 4\}$ (if n could be 0, then y would be the identity, which is not possible).

This implies $G = \{x^i y^j : 0 \le i \le 1, 0 \le j \le 4\}$, with $x^2 = 1$, $y^5 = 1$, $yx = xy^n$. This completely determines the multiplication table, since it tells us how to swap x and y and reduce any product to the form xy^j.

It looks like we have 4 choices for G, but we can restrict n even further. Given $xyx^{-1} = y^n$, we have $y = x^2yx^{-2} = x(xyx^{-1})x^{-1} = (xyx^{-1})^n = (y^n)^n = y^{n^2}$, so $n^2 \equiv 1 \pmod 5$. So it turns out that $n \neq 2, 3$ and we must have $n \in \{1, 4\}$. This reduces the number of choices we have for G down to 2.

When $n = 1$, we have $xy = yx$, so G is abelian and $G \cong C_2 \times C_5$. In this case, $n_2 = 1$ and $n_5 = 1$ (which is the same case that we had in the previous example). When $n = 4$, $xy = y^{-1}x$, this matches the multiplication rule for the dihedral group, so $G \cong D_5$. In this case, $n_2 = 5$ (s, sr, sr^2, sr^3, sr^4) and $n_5 = 1$.

Example 22.8

Let's generalize further. Consider any group G with $|G| = pq$ for primes p, q and $p < q$.

We have $n_q \equiv 1 \pmod{q}$ and $n_q \mid p$, implying $n_q = 1$, so there is one unique Sylow q-subgroup. As before, this subgroup is normal. We also have $n_p \equiv 1 \pmod{p}$ and $n_p \mid q$. If $q \not\equiv 1 \pmod{p}$, this implies $n_p = 1$, giving a unique Sylow p-subgroup. Using the same arguments that we used in Example 22.5, this gives us that $G \cong C_p \times C_q$.

Now consider when $q \equiv 1 \pmod{p}$. In this case, we have $n_p = 1$ or $n_p = q$. Using the same arguments that we used in Example 22.7, this gives us an additional possibility for G:

$$G = \{x^i y^j : 0 \le i \le p-1, 0 \le j \le q-1\},$$

with $x^p = 1$, $y^q = 1$, and $yx = xy^n$ for some $n \in \{1, 2, \ldots, q-1\}$. As before, we can reduce our possibilities for n. Given $xyx^{-1} = y^n$, we have $y = x^p y x^{-p} = x(x \ldots (x(xyx^{-1})x^{-1}) \ldots x^{-1})x^{-1} = y^{n^p}$, so we must have $n^p \equiv 1 \pmod{q}$.

When $n \equiv 1 \pmod{q}$, G is abelian, so we get $C_p \times C_q$ again. Now we claim that all $n \not\equiv 1 \pmod{q}$, give the same group. The multiplicative group $\mathbb{F}_q^\times$ has order $q-1$ and is cyclic. Since $p \mid q-1$, there exists a subgroup with order p generated by some n: $\{1, n, n^2, \ldots, n^{p-1}\}$, all of which are roots to $n^p \equiv 1 \pmod{q}$. But now, when we consider mapping the generator $x \mapsto x^i$, we have $x^{ip} = 1$ and

$$yx^i = xy^n x^{i-1} = \ldots = x^i y^{n^i},$$

so this amounts to n and n^i yielding isomorphic groups for all i. Thus, there are only two possibilities for G.

23 November 4, 2022

23.1 Proving Sylow's Theorem

Recap from last time:

Definition 23.1

Let G be finite group and p a prime number such that $p^k \mid |G|$ but $p^{k+1} \nmid |G|$. Then, a **Sylow p-subgroup** of G is a subgroup of order p^k.

Theorem 23.2 (Sylow's Theorem)

For any finite group G and prime p,

(1) G has a Sylow p-subgroup.

(2) All the Sylow p-subgroups in G are conjugate.

(3) Every p-subgroup (any subgroup whose order is a power of p) is contained in some Sylow p-subgroup.

(4) The number of Sylow p-subgroups is $1 \mod p$ and divides $|G|/p^k$.

Today, we'll prove all four parts of this theorem. Proving this theorem is good culmination of everything that we have learned so far about group actions in this course.

Let's start with (1), which we proved on the problem set. We'll take a different approach here.

Proof. Consider the action of G by left multiplication on subsets (not necessarily subgroups) of G of size p^k. Suppose we find some subset $U \subseteq G$ with $|U| = p^k$ such that the orbit size is not divisible by p.

If so, let $H = \text{stab}_G(U) = \{h \in G : hU = U\}$. Then H is a p-group (i.e., its order is a power of p). We have that $Hu \in U \forall u \in U$, so U is partitioned into right cosets of H. Since every right coset has the same size ($|H|$), the size of H must divide the size of U, so H is a p-group. We also have $|G| = \underbrace{|GU|}_{v_p=0}|H| \implies p^k \mid |H|$, so H is a Sylow p-subgroup.

Now we are left to show that U exists. The number of subsets of size p^k is equal to $\binom{mp^k}{p^k}$. Since the orbits partition this set of subsets, if this quantity isn't divisible by p, some orbit size isn't, so we are left to show that this is not divisible by p.

Lemma 23.3

$$\binom{pa}{pa} \equiv \binom{a}{b} \pmod{p}.$$

Proof. By binomial expansion,

$$\binom{pa}{pb} = [x^{pb}](1+x)^{pa}$$

We know that $(1+x)^p = 1+x^p \pmod{p}$, since $\binom{p}{i} \equiv 0 \pmod{p}$ for all $1 \le i \le p-1$. Therefore,

$$[x^{pb}](1+x)^{pa} \equiv [x^{pb}](1+x^p)^a = \binom{a}{b}.$$

□

Applying the lemma k times,

$$\binom{mp^k}{p^k} \equiv m \pmod{p},$$

but $(p,m) = 1$, so we're done. □

Now we'll prove (2) and (3) together, by showing the following the following reformulation:

Lemma 23.4

Let H be a Sylow p-subgroup of G. Let K be any p-subgroup. Then $K \subseteq gHg^{-1}$ for some $g \in G$.

This is enough to show (2), because K can be any Sylow p-subgroup. This is also enough to show (3), because gHg^{-1} has the same size as H, so gHg^{-1} is also a Sylow p-subgroup. Note that we know H exists by (1).

Proof. Consider the action of K by left multiplication on the left cosets of H, i.e., $k \cdot gH = kgH$. It turns out that proving this lemma amounts to proving that there exists a fixed point in K.

We know $p \nmid |G/H| = m$, and we also know that every orbit size divides $|K| = p^\ell$. Since the orbits partition $|G/H|$, there must be an orbit of size 1, i.e., there exists a coset gH such that $kgH = gH$ for all $k \in K$.

$kgH = gH \forall k \in K \iff g^{-1}kgH = H \forall k \in K \iff g^{-1}kg \in H \forall k \in K \iff g^{-1}Kg \subseteq H \iff K \subseteq gHg^{-1}$, so we're done. □

Finally, let's prove (4).

Proof. Let X be the set of Sylow p-subgroups. We want to show that $|X| \mid m$, and $|X| \equiv 1 \pmod p$. By (2), X is a single orbit under conjugation by G.

By the orbit-stabilizer theorem,

$$p^k m = |G| = |X||N(H)|,$$

for some $H \in X$, where the normalizer of H

$$N(H) = \text{stab}_G(H) = \{g \in G : gHg^{-1} = H\}.$$

But H is a subgroup of $N(H) \implies p^k = |H| \mid |N(H)|$. Thus, $m = |X| \cdot \left(|N(H)|/p^k\right)$, so $|X|$ divides m.

Now, consider the conjugation action of H on X. There could be many orbits, since we are restricting our first action to elements of H, rather than all elements in G. All orbit sizes are powers of p, since they must divide $|H|$. Note that H itself has an orbit of size 1. We claim that this is the only orbit of size 1, which is enough to imply $|X| \equiv 1 \pmod p$.

Lemma 23.5
$\forall H, H' \in X, H \subseteq N(H') \iff H' = H \iff \{H'\}$ is an H orbit.

Proof. H, H' are both Sylow p-subgroups of the normalizer of H'. By (2), there exists $n \in N(H')$ s.t. $H = nH'n^{-1}$. By definition, the normalizer fixes H', so $H = H'$. □

For any $H' \in X$ that has an orbit of size 1, $H \subseteq N(H')$, since all elements in H are fixed points for H'. Thus, the lemma implies $H' = H$, so we're done. □

23.2 Addendum: simple groups with order 60

Addendum: the following was not covered in lecture, but I wanted to include it anyways, because it feels like a solid way to wrap up our discussions on orbits, stabilizers, and the Sylow's Theorem.

Theorem 23.6

The only simple group with order 60 is A_5.

This proof extends the logic that we used in Theorem 20.7, where we proved that A_5 was simple.

Proof. Let G be a simple group with order 60. Our goal is to show that $G \cong A_5$.

First, suppose that G acts non-trivially on some set S with order less than order to 5. By the same logic that we used in our proof of Theorem 20.7, this implies an injective homomorphism from G to the permutations of S, which is possible if and only if $|S| = 5$, in which case $G \cong A_5$.

Now, we go about constructing S. Assume for the sake of contradiction that $G \ncong A_5$. By Sylow's Theorem, $n_2 \mid 15 \implies n_2 \in \{1,3,5,15\}$. If $n_2 = 1$, then the unique Sylow 2-subgroup is normal, which breaks the assumption that G is simple. If $n_2 = 3$, or $n_2 = 5$, then the conjugation action from G onto the set of Sylow 2-subgroups is well-defined (Sylow groups are closed under conjugation, by Sylow's theorem), and certainly non-trivial. Thus, our argument in the previous paragraph breaks the assumption that $G \cong A_5$. We can use a similar line of reasoning to show that $n_3 \equiv 1 \pmod 3$ and $n_3 \mid 20 \implies n_3 = 10$.

Since all Sylow 3-subgroups are cyclic, they all intersect trivially. Now, suppose all Sylow 2-subgroups intersected trivially. If so, then the number of non-identity elements contained in the union of these subgroups is $3 \cdot 15 = 45$. But the number of non-identity elements contained in the unions of the Sylow 3-subgroups is $2 \cdot 10 = 20$, in which case the total number of non-identity elements exceeds the size of G.

Therefore, there exists distinct $H,\ H' \in G$ which are Sylow 2-subgroups and intersect non-trivially. Denote G^* the subgroup of G generated by the elements in $H \cup H'$. Since H is a strict subgroup of G^*, $|G^*| = 4k$ for some $k > 1$. Since we also have $|G^*| \mid |G|$, $k \in \{3,5,15\}$.

Finally, let $x \in H \cap H' - \{1\}$. Since $|H| = |H'| = 4$, they are abelian, and x commutes with everything in both groups. This implies that x commutes with everything in G^*, so $\langle x \rangle$ is a normal subgroup in G^*. If $k = 15$, then $G^* = G$, so this is not possible. Therefore, $k \in \{3, 5\}$. In either case, the action from G onto G/G^* defined by $(g_1, g_2G^*) \mapsto g_1g_2G^*$ is well-defined and non-trivial, so we are done by our first paragraph. $\square$

24 November 7, 2022

24.1 Bilinear Forms

Definition 24.1

Let V vector space over F. A **bilinear form** on V is a function $V \times V \to F$, $(x, y) \mapsto \langle x, y \rangle$ such that

- $\langle \lambda x, y \rangle = \langle x, \lambda y \rangle = \lambda \langle x, y \rangle$
- $\langle x_1 + x_2, y \rangle = \langle x_1, y \rangle + \langle x_2, y \rangle$
- $\langle x_1, y_1 + y_2 \rangle = \langle x, y_1 \rangle + \langle x, y_2 \rangle$

for all $\lambda \in F$, $x_1, x_2, y_1, y_2, x, y \in V$.

The first condition takes care of scalar multiplication, while the second and third conditions imply that our function is linear in both the first and second variable, hence “bilinear".

Proposition 24.2

Let $V = F^n$. For any $A \in F^{n \times n}$, $\langle x, y \rangle = x^t A y$ is a bilinear form. Moreover, all bilinear forms on F^n are of this form.

Proof. It is easy to verify that functions of this form are bilinear forms, so let's prove the other direction.

Consider the standard basis $e_1, \ldots, e_n$, where

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_i x_i e_i \quad \text{and} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_i y_i e_i.$$

Then our bilinear form

$$\begin{aligned}\left\langle \sum_i x_i e_i, \sum_j y_j e_j \right\rangle &= \sum_i x_i \left\langle e_i, \sum_j y_j e_j \right\rangle \\ &= \sum_{i,j} x_i y_j \left\langle e_i, e_j \right\rangle \\ &= x^t A y,\end{aligned}$$

where A is uniquely defined by $A_{ij} = \langle e_i, e_j \rangle$.

$\square$

What happens when we change our basis? Let

$$B = \begin{pmatrix} | & & | \\ b_1 & \dots & b_n \\ | & & | \end{pmatrix}$$

be the basis matrix transforming $\{e_1, \dots, e_n\}$ to $\{b_1, \dots, b_n\}$, i.e., $Be_i = b_i$. Then we have

$$\langle Bx, By \rangle = (Bx)^t ABy = x^t (B^t AB) y,$$

so the effect of changing our basis was effectively $A \mapsto B^t AB$.

Note the similarities between binary forms and linear operators. Both can be represented by matrices, but the basis transformation for a linear operator is conjugation, while the basis transformation for a binary form is $A \mapsto B^t AB$. **These are very different!**

Definition 24.3
A bilinear form $\langle \cdot, \cdot \rangle$ is **symmetric** if $\langle x, y \rangle = \langle y, x \rangle \quad \forall x, y \in V$.

Since $\langle x, y \rangle \in F$, $\langle x, y \rangle^t = \langle x, y \rangle$, so we can say that $x^t Ay = (x^t Ay)^t = y^t A^t x$. In other words, a binary form is symmetric if and only if $A = A^t$. This property of symmetry is preserved under the change of basis $A \mapsto B^t AB$, as we should expect.

24.2 Inner Products and Hermitian forms

Now let's look at the real numbers, $F = \mathbb{R}$. Over this field, inequalities are well-defined.

Definition 24.4

A bilinear form $\langle\cdot,\cdot\rangle$ is **positive semidefinite** if $\langle x,x\rangle \geq 0 \quad \forall x \in V$. Additionally, we say it is **positive definite** if also $\langle x,x\rangle = 0 \iff x = 0$.

Definition 24.5

An **inner product** on a real vector space is a symmetric, positive definite bilinear form.

With inner products, you can do geometry, define lengths, angles, etc. Over the next few lectures, we are going to classify all the possible inner products. We'll find that they're not all that different from the normal dot product. In fact, we'll find that they're all isomorphic to the dot product under suitable coordinates.

Example 24.6

Let's attempt to characterize the qualities of different binary forms

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle$$

$(x,y) \mapsto ?$	bilinear	symmetric	positive semidefinite	positive definite
$x_1 + y_1$	X			
$x_1 y_2$	Y	X		
$x_1 y_1 - x_2 y_2$	Y	Y	X	
$x_1 y_1$	Y	Y	Y	X
$x_1 y_1 + x_2 y_2$	Y	Y	Y	Y

The first row is not a bilinear form, since it does not satisfy $\langle \lambda x, y\rangle = \langle x, \lambda y\rangle = \lambda\langle x,y\rangle$. The second row is a bilinear form, but is not symmetric in x and y. The third row has this symmetry, but is not positive semidefinite, since $\langle x,x\rangle$ when $x_2 > x_1$ is strictly negative. The third row is positive semidefinite, but not positive definite, since $x_1 = 0$ is sufficient for x to have length 0. Finally, the last row is the usual dot product, so it satisfies all conditions for an inner product as we expect.

Now, let's generalize the Hermitian inner product.

Definition 24.7

Let V be a vector space over $\mathbb{C}$. We say that $\langle \cdot, \cdot \rangle$ is a **Hermitian form** on V if $\langle x, y \rangle$ is linear in y and $\langle y, x \rangle = \overline{\langle x, y \rangle} \quad \forall x, y \in V$.

Note that this definition implies conjugate linearity in x, i.e., $\langle \lambda x, y \rangle = \overline{\lambda} \langle x, y \rangle$. By the way, Hermitian is pronounced "her–mee–shun".

Definition 24.8

The **adjoint** matrix A^* of any $A \in \mathbb{C}^{n \times n}$ is given by $A^* = \overline{A}^t$.

Note that adjoint matrices follow many of the same rules that transpose matrices follow. For example, $(AB)^* = (\overline{AB})^t = B^* A^*$.

Therefore, Hermitian forms are almost the same as bilinear forms, but with extra conjugation. In particular, all Hermetian forms also have the matrix representation $\langle x, y \rangle = x^* A y$, but with the additional constraint that $A^* = A$, since we need $\langle x, y \rangle = \overline{\langle y, x \rangle} = \langle y, x \rangle^* = x^* A^* y$. Like bilinear forms, the change of basis for Hermitian forms maps $A \mapsto B^* A B$.

Since $\overline{\langle x, x \rangle} = \langle x, x \rangle$, we have $\langle x, x \rangle \in \mathbb{R}$, so inequalities are well-defined and we can use the same definitions for **positive semidefinite** and **positive definite** as before.

Definition 24.9

A **Hermitian inner product** is a positive definite hermitian form.

More on Hermitian inner products later. Now, let's look at some applications of bilinear forms.

Example 24.10

We can use bilinear forms to study quadratic functions. Consider any second degree polynomial in $x_1, \ldots, x_n$.

We can represent our function in the following way:

$$\underbrace{x^t A x}_{\text{quadratic terms}} + \underbrace{b^t x}_{\text{linear terms}} + \underbrace{c}_{\text{constants}}, \text{ with } A \in F^{n \times n}, b \in F^n, c \in F.$$

Note that since we have $x^tAx = (x^tAx)^t = x^tA^tx$, we can replace A with $(A+A^t)/2$ and our expression still holds, as long as $2 \neq 0$ in our field. Since $(A+A^t)/2$ is symmetric, we may therefore assume that it is always true that $A = A^t$.

For example, we may have

$$3x^2 + 4xy + 5y^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

where $A = \begin{pmatrix} 3 & 1 \\ 3 & 5 \end{pmatrix}$. Using $(A+A^t)/2$ instead, we find that the symmetric matrix $\begin{pmatrix} 3 & 2 \\ 2 & 5 \end{pmatrix}$ also works.

Example 24.11

Using the same idea, we can also use bilinear forms to represent the second degree taylor polynomial of any function $f : \mathbb{R}^n \to \mathbb{R}$ at $x_0 = \mathbb{R}^n$.

In this case, we can let

$$\begin{aligned} A_{ij} &= \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0), \\ b &= \nabla f(x_0), \\ c &= f(x_0). \end{aligned}$$

Since partial derivatives commute, A is symmetric!

25 November 9, 2022

25.1 Unitary Group

Let $\langle x, y \rangle = x^t y$ be the usual inner product.

The set of orthogonal matrices can be defined by the set of all matrices M satisfying $\langle Mx, My \rangle = \langle x, y \rangle$ for $x, y \in \mathbb{R}^n \iff M^tM = I_n$. In other words, we can write

$$O(n) = \{M \in \mathbb{R}^{n\times n} : M^tM = I_n\}.$$

Let's generalize to the Hermitian inner product. Let $\langle x, y\rangle = \overline{x}^t y$ on $\mathbb{C}^n$. Then, we have $\langle Mx, My\rangle = \langle x, y\rangle \iff x^*M^*My = x^*y \iff M^*M = I_n$, so this turns out to be analogous to the real case.

Definition 25.1
The **unitary group** is the set of complex matrices

$$U(n) = \{M \in \mathbb{C}^{n\times n} : M^*M = I_n\}.$$

Proposition 25.2
The eigenvalues of any Hermitian matrix are real.

Proof. Suppose $A^* = A$, $Av = \lambda v$, $v \neq 0$ for $\lambda \in \mathbb{C}, v \in \mathbb{C}^n$. Then we must have $v^*Av \in \mathbb{R}$, since $(v^*Av)^* = v^*A^*v = v^*Av$. We can also view this product as a Hermitian form.

We also know $v^*v \in \mathbb{R}$, since

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \implies v^*v = \sum_{j=1}^{n} |v_j|^2.$$

Thus, $v^*Av = \lambda v^*v$ implies

$$\lambda = \frac{v^*Av}{v^*v} \in \mathbb{R}.$$

□

25.2 Degeneracy

Standing assumption for the rest of this lecture: Let V be a finite dimensional real or complex vector space. Let $\langle\cdot,\cdot\rangle$ be a symmetric or Hermitian form on V, which is not assumed to be positive definite.

Definition 25.3
For any $v, w \in V$, we say they are **orthogonal**, or that $v \perp w$, if $\langle v, w\rangle = 0 \iff \langle w, v\rangle = 0$.

Definition 25.4
For any subspace W, we define the orthogonal subspace

$$W^\perp = \{v \in V : v \perp w \quad \forall w \in W\}$$

General forms yield interesting examples of orthogonal vectors. For example, vectors may be orthogonal to itself, or to the entire vector space.

Example 25.5
Let $V = \mathbb{R}^2$.

- If $\langle x, y\rangle = x^t \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} y$, then $\begin{pmatrix} 1 \\ 1 \end{pmatrix} \perp \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.
- If $\langle x, y\rangle = x^t \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} y$, then $\begin{pmatrix} 0 \\ 1 \end{pmatrix} \in (\mathbb{R}^2)^\perp$.

Definition 25.6
The **null space** are the vectors orthogonal to everything, i.e., $V^\perp$. We call vectors in the null space as **null vectors**. We say that our form is **degenerate** if $V^\perp \neq \{0\}$.

Let W be a subspace of V. We say that $\langle \cdot, \cdot \rangle$ is degenerate on W when $W \cap W^\perp \neq \{0\}$.

These definitions should intuitively align with what we already know about null spaces and kernels. For example, our form is degenerate if and only if the determinant of the corresponding matrix is 0, since invertible matrices automatically zero out any null vector.

Lemma 25.7
If $\langle \cdot, \cdot \rangle$ is non-degenerate and $x, y \in V$ satisfies $\langle x, v\rangle = \langle y, v\rangle$ for all $v \in V$, then $x = y$.

Proof. By linearity, we know $\langle x-y, v\rangle = 0$ for all $v \in V$. This implies $x-y \in V^\perp = \{0\}$, so $x = y$. $\square$

Proposition 25.8

If A is a matrix for $\langle\cdot,\cdot\rangle$ with respect to a basis, then null vectors form $\ker A$, or the null space of A itself.

Proof. $\langle x, y\rangle = x^* A y$ if x, y are coordinate vectors with respect to our basis. Then,

$$y \in V^\perp \iff x^* A y = 0 \quad \forall x \iff Ay = 0,$$

which is true by plugging in elementary basis vectors for x. □

Theorem 25.9

Let W be a subspace of V. Then

$$V = W \oplus W^\perp \iff \langle\cdot,\cdot\rangle \text{ non-deg on } W.$$

Recall that being non-degenerate on W means that $W \cap W^\perp = \{0\}$, and that the direct sum means that each $v \in V$ can be written uniquely as a sum of elements of W and $W^\perp$.

Proof. $V = W \oplus W^\perp$ if and only if $V = W + W^\perp$ and $W \cap W^\perp = \{0\}$. In other words, every vector in V can be written uniquely as the sum of vectors in $W, W^\perp$ if and only if they can be written possibly non-uniquely, and the intersection is null, in which case sums are unique. This implies the forward direction.

So now we need to show that $W \cap W^\perp = \{0\} \implies V = W + W^\perp$. First, pick a basis of W, and extend it to a basis of V, so that the matrix M for $\langle\cdot,\cdot\rangle$ in this basis is given by

$$M = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right),$$

with A the matrix for $\langle\cdot,\cdot\rangle$ on W. Since $W \cap W^\perp = \{0\}$, the kernel of A is trivial by our previous proposition, so A is invertible. Now, if $B = 0$, we would be done, since we would have that the basis vectors not in W are all orthogonal to the basis vectors in W, and therefore $V = W + W^\perp$. It turns out that we can change our original basis such that this is true.

Consider another basis matrix

$$B = \begin{pmatrix} I & Q \\ 0 & I \end{pmatrix},$$

with the intuition that the first "column" of B preserves the basis vectors in W, and the second "column" of B adds some basis vectors in W to every other vector in M.

Under this new basis, our new matrix for $\langle \cdot, \cdot \rangle$ becomes

$$B^* M B = \begin{pmatrix} I & 0 \\ Q^* & I \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I & Q \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & B \\ \sim & \sim \end{pmatrix} \begin{pmatrix} I & Q \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & AQ + B \\ \sim & \sim \end{pmatrix}.$$

Since A is invertible, we can let $Q = -A^{-1}B$, in which case the basis vectors in W are all orthogonal to the basis vectors not in W under our new basis, so we're done. $\square$

No lecture on Friday due to Veteran's day.

26 November 14, 2022

26.1 Classifying symmetric/Hermetian forms

Our goal today is to get a better handle on classifying symmetric / Hermetian forms. Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$. Let $\langle \cdot, \cdot \rangle$ be a symmetric (if V over $\mathbb{R}$) or Hermetian (if V over $\mathbb{C}$) form.

Theorem 26.1

(1) Let W be a subspace of V. Then

$$V = W \oplus W^\perp \iff \langle \cdot, \cdot \rangle \text{ non-deg on } W.$$

(2) Let W be a subspace of V. Then if $\langle \cdot, \cdot \rangle$ is non-deg on V and W, it is not deg on $W^\perp$.

We proved (1) last lecture. Here, we prove (2).

Proof. By (1), $V = W \oplus W^{\perp}$, so we can express the matrix for $\langle\cdot,\cdot\rangle$ as a block matrix

$$M = \left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array}\right)$$

where A is the matrix for $\langle\cdot,\cdot\rangle$ on W, and B is the matrix for $\langle\cdot,\cdot\rangle$ on $W^{\perp}$. Because our operator is non-deg on both V and W, M and A both have non-zero determinant. On the other hand, $\det M = \det A \cdot \det B$, so the determinant of B is also non-zero; therefore, our form is non-deg on $W^{\perp}$. □

Lemma 26.2

If $\langle\cdot,\cdot\rangle$ is not identically zero, then $\langle v,v\rangle \neq 0$ for some $v \in V$.

Proof. Suppose $\langle x,y\rangle \neq 0$. Note that replacing y with cy for any $c \in \mathbb{C}$ has the effect of rotating $\langle x,y\rangle$ in the complex plane by the argument of c. Therefore, we may assume $\langle x,y\rangle \in \mathbb{R}$, since we can rotate it to be real if necessary. So, we may assume $\langle x,y\rangle = \langle y,x\rangle$. Then,

$$\begin{aligned}\langle x+y, x+y\rangle &= \langle x,x\rangle + \langle y,y\rangle + \langle x,y\rangle + \langle y,x\rangle \\ &= \langle x,x\rangle + \langle y,y\rangle + \underbrace{2\langle x,y\rangle}_{\neq 0},\end{aligned}$$

which implies at least one of $\langle x+y,x+y\rangle$, $\langle x,x\rangle$, or $\langle y,y\rangle$ is not zero. □

Note that this proof implicitly uses the polarization identity that we used in Lectures 12 and 13, i.e., $\langle x,y\rangle = (\langle x+y,x+y\rangle - \langle x,x\rangle - \langle y,y\rangle)/2$. Here, we framed it differently because of the need to deal with Hermitian forms acting weird.

Theorem 26.3

V has an orthogonal basis $b_1,\ldots,b_n$, i.e., $\langle b_i,b_j\rangle = 0$ for all $i \neq j$.

Proof. We proceed with induction on $\dim V$.

If $\langle\cdot,\cdot\rangle$ is identically zero, we're done, since any basis is automatically orthogonal. Otherwise, there exists some $v \in V$ with $\langle v,v\rangle \neq 0$. Let $W = \operatorname{span} v$.

Then $\langle\cdot,\cdot\rangle$ is nondeg on the one-dimensional subspace W, which implies $V = W \oplus W^{\perp}$. By induction, $W^{\perp}$ has an orthogonal basis. Adding v into our basis, we're finished. □

Corollary 26.4

There exists an orthogonal basis $b_1, \dots, b_n$ with $|b_i| \in \{0, \pm 1\}$ for all i.

Proof. Take any orthogonal basis. For all b_i, if $|b_i| \neq 0$, we may replace it with $b_i/\sqrt{|\langle b_i, b_i\rangle|}$. □

Note that this does not necessarily eliminate the case when $\langle b_i, b_i \rangle = -1$, since

$$\left| b_i/\sqrt{|\langle b_i, b_i\rangle|} \right| = \langle b_i, b_i\rangle / |\langle b_i, b_i\rangle| \in \{\pm 1\}.$$

So, $\langle \cdot, \cdot \rangle$ has a basis with matrix

$$\begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & -I_{n_{-1}} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $n_t = \{\text{number of } b_i : |b_i| = t\}$, and $\dim V = n_0 + n_1 + n_{-1}$. In this basis,

$$\langle x, y \rangle = \sum_{i=1}^{n_1} \overline{x_i} y_i - \sum_{j=n_1+1}^{n_1+n_{-1}} \overline{x_j} y_j.$$

26.2 Sylvester's Law of Inertia

Definition 26.5

The **signature** of the form $\langle \cdot, \cdot \rangle$ is the triple (n_1, n_0, n_{-1}).

It turns out that this definition is well-defined by the following theorem.

Theorem 26.6 (Sylvester's Law of Inertia)

Each form $\langle \cdot, \cdot \rangle$ gives a unique signature which is independent of $b_1, \dots, b_n$.

Prof. Cohn notes that Sylvester is responsible for lots of naming schemes in Algebra, and if we ever come across something that sounds strange, it was probably named by Sylvester. For example, he came up with “determinant" in linear algebra, “syzygy" in abstract algebra, and “totient" in number theory, the latter of which is definitely just a made-up word that he used to sound fancy. The name of this particular theorem sounds crazy but he's using “inertia" to mean unchanging, i.e., being inert, so it's not so bad.

Before proving the theorem, here is an interesting connection between Sylvester's Law of Inertia and a pretty fundamental result from multivariable calculus.

Example 26.7

Sylvester's Law of Inertia implies the second derivative test. Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable.

Let x_0 be a critical point for f, i.e., $\nabla f(x_0) = 0$. From our previous discussions on the bilinear form representation of a second degree taylor polynomial, we can write

$$T_2(x) = f(x_0) + (x - x_0)^t H(x - x_0),$$

where H is the **Hessian matrix** given by

$$H = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \right)_{1 \le i,j \le n}.$$

Since partial derivatives commute, H is symmetric. Therefore, Sylvester's Law of Inertia implies that we can choose coordinates so that H is diagonal (with n_1 1s, n_{-1} -1s, and n_0 0s on the diagonal), in which the second degree taylor polynomial becomes

$$f(x_0) + \sum_{i=1}^{n_1} x_i^2 - \sum_{j=n_1+1}^{n_1+n_{-1}} x_j^2.$$

When $n_0 = 0$, we say that x_0 is a nondegenerate critical point of f, because our matrix in this case is nondegenerate, since it is invertible. Our form is positive semidefinite if and only if $n_{-1} = 0$, and positive definite if and only if $n_0 = n_{-1} = 0$.

Let's consider all nondegenerate critical points. When $n_{-1} = 0$, we have a local minimum, since shifting our coordinates can only increase the value of our taylor polynomial (positive definite). $n_{-1} = 1$ gives us a saddle point with one direction of decrease, i.e., the one coordinate corresponding to the negative eigenvalue. It can also be thought of as a mountain pass between two minima, or a potential barrier between stable states given a reaction pathway from one state to another. $n_{-1} = 2$ defines a saddle point with two directions of decrease, or a barrier to transform one reaction pathway to another. $n_{-1} = 3$ defines a barrier to transform one transformation of a reaction pathway to another, to another transformation of a reaction pathway to another. And so on.

Just like the normal second derivative test, we can't say anything about degenerate critical points, because we do not have information about how f behaves when we adjust coordinates corresponding to null eigenvalues. **Add image**

26.3 Proving Sylvester's Law of Inertia

Now we prove Sylvester's Law of Inertia.

Proof. Let $V = V_1 \oplus V_0 \oplus V_{-1}$, each corresponding to subspace spanned by the three types of basis vectors.

We first claim that V_1 is the max dimensional subspace on which our form is positive definite. V_1 is positive definite, because its matrix is just the identity. Any higher dimensional subspace has to intersect $V_0 \oplus V_{-1}$, but every vector in $V_0 \oplus V_{-1}$ has non-positive norm (i.e., negative semi-definite). Therefore, every subspace $W \subseteq V$ on which $\langle \cdot, \cdot \rangle$ is positive definite has $\dim W \leq n_1$.

An analogous argument yields that n_{-1} is the dimension of the largest subspace on which our form is negative definite, so every subspace $W \subseteq V$ on which $\langle \cdot, \cdot \rangle$ is negative definite has $\dim W \leq n_{-1}$.

So, suppose V has two signatures (n_1, n_0, n_{-1}) and (n_1', n_0', n_{-1}'). Applying our inequalities from the first signature to the second gives $n_1 \leq n_1'$ and $n_{-1} \leq n_{-1}'$. On the other hand, we can also apply them the other way, giving $n_1' \leq n_1$ and $n_{-1}' \leq n_{-1}$, so $n_1 = n_1'$ and $n_{-1} = n_{-1}'$. Finally, $n_0 = n_0'$ from $\dim V = n_0 + n_1 + n_{-1} = n_0' + n_1' + n_{-1}'$, so we're done. $\square$

27 November 16, 2022

27.1 Euclidean/Hermetian spaces

Last lecture, we classified all symmetric ($\mathbb{R}$) / Hermetian ($\mathbb{C}$) forms. In particular, we showed that there exists a basis such that the matrix for $\langle \cdot, \cdot \rangle$ is diagonal with entries $\pm 1, 0$. Moreover, Sylvester's Law of Inertia showed that each form gives a unique signature.

Definition 27.1
A **Euclidean space** is a finite dimensional real vector space with an inner product. A **Hermetian space** is a finite dimensional complex vector space with a Hermetian inner product.

By our classification, there always exists an orthonormal basis for these spaces for which the matrix corresponding to the inner product is the identity. If $n_{-1} > 0$, then the form can't be positive semidefinite, and if $n_0 > 0$, then the form can't be positive definite.

Therefore, inner products on these spaces are all the same as the usual dot product on $\mathbb{R}^n$ or Hermetian inner product on $\mathbb{C}^n$, up to a change of basis. In other words, $A \in \mathbb{C}^{n\times n}$ gives a Hermetian inner product $\iff A = B^* I_n B = B^* B$ for some $B \in GL_n(\mathbb{C})$. (B must be invertible, since changing the basis of a matrix is an isomorphism from our original vector space to itself.) The real case is analogous; we must have $A = B^t B$ for some $B \in GL_n(\mathbb{R})$.

Prof. Cohn addresses the fact that it's probably confusing why we continue to separate real and complex cases when, for everything that we have been doing so far, the exact same results hold with the exact same proofs. This will no longer be the case once we get to spectral theory.

Theorem 27.2 (Sylvester's criterion)
Suppose $A \in \mathbb{C}^{n\times n}$ and $A^* = A$. Then A is positive definite if and only if for all $k = 1, \dots, n$,

$$\det(A_{ij})_{1\le i,j\le k} > 0.$$

The real case holds analogously.

Proof. First, suppose that A is positive definite, so it is equal to B^*B for some $B \in GL_n(\mathbb{C})$ by our classification. Then,

$$\begin{aligned}\det A &= \det(B^*)\det(B)\\ &= \overline{\det(B)}\det(B)\\ &= |\det(B)|^2 > 0.\end{aligned}$$

Now consider what happens when we restrict our basis to the first k vectors. Since A was positive definite, the new $k \times k$ matrix formed by restricting our basis (in the upper left corner of A) must also be positive definite. Therefore, the same argument holds to show that the determinant for all such matrices are strictly positive.

Now, suppose $\det(A_{ij})_{1\le i,j\le k} > 0$ for all k. By assumption, $A_{11} > 0$. We can use row operations to zero out the rest of the first column in A, and then use column operations to zero out the rest of the first row in A.

This is equivalent to transforming A to a different basis. If we let $v_1, \ldots, v_n$ be our original basis, with v_1 corresponding to the row/column containing A_{11}, each row/column operation is equivalent to replacing the basis vector v_i with $v_i + \alpha v_1$, with α chosen such that $\langle v_i + \alpha v_1, v_1 \rangle = 0$. Therefore, our row/column operations has the effect of transforming $A \mapsto B^* AB$ for some $B \in GL_n(\mathbb{C})$.

$$\overbrace{\begin{pmatrix} A_{11} & \cdots & \cdots & \cdots & A_{1n} \\ \vdots & & & & \\ \vdots & & \ddots & & \\ \vdots & & & & \\ A_{n1} & & & & \end{pmatrix}}^{v_1 \qquad \cdots \qquad v_n} \mapsto \begin{pmatrix} A_{11} & 0 & \cdots & 0 & \cdots \\ 0 & & & & \\ \vdots & & M_k & & \\ 0 & & & & \\ \vdots & & & & \end{pmatrix}$$

Our transformed matrix still satisfies the determinant property, since determinants are invariant to row and column operations. Let M be the matrix in the lower right corner of A after this transformation. Now, for any $M_k = (M_{ij})_{1\le i,j\le k}$,

$$\det\left(\begin{pmatrix} A_{11} & 0 \\ 0 & M_k \end{pmatrix}\right) > 0 \implies \det M_k > 0,$$

where the first inequality holds because our transformed A still satisfies the determinant property. Therefore, M also satisfies the determinant property, so we can induct downward to show that there exists a basis for A under which it is diagonal with only positive elements along the diagonal. Since this is true if and only if A is positive definite, we are done. □

Misconception ⚠

Suppose $A \in \mathbb{C}^{n\times n}$ and $A^* = A$. Then A is positive semidefinite if and only if for all $k = 1, \ldots, n$,

$$\det(A_{ij})_{1\le i,j\le k} \ge 0.$$

The real case holds analogously.

This seems like it really should be true, but unfortunately, it's not. Here is a concrete counterexample:

$$A = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}.$$

A is not positive semidefinite, since $n_{-1} > 0$. On the other hand, both determinants are non-negative, so the claim above must be false. The problem stems from the fact that we are no longer able to induct downwards in the same way that we did before; $A_{11} \cdot \det M_k \ge 0$ tells us nothing about the sign of $\det M_k$ when $A_{11} = 0$.

27.2 Gram-Schmidt orthogonalization

Suppose V is Euclidean ($\mathbb{R}$) / Hermetian ($\mathbb{C}$). All subspaces are automatically non-degenerate; for any $W \subseteq V$, $w \in W$, and $w \ne 0$, then $\langle w, w\rangle > 0$.

Tangent: special relativity operates on four dimensional vectors with signature $(3,0,1)$ so you have to worry about degeneracy, something about light cones.

For any subspace $W \subseteq V$, let W have orthogonal basis $w_1, \ldots, w_k$. We can orthogonally project V to W as follows. Since our form is non degenerate, $V = W \oplus W^\perp$, so we can write

$$v = \underbrace{c_1 w_1 + \ldots + c_k w_k}_{\in W} + \underbrace{u}_{\in W^\perp} \qquad \forall v \in V.$$

Then, we can project v to $v - u$.

How do we get an orthonormal basis for W in the first place? We can use a process known as **Gram-Schmidt orthogonalization**. Start with any basis $w_1, \ldots, w_k$. For each i from 1 to k, replace

$$w_i \mapsto w_i - \sum_{j=1}^{i-1} \langle w_j, w_i\rangle w_j,$$

and then normalize w_i. After i replacements, $\{w_1, \ldots, w_i\}$ are orthonormal.

Here is a brief sketch for why this works. For each new basis vector w_i that we add, we can express as some linear combination of the other (transformed) basis vectors, $\alpha_1 w_1 + \ldots + \alpha_{i-1} w_{i-1}$. Then,

$$\begin{aligned}\langle w_i - \sum_{k=1}^{i-1} \alpha_k w_k, w_j \rangle = 0 &\iff \langle w_i, w_j \rangle - \alpha_j \langle w_j, w_j \rangle = 0 \\ &\iff \alpha_j = \frac{\langle w_i, w_j \rangle}{\langle w_j, w_j \rangle} = \langle w_i, w_j \rangle\end{aligned}$$

must hold for all $1 \le j \le i-1$. The last equality holds since we are normalizing our vectors after each iteration of the algorithm.

In matrix form, Gram-Schmidt orthogonalization takes a basis $v_1, \ldots, v_n$ of $\mathbb{R}^n$ and transforms it into an orthonormal basis $u_1, \ldots, u_n$ such that $u_i \in \operatorname{span}\{v_1, \ldots, v_i\}$ for all i.

In other words, if we have

$$V = \begin{pmatrix} | & & | \\ v_1 & \ldots & v_n \\ | & & | \end{pmatrix} \quad \text{and} \quad U = \begin{pmatrix} | & & | \\ u_1 & \ldots & u_n \\ | & & | \end{pmatrix},$$

then $U = VA$ for some upper triangular A (by Gram-Schmidt). This means that we can factor any $V \in GL_n(\mathbb{R})$ as UA^{-1} for some $U \in O_n(\mathbb{R})$ and A^{-1} upper triangular. This is called **QR factorization**.

28 November 18, 2022

28.1 The Spectral theorem

Standing assumptions: let V be a Hermetian space (has an inner product), which we proved last lecture can be viewed as $\mathbb{C}^n$ with the usual Hermetian form using any orthonormal basis of V.

Definition 28.1
The **adjoint** of any linear operator $T : V \to V$ is $T^* : V \to V$, which is defined so that if M is the matrix of T with respect to some basis, M^* is the matrix of T^* with respect to the same basis.

Adjoint linear operators are basis independent. If $M' = B^{-1}MB$, with $B \in U_n$, i.e., $B^{-1} = B^*$, then

$$(M')^* = B^* M^* B = B^{-1} M^* B,$$

so the original two linear operators are adjoint with respect to any basis.

Definition 28.2
T is **Hermetian** if $T^* = T$, **unitary** if $T^*T = I_n$, and **normal** if $TT^* = T^*T$. Hermetian and unitary matrices are also normal.

In general, $\langle Tv, w\rangle = (Mv)^* w = v^* M^* w = \langle v, T^* w\rangle$. Therefore, we can say that:

- Hermetian forms satisfy
$$\langle Tx, y\rangle = \langle x, Ty\rangle.$$
- Unitary forms satisfy
$$\langle Tx, Ty\rangle = \langle x, y\rangle.$$
- Normal forms satisfy
$$\langle Tx, Ty\rangle = \langle x, T^*Ty\rangle = \langle x, T^*Ty\rangle = \langle T^*x, T^*y\rangle.$$

Theorem 28.3 (Spectral theorem)
If $T : V \to V$ is normal, then there exists an orthonormal basis of V consisting of eigenvectors of T.

This is a pretty powerful statement. If T is normal, not only do we get that T is diagonalizable, but we also get that the basis under which its matrix is diagonal is orthonormal.

Further, we can show that the spectral theorem is as strong as possible; i.e., that the converse is true. Given an orthonormal basis of V consisting of eigenvectors of

T, then the matrix for T under this basis is given by $M = P^*\Lambda P$, where $P \in U_n$ and

$$\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

The fact that P is unitary implies that our basis is orthonormal, so each eigenvalue has magnitude 1. Now, if $M = P^*\Lambda P$, then $M^* = P^*\Lambda^* P$ and

$$\Lambda^* = \begin{pmatrix} \overline{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \overline{\lambda_n} \end{pmatrix}.$$

So, we must have

$$MM^* = P^*\Lambda\Lambda^* P = P^*\Lambda^*\Lambda P = M^*M,$$

since diagonal matrices commute, implying that M is normal. Let's first build up some key lemmas before proving the spectral theorem.

Definition 28.4
Subspace $W \subseteq V$ is ***T*-invariant** if $T(W) \subseteq W$.

Lemma 28.5
If W is T-invariant, then $W^\perp$ is T^*-invariant. The converse is also true.

Proof. Suppose $v \in W^\perp$. In other words, $\langle v, w\rangle = 0$ for all $w \in W$. Then, for all $w \in W$

$$\langle T^*v, w\rangle = \langle v, \underbrace{Tw}_{\in W}\rangle = 0.$$

Thus $T^*v \in W^\perp$.

This works in both directions. If $W^\perp$ is T^*-invariant, then $(W^\perp)^\perp = W$ is $(T^*)^* = T$-invariant. □

Our next lemma gives us a concrete relationship between the eigenvectors of T and T^*.

Lemma 28.6

If T is normal and $v \in V$ is an eigenvector of T with eigenvalue λ, then $T^*v = \overline{\lambda}v$.

Proof. First, we show that this works when $\lambda = 0$, i.e., $Tv = 0$. In this case,

$$0 = \langle Tv, Tv\rangle = \langle T^*v, T^*v\rangle,$$

implying $T^*v = 0$ by positive definiteness.

In general, suppose T has eigenvalue λ and corresponding eigenvector v. Define $S = T - \lambda I$, so that $Sv = 0$. Then, $S^* = T^* - \overline{\lambda}I$, and

$$\begin{aligned} SS^* &= (T - \lambda I)(T^* - \overline{\lambda}I) \\ &= TT^* - \lambda T^* - \overline{\lambda}T + |\lambda|^2 I \\ &= T^*T - \lambda T^* - \overline{\lambda}T + |\lambda|^2 I \\ &= (T^* - \overline{\lambda}I)(T - \lambda I) = S^*S, \end{aligned}$$

where the third equality follows from the fact that T is normal. This implies that S is normal, so we can apply the same argument for $\lambda = 0$ used above to get $S^*v = 0 \iff \overline{\lambda}$ is an eigenvalue for T^* with the same corresponding eigenvector. We're done! □

Now, we're ready to prove the spectral theorem.

Proof. We proceed with induction on $\dim V$.

There exists an eigenvector v of T with eigenvalue λ. Since we're working over the complex numbers, the characteristic polynomial always has a root, so this eigenvector is guaranteed. By our second Lemma,

$$Tv = \lambda v \implies T^*v = \overline{\lambda}v.$$

This implies that the one dimensional subspace $W = \mathbb{C}v$ is both T-invariant and T^*-invariant. By our first lemma, this means that $W^\perp$ is T^*-invariant and T-invariant as well.

Therefore, $T|_{W^\perp}$ (the restriction of T to $W^\perp$) is a linear operator on $W^\perp$, whose adjoint is $T^*|_{W^\perp}$. Since $W^\perp$ is the complement of a one-dimensional subspace of V,

$\dim W^\perp = \dim V - 1$. Moreover, T is still normal with respect to $W^\perp$, since T and T^* commutes. Therefore, there exists an orthonormal basis for $W^\perp$ with eigenvectors of $T|_{W^\perp}$ by our induction hypothesis. Now we can just rescale $v \in W$ and add it to the basis to finish. □

This is the first time that we see a divergence in the theory between real and complex vector spaces. The spectral theorem does not hold for Euclidean spaces, because we are no longer guaranteed that every operator T has a real eigenvalue (the first step of our proof). For example, one implication of the spectral theorem is that all unitary operators are conjugate to a diagonal operator. The analogous statement for Euclidean spaces is false; it is not true that all orthogonal operators are diagonalizable.

29 November 21, 2022

Today, we'll be talking about a few different applications of the Spectral theorem. In contrast to the Jordan canonical form, this theorem is surprisingly useful in lots of different ways.

Recall the complex and real variants of the spectral theorem:

Theorem 29.1 (Spectral theorem)
(Complex): Every normal operator on a Hermetian space has an orthonormal basis of vectors. (Real): Every symmetric operator on a Euclidean space has an orthonormal basis of vectors.

29.1 Quadric Hypersurfaces (Conic Sections)

A **quadric hypersurface** is a solution to any quadratic equation in n variables. Recall that functions representing these surfaces can be written as

$$f(x) = x^t A x + b^t x + c,$$

where $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$ and symmetric, $b \in \mathbb{R}^n$, and $c \in R$. We can say that A is symmetric, since $x^t A x = (x^t A x)^t = x^t A^t x$, so we can let $A \mapsto (A + A^t)/2$. The spectral theorem implies that A is diagonalizable.

When $n = 2$, our function is just a conic section. Since A is diagonalizable, we may write our equation as $ax^2 + bx^2 + cx + dy + e = 0$ (eliminating the xy term). If $a, b \neq 0$, we can complete the square, e.g.,

$$ax^2 + cx = \left(x + \frac{c}{2a}\right)^2 - \frac{c^2}{4a}.$$

So we can eliminate cross terms with rotation (basis in $O_2(\mathbb{R})$), and linear terms by translation whenever there is a non-zero quadratic term, which reduces the equation to something we might see in high school.

The same thing works for the general case. When A is invertible, we can completely eliminate its linear terms by translation:

$$\begin{aligned}
&(x + x_0)^t A(x + x_0) + b^t(x + x_0) + c = 0 \\
\iff\ &(x^t Ax + x^t Ax_0 + x_0^t Ax + x_0^t Ax_0) + (b^t x + b^t x_0) + c = 0 \\
\iff\ &\ldots + (2Ax_0 + b)^t x + \ldots = 0,
\end{aligned}$$

so we can let $x_0 = -(bA^{-1})/2$ to eliminate linear terms (as before, A is diagonal, so it only contributes squared terms).

29.2 Principal Component Analysis (PCA)

Let $X_1, \ldots, X_n$ be random variables. Let's say we have lots of data, so we can estimate means and covariances for these random variables. Without loss of generality, subtract the mean from each value, so $\mathbb{E}(X_i) = 0$.

Definition 29.2
The **covariance matrix** $M \in \mathbb{R}^{n\times n}$ is given by $M = (\mathbb{E}(X_i X_j))_{1\leq i,j\leq n}$. If we let $X = \begin{pmatrix} X_1 & \ldots & X_n \end{pmatrix}^t$, this is equivalent to $M = \mathbb{E}(XX^t)$.

Note that M is symmetric and positive semidefinite.

Lets consider a linear change of variables, $c_1X_1 + \ldots c_nX_n = c^tX$, where $c = \begin{pmatrix} c_1 & \ldots & c_n \end{pmatrix}^t$. The goal of PCA is to choose $c \neq 0$ to maximize the variance, without making c large, i.e., to maximize

$$\frac{\operatorname{var}(c^tX)}{c^tc} = \frac{\mathbb{E}(c^tX \cdot (c^tX)^t) - \mathbb{E}(c^tX)}{c^tc} = \frac{c^tMc}{c^tc}.$$

The idea is that by preserving the variance of our data, we do not "lose information". By the spectral theorem, there is an orthonormal basis $v_1, \dots, v_n$ for M such that $Mv_i = \lambda_i v_i$. Assume that they are ordered, so that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$.

Suppose $c = \sum \alpha_i v_i$, which implies $c^t c = \sum_i \alpha_i^2$, since $\{v_i\}$ are orthonormal. Then,

$$Mc = \sum_i \alpha_i \lambda_i v_i \implies c^t M c = \sum_i \alpha_i^2 \lambda_i \implies \frac{c^t M c}{c^t c} = \frac{\sum_i \alpha_i^2 \lambda_i}{\sum_i \alpha_i^2}$$

This is a weighted average of λ_i with weights proportional to α_i^2. In other words,

$$\frac{c^t M c}{c^t c} \leq \lambda_1,$$

with equality if and only if c is an eigenvector with eigenvalue λ_1.

This gives way to the following formula:

Proposition 29.3 (Rayleigh-Ritz Formula)

$$\lambda_k = \max_{\substack{c \in \mathbb{R}^n \setminus \{0\} \\ c \perp v_i \text{ for } i < k}} \left(\frac{c^t M c}{c^t c} \right)$$

Rayleigh-Ritz is often taken as a direct corollary of the spectral theorem. The formula tells us that the largest eigenvector maximizes our variance, as we have shown. The second largest eigenvector maximizes our variance, subject to the constraint that it is distinct (i.e., perpendicular) from the first eigenspace. And so on.

In practice, this is a very useful formula. For example, dimension reduction for large datasets. We want to throw out as much information as possible without losing information that we care about, in order to increase the interpretability of the data. PCA suggests that the best way to do this is to keep the top eigenvectors. If we have a million random variables, but we wish we only had fifty, then we should look at the random variables corresponding to the top fifty eigenvectors, and see if this gives us enough information.

In practice, a real life example is its use in facial recognition software. Each individual pixel in an image has some number of parameters corresponding to its RGB value. By using PCA, we can generate a set of "eigenfaces" corresponding to the most distinctive features of a particular set of faces, limiting the total number of dimensions required to e.g. train a neural network on a particular dataset.

29.3 Singular Value Decomposition (SVD)

SVD is a technique that can be applied to general data matrices in the same way that PCA can be applied to covariance matrices.

Say we have a very large data matrix

$$\mathbf{X} \in \mathbb{R}^{N \times n},$$

where $\mathbf{X}_{ij}$ represents the ith sample for X_j. As before, subtract means. Then,

$$M = \frac{1}{N}\mathbf{X}^t\mathbf{X}$$

can (sort of) be seen as a covariance matrix. The difference here from PCA is that this matrix can be rectangular, so we apply SVD instead.

Definition 29.4

For $A \in \mathbb{R}^{m \times n}$, its **Singular Value Decomposition (SVD)** is given by $A = V\sum W^t$, where $V \in O_m(\mathbb{R})$, $W \in O_n(\mathbb{R})$, and $\Sigma \in \mathbb{R}^{m \times n}$ is a diagonal matrix with diagonal entries $\sigma_1 \geq \sigma_2 \geq \ldots \geq 0$, which are called the **singular values**.

Proof of existence:

Proof.

$$A^tA = W\Sigma^t\Sigma W^t$$

$$AA^t = V\Sigma\Sigma^t V^t$$

These are both the spectral diagonalizations. $\Sigma\Sigma^t$ and $\Sigma^t\Sigma$ are both diagonal with eigenvalues $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq 0$ excluding some extra zeros when the dimensions are not balanced, so the existence of the SVD is equivalent to saying that A^tA and AA^t have the same set of eigenvalues.

A^tA is symmetric, so the spectral theorem implies that there exists an orthonormal basis $\{w_i\}$ of eigenvectors for A^tA. **bleh add later**. $\square$

29.4 Addendum: Moore-Penrose pseudoinverse

Addendum: this was not covered in lecture, but is another relevant application of SVD.

Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^n$. SVD gives us a way to construct a least-squares solution $x^* \in \mathbb{R}^m$ to $Ax = b$, i.e., that $|Ax^* - b|$ is minimized amongst all possible x. Another way to view this is that the least-squares solution is the best approximation to solving $Ax = b$, even when there is not necessarily an exact solution.

Let S be the column space of A, which is some subspace of $\mathbb{R}^n$. For any $v \in \mathbb{R}^n$, let $v|_S$ be the components of v restricted to S (i.e., a projection). A least squares solution to the equation $Ax = b$ will satisfy $(Ax^*)|_S = b|_S$, since $b|_S$ is necessarily in S and therefore reachable by A. Moreover, $b|_{S^\perp}$ is inaccessible to A, so this is the best we can do. Now, if we split A into its SVD, we find

$$b|_S = (Ax^*)|_S = (V\Sigma W^t x^*)|_S,$$

The restriction of $V\Sigma W^t x^*$ to S is the restriction of Σ to its non-zero diagonal elements. Therefore, if we define $\Sigma^\dagger \in \mathbb{R}^{m\times n}$ to be diagonal with

$$(\Sigma^\dagger)_{ii} = \begin{cases} 1/(\Sigma)_{ii} & (\Sigma)_{ii} \neq 0 \\ 0 & \text{else,} \end{cases}$$

then we find that $x^* = W\Sigma^\dagger V^t b = A^\dagger b$ is a least-squares solution.

Definition 29.5

The matrix $A^\dagger$ is called the **Moore-Penrose pseudoinverse** for A.

Every matrix in $\mathbb{R}^{m\times n}$ has this pseudoinverse given by its SVD, as shown above. Is this pseudoinverse unique? By our original stipulation $b|_S = (Ax^*)|_S$, we want $AA^\dagger|_S = I_n$ (when A is invertible, $A^\dagger = A^{-1}$, hence the name). If we additionally stipulate that $AA^\dagger A = A$, $A^\dagger AA^\dagger = A^\dagger$, and $AA^\dagger$, $A^\dagger A$ are both symmetric in order to remove the restriction from S, then the SVD construction for $A^\dagger$ is unique. Here is a quick sketch.

Proof. Suppose $A^\dagger$ and $B^\dagger$ were two pseudoinverses for A. Then $(AA^\dagger - AB^\dagger)^2 = A(A^\dagger - B^\dagger)A(A^\dagger - B^\dagger) = (A - A)(A^\dagger - B^\dagger) = 0$, and similarly $(A^\dagger A - B^\dagger A)^2 = 0$. But, since both are symmetric matrices, $AA^\dagger = AB^\dagger$ and $A^\dagger A = B^\dagger A$ by here. Thus $B^\dagger = B^\dagger AB^\dagger = A^\dagger AB^\dagger = A^\dagger AA^\dagger = A^\dagger$, as desired. $\square$

30 November 28, 2022

30.1 Geometries and linear groups

We have a number of different infinite matrix groups:

$$GL_n, SL_n, O_n, U_n, \dots$$

over infinite fields:

$$\mathbb{R} \text{ (reals)}, \mathbb{C} \text{ (complex)}, \mathbb{H} \text{ (quaternions)}, \dots$$

For the next few lectures, we'll spend some time focusing on better understanding these matrix groups over these fields. Which groups are out there? What is their structure?

In 1872, Felix Klein proposed the **Erlangen Program**: that studying geometries is somehow equivalent to studying different groups of symmetries.

One example that we have already studied in depth is the group of isometries:

$$\text{Euclidean geometries in } \mathbb{R}^n \longleftrightarrow M_n = O_n(\mathbb{R}) \ltimes \mathbb{R}^n.$$

Here are some others:

Example 30.1

$O_{p,q}(\mathbb{R})$ is the symmetry group of the form with signature $(p, 0, q)$. In other words,

$$O_{p,q}(\mathbb{R}) = \left\{ A \in \mathrm{GL}_n(\mathbb{R}) : \underbrace{A^t \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix} A = \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix}}_{A \text{ preserves the form w/ this matrix}} \right\}.$$

$O_{n,0}(\mathbb{R})$ is the usual orthogonal group. $O_{3,1}(\mathbb{R})$ plays an important role in Lorentz transformations and hyperbolic geometries (equations of the form $x^2 + y^2 + z^2 - t^2$).

Example 30.2

$$SO_2(\mathbb{R}) \cong \mathbb{R}/(2\pi\mathbb{Z}) \cong \{z \in \mathbb{C} : |z| = 1\}.$$

In other words, we can think about $SO_2(\mathbb{R})$ as the unit circle in the complex plane. Its noteworthy that this group is abelian, which is sort of lucky and does not generalize. This group is important to understand, but a bit simple, so let's see what happens in higher dimensions.

Example 30.3

What does SO_3 look like geometrically?

How to describe elements of $SO_3(\mathbb{R})$:

- Pick a rotation axis $a \in \mathbb{R}^3$, $|a| = 1$
- Pick an angle θ to rotate about our pole

This leads us to a few intuitive characterizations for the geometry of SO_3.

Definition 30.4

An ***n*-sphere** is a set of points

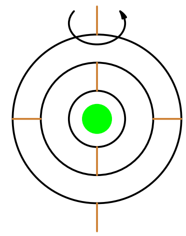
$$S^n = \{a \in \mathbb{R}^{n+1} : |a| = 1\}.$$

We define it this way because an n-sphere is a topologically n-dimensional object embedded in an $(n+1)$-dimensional space. For example, the surface of a 2-sphere is 2-dimensional, but the sphere itself is embedded in 3-dimensional space.

The characterization for SO_3 above amounts to choosing two elements, one from S^2, and one from S^1. Does this mean that $SO_3(\mathbb{R})$ is topologically equivalent to $S^2 \times S^1$? The answer is no:

- If $\theta = 0$, all a are equivalent (non-unique identity)
- (a, θ) is equal to $(-a, -\theta)$.

What if we try to instead define a from two angles (e.g., think polar coordinates)? This gives a description of any element in SO_3 as a combination of three angles, which are also called **Euler angles**. It turns out that SO_3 can't be characterized by $(S^1)^3$ either, for reasons that are similar to the first. We can try to visualize degenerate cases with an analogy to a concept in engineering:



Each ring here is called a **gimbal**, which is mounted to an axis by which it can rotate. The direction of rotation for the outermost gimbal is shown, with each subsequent gimbal having alternating perpendicular axes of rotation.

Now, we mount an object inside of our gimbals. Prof. Cohn used a frog, so we'll use a green circle here. This set of three gimbals should in principal allow the frog to achieve any orientation, with each orientation of the frog corresponding to a unique set of three angles of the gimbals. This would be analogous to SO_3 being topologically equivalent to $(S^1)^3$. In practice, there are some degenerate cases, like in the illustration, when all three gimbals are aligned so that the configuration is two-dimensional. In this configuration, there is no immediate way to rotate the frog with respect to a pole going directly into/out of the page, which is a problem known as **gimbal lock**. Ultimately, this shows that SO_3 can't be equivalent to $(S^1)^3$.

30.2 Fundamental Groups

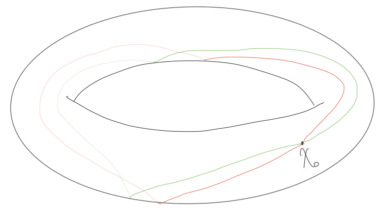
Definition 30.5

Given a topological space X, and some point $x_0 \in X$, the **fundamental group** for this point in space is defined as the set

$$\pi_1(X, x_0) = \{\text{loops based at } x_0 \text{ up to deformation}\},$$

with the composition group action.

A "loop" is a closed path. The fundamental group counts paths which are deformed in some way as the same path.



For example, the red and green loops on this torus are considered the same, because they have the same fundamental structure, they're just "deformed".

Example 30.6

$$\pi_1(S^1, \star) \cong \mathbb{Z}.$$

(We use $\star$ for our element because it does not matter where we start our loop). The identity of this group is "do nothing". Each element in the group is some integer number of rotations starting and ending at this point, so we have a single generator, going around the circle once.

Example 30.7

$$\pi_1(\text{torus}, \star) \cong \mathbb{Z}^2.$$

The two generators are the number of times the loop goes around the donut, and the number of times the loop goes around the ring of the donut.

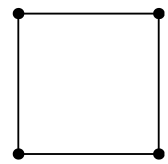
Example 30.8

$$\pi_1(SO_3(\mathbb{R}), I) \cong \mathbb{Z}/2\mathbb{Z}.$$

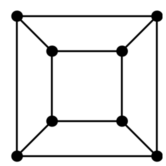
Prof. Cohn demonstrates this by holding his phone flat in one hand, and rotating it by a full loop, which results in his arm being twisted 180°. Composing this rotation with itself does not result in his arm being "twice as twisted", but instead it results in his arm returning to its normal position. In this way, there is no way for his arm to become "twice as twisted" once it returns to its identity position; rather, it just untwists itself, which gives some intuition for why the fundamental group is $\mathbb{Z}/2\mathbb{Z}$.

30.3 Cubes in $\mathbb{R}^n$

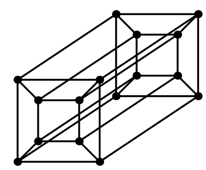
$n = 1$



$n = 2$



$n = 3$



$n = 4$

Moving up from one dimension to the next can be thought of as shifting over a copy of the current cube, and then connecting corresponding vertices between the original and the copy. Each time we do this, the number of vertices doubles. The number of facets (i.e., the number of $(n-1)$-dimensional faces) increases by 2, because we get one facet for each of the current facets when the copy is shifted, as well as a new "front" and "back" end of the new cube.

In other words, an n-dimensional cube has 2^n vertices and $2n$ facets. If its side length is 1, then its diameter is $\sqrt{1^2 + 1^2 + \ldots + 1^2} = \sqrt{n}$. Note the exponential growth in the number of vertices, compared to the sublinear growth of the diameter of the cube. A cube in $n = 10^6$ dimensions has a really, really large number of vertices, but a diameter of only 1000, which becomes tricky to visualize, especially given the fact that hypercubes must be convex.

31 November 30, 2022

31.1 Spheres

Recall from last time the definition for an $(n-1)$-sphere:

Definition 31.1

$$S^{n-1} = \{s \in \mathbb{R}^n : |x| = 1\}.$$

The naming convention is due to the fact that a sphere is an $(n-1)$ dimensional surface embedded in n dimensions. If we allow $|x| \leq 1$, instead of $|x| = 1$, then we call this a **ball**, where the volume inside the sphere is filled as well.

How does volume work in $\mathbb{R}^n$? Consider a ball with radius r, and another ball centered at the same point with radius $0.9r$. The ratio of the volumes of the two balls is $(0.9r)^n/r^n = 0.9^n$, which tends towards 0 as $n \to \infty$. This tells us that in higher dimensions, the volume of a ball is concentrated near its boundary. The rate at which volume packs into a shell near the boundary with increasing n is given by

$$\lim_{n\to\infty}\left(1-\frac{c}{n}\right)^n = e^{-c},$$

so the same amount of volume is packed into a shell with width c/n as n grows large.

31.2 Special unitary group SU_2

Recall the definition for the special unitary group SU_2:

Definition 31.2

$$SU_2 = \{A \in \mathbb{C}^{2\times 2} : A^*A = I_2, \det A = 1\}.$$

This may seem unrelated to SO_3, but it is ultimately helpful to understand, and we'll get back to SO_3 later. Let's try to characterize elements of SU_2. Given

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in SU_2,$$

we know $\alpha\delta - \beta\gamma = 1$, and

$$\begin{pmatrix} \overline{\alpha} & \overline{\gamma} \\ \overline{\beta} & \overline{\delta} \end{pmatrix} = A^* = A^{-1} = \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix}.$$

Therefore, $\overline{\alpha} = \delta$, and $\overline{\delta} = -\beta$, so

$$SU_2 = \left\{\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} : \alpha, \beta \in \mathbb{C}, |\alpha|^2 + |\beta|^2 = 1.\right\}$$

Now, each complex number has two degrees of freedom:

$$\alpha = w + ix$$
$$\beta = y + iz,$$

for $w, x, y, z \in \mathbb{R}$. This is four-dimensional, but we also have one more restriction: $|\alpha|^2 + |\beta|^2 = 1 \implies w^2 + x^2 + y^2 + z^2 = 1$. So, we end up with a 3-sphere (with three degrees of freedom):

$$S^3 = \{(w, x, y, z) \in \mathbb{R}^4 : w^2 + x^2 + y^2 + z^2 = 1\}.$$

Example 31.3
When is S^k a group?

$S^0 = \{\pm 1\}$ is a group. We showed last lecture that S^1 is topologically equivalent to SO_2. The group representation is given by $S^1 = \{z \in \mathbb{C} : |z| = 1\}$, which is just the unit circle. We've also just shown that S^3 is topologically equivalent to SU_2. It turns out that these are the only three examples. Now, how would we write S^3 as a group?

31.3 Quaternions

Definition 31.4
The **quaternions** are a vector space

$$\mathbb{H} = \mathbb{R} \oplus \mathbb{R}_i \oplus \mathbb{R}_j \oplus \mathbb{R}_k,$$

i.e., with basis elements $1, \hat{\imath}, \hat{\jmath}, \hat{k}$. The interaction between each basis element is given by $\hat{\imath}^2 = \hat{\jmath}^2 = \hat{k}^2 = -1$, and

$$\hat{\imath}\hat{\jmath} = \hat{k} = -\hat{\jmath}\hat{\imath}$$
$$\hat{\jmath}\hat{k} = \hat{\imath} = -\hat{k}\hat{\jmath}$$
$$\hat{k}\hat{\imath} = \hat{\jmath} = -\hat{\imath}\hat{k}.$$

The $\mathbb{H}$ stands for **Hamilton**, after William Rowan Hamilton, who first described quaternions.

If $\alpha = w + x\hat{\imath} + y\hat{\jmath} + z\hat{k}$, then its length is $|\alpha| = \sqrt{x^2+y^2+z^2+w^2}$. It turns out that S^3 is just the group of unit quaternions, i.e.,

$$S^3 = \{\alpha \in \mathbb{H} : |\alpha| = 1\}.$$

Given $\alpha = w + ix$, $\beta = y + iz$,

$$\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & -\overline{\alpha} \end{pmatrix} = w \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + x \cdot \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} + y \cdot \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + z \cdot \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

for any element in SU_2. We can create a correspondence

$$(1, \hat{\imath}, \hat{\jmath}, \hat{k}) \longleftrightarrow \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \right)$$

so that SU_2 is just $w + x\hat{\imath} + y\hat{\jmath} + z\hat{k}$, i.e., the group of unit quaternions. Its easy to check that these basis elements work.

Example 31.5

In what ways are quaternions similar to $\mathbb{C}$?

In the same way that complex numbers generalize $\mathbb{R}$ by introducing another degree of freedom, quaternions do the same for $\mathbb{C}$ by "combining" complex numbers.

For $\alpha = w + x\hat{\imath} + y\hat{\jmath} + z\hat{k} \in \mathbb{H}$, we define the **quaternion conjugate** $\overline{\alpha} = w - x\hat{\imath} - y\hat{\jmath} - z\hat{k}$. Conjugates for quaternions and complex numbers behave similarly:

$$\begin{aligned} \overline{\alpha + \beta} &= \overline{\alpha} + \overline{\beta} \\ \overline{\alpha\beta} &= \overline{\beta}\,\overline{\alpha} \end{aligned}$$

Unlike complex multiplication, quaternion multiplication is not commutative, so we have to be careful.

Lengths multiply as usual. Note $|\alpha|^2 = \alpha\overline{\alpha} = w^2 + x^2 + y^2 + z^2$, so

$$\begin{aligned} |\alpha\beta|^2 &= \alpha\beta\overline{\alpha\beta} = \alpha(\beta\overline{\beta})\overline{\alpha} \\ &= \alpha\overline{\alpha}\beta\overline{\beta} = |\alpha|^2|\beta|^2, \end{aligned}$$

where the third equality is true by the fact that $(\beta\overline{\beta}) \in \mathbb{R}$, so it commutes normally. This shows that $|\alpha\beta| = |\alpha||\beta|$.

31.4 $\mathbb{H}$ is a Division Algebra

Given $\alpha \in \mathbb{H}$ and $\alpha \neq 0$, then α has an inverse given by

$$\alpha^{-1} = \frac{\overline{\alpha}}{|\alpha|^2},$$

since $\alpha\alpha^{-1} = \alpha\overline{\alpha}/|\alpha|^2 = 1$. In this way, $\mathbb{H}$ is just like a normal field: addition commutes, is associative, and has an inverse; both distributive laws hold; additive and multiplicative identites both exist; multiplication is associative and we've shown that multiplication has an inverse. The only thing missing is that multiplication is not commutative.

We call $\mathbb{H}$ a **division algebra**, or alternatively a **skew field**. Which division algebras contain $\mathbb{R}$? It turns out that $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$ are the only examples. Sometimes, the definition of a division algebra also includes non-associative algebras, in which $\mathbb{O}$ (the octonians) also contains $\mathbb{R}$.

Example 31.6

What does $|\alpha||\beta|^2 = |\alpha\beta|^2$ look like over the different division algebras containing $\mathbb{R}$?

Over $\mathbb{C}$, if we let $\alpha = w + ix$ and $\beta = y + iz$, this amounts to

$$(w^2 + x^2)(y^2 + z^2) = (wy - xz)^2 + (wz + xy)^2.$$

Over $\mathbb{H}$,

$$\begin{aligned}(a_1^2 + a_2^2 + a_3^2 + a_4^2)(b_1^2 + b_2^2 + b_3^2 + b_4^2) &= (a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4)^2 \\ &+ (a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3)^2 \\ &+ (a_1b_3 + a_3b_1 + a_2b_4 - a_4b_2)^2 \\ &+ (a_1b_4 + a_4b_1 + a_2b_3 - a_3b_2)^2.\end{aligned}$$

We can do the same thing over $\mathbb{O}$, but its pretty nasty. Note that the right hand side of each expression is a bilinear form in the original input vectors α, β. These so-called bilinear identities imply that any product of two sums of n squares is also the sum of n squares itself, at least for $n = 1,2,4,8$. This fails when e.g. $n = 3$:

$$(1^1 + 1^1 + 1^1)(2^2 + 1^2 + 0^2) = 15 \neq x^2 + y^2 + z^2 \quad \text{for } x, y, z \in \mathbb{Z}.$$

Theorem 31.7 (Hurwitz)
A bilinear identity for sums of n squares exist only when $n = 1,2,4,8$.

What does linear algebra look like over a skew field? It's possible, but requires care, and we'll cover this a bit later. Prof. Cohn also promises to eventually connect all of this back to SO_3.

32 December 2, 2022

32.1 Linear algebra over a skew field

Last time, we found a nice characterization for SU_2:

$$SU_2 = \left\{ \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & -\overline{\alpha} \end{pmatrix} : \alpha, \beta \in \mathbb{C}, |\alpha|^2 + |\beta|^2 = 1 \right\} \cong \{\alpha \in \mathbb{H} : |\alpha| = 1\}.$$

We called $\mathbb{H}$ a **skew field** because it satisfies all the properties of a normal field, with the exception of commutative multiplication. Let's attempt to do linear algebra over $\mathbb{H}$.

$\mathbb{H}$ can act on $\mathbb{H}^n$ by left or right multiplication. Let's choose left scalar multiplication. Consider the map given by $x \mapsto \alpha x = f(x)$ for $x \in \mathbb{H}$. This is (generally) not a linear transformation. In order for our map to be a linear transformation, we need $f(x + y) = f(x) + f(y)$, which we have. We also need $f(\beta x) = \beta(f(x)) \iff \alpha(\beta x) = \beta(\alpha x)$. But by letting $x = 1$ this is only true when $\alpha\beta = \beta\alpha$. Since $\mathbb{H}$ isn't commutative, this does not necessarily hold.

On the other hand, let's see what happens when $\mathbb{H}$ acts on $\mathbb{H}^n$ by right multiplication. Let $f(x) = x\alpha$. Now, since $\mathbb{H}$ is associative, $(\beta x)\alpha = \beta(x\alpha)$, so $f(\beta x) = \beta f(x)$, so we have a linear transformation.

Why are quaternions important? Philosophically, one-third of mathematics is controlled by quaternions. For instance, consider the classification of compact Lie groups (Prof. Cohn emphasizes the Lie should be pronounced as "lee"). A lie group is both a group and a manifold. A manifold is a space that looks locally like $\mathbb{R}^n$. Compact approximately means closed and bounded, which means that every sequence of points has a convergence subsequence.

Theorem 32.1
Compact, connected Lie groups are always a quotient by a finite central subgroup of some product of groups isomorphic to

- S^1
- Spin_n ($\text{Spin}_n/\{\pm 1\} \cong SO_n$)
- SU_n
- The quaternionic unitary group $U_n(\mathbb{H}) \subseteq GL_n(\mathbb{H})$, defined as the set of matrices which preserves the quaternionic Hermetian form (we will define this later)
- G_2, F_4, E_6, E_7, E_8

This is imprecise, but the idea is that $\mathbb{R}, \mathbb{C}$, and $\mathbb{H}$ contribute equally towards generating linear groups; therefore, $\mathbb{H}$ is just as important and useful as the other two, even if it just seems more abstract.

32.2 Rotations in $\mathbb{R}^4$

We can also view $\mathbb{H}$ as $\mathbb{R}^4$, which will ultimately lead us back to SO_3 and the reason why we started looking at quaternions in the first place. So far, we have a natural way to compute the norm of any $\alpha \in \mathbb{H}$, so just like $\mathbb{R}$ and $\mathbb{C}$, we should also be able to construct a quaternion Hermetian form. In particular, let

$$\begin{aligned}
\langle \alpha, \beta \rangle &= \text{Re}(\alpha\overline{\beta}) \\
&= \text{Re}((\alpha_1 + \alpha_2\hat{\imath} + \alpha_3\hat{\jmath} + \alpha_4\hat{k})(\beta_1 - \beta_2\hat{\imath} - \beta_3\hat{\jmath} - \beta_4\hat{k})) \\
&= \alpha_1\beta_1 + \alpha_2\beta_2 + \ldots, \alpha_4\beta_4.
\end{aligned}$$

Proposition 32.2
If $|\alpha| = 1$, then $\alpha, \hat{\imath}\alpha, \hat{\jmath}\alpha, \hat{k}\alpha$ is an orthonormal basis of $\mathbb{H}$.

Proof. For $\hat{\imath}\alpha$ and $\hat{\jmath}\alpha$,

$$\langle \hat{\imath}\alpha, \hat{\jmath}\alpha \rangle = \mathrm{Re}(\hat{\imath}\alpha\overline{\alpha}(-\hat{\jmath})) = \mathrm{Re}(\hat{\imath}|\alpha|^2(-\hat{\jmath})) = \mathrm{Re}(-\hat{\imath}\hat{\jmath}) = \mathrm{Re}(-\hat{k}) = 0.$$

Similar calculations hold for the other pairs of vectors. □

The consequence of this is that $x \mapsto x\alpha$ and $x \mapsto \alpha x$ are orthogonal transformations of $\mathbb{R}^4$ if $|\alpha| = 1$ (we view $\mathbb{H}$ as $\mathbb{R}^4$). The same holds for $x \mapsto \overline{x}\alpha$ and $x \mapsto \alpha\overline{x}$. This gives us the following:

Theorem 32.3
$\forall \alpha, \beta \in \mathbb{H}$ with $|\alpha| = |\beta| = 1$, $x \mapsto \alpha x \beta$ and $x \mapsto \alpha\overline{x}\beta$ are in $O_4(\mathbb{R})$.

This is a direct result of our previous Proposition. We can think of $x \mapsto \alpha x \beta$ as our "orientation-preserving" maps, i.e., they are composed of an even number of reflections and have determinant $+1$. We can think of $x \mapsto \alpha\overline{x}\beta$ as our "orientation-reversing" maps, i.e., they are composed of an odd number of reflections and have determinant -1.

Theorem 32.4
And this is all of $O_4(\mathbb{R})$.

Proof. We take for granted that $O_4(\mathbb{R})$ is generated by reflections in hyperplanes, which we proved in Problem Set 6. Now, the key idea is that

$$x \mapsto -\alpha\overline{x}\alpha$$

is a reflection in the hyperplane perpendicular to α. We can verify this:

$$\begin{aligned} \alpha &\mapsto (-\alpha\overline{\alpha}\alpha) = (-|\alpha|^2\alpha) = -\alpha \\ \hat{\imath}\alpha &\mapsto (-\alpha\overline{\hat{\imath}\alpha}\alpha) = (-\alpha\overline{\alpha}(-\hat{\imath})\alpha) = \hat{\imath}\alpha, \end{aligned}$$

and similarly for $\hat{\jmath}$, $\hat{k}$ (sign preserved). Since $O_4(\mathbb{R})$ is generated by reflections, any element is equal to some

$$\left(x \mapsto \prod_i (-\alpha_i \overline{x} \alpha_i)\right) = \left(x \mapsto -\alpha_1 \left(\prod_{i>1} (-\alpha_i \overline{x} \alpha_i)\right) \alpha_1\right) = \dots$$

where a chain of unit α grows on each side of x and the orientation of x flips for each new α (this is somewhat abusive notation, we're mixing up composition and actual multiplication). Therefore, every element in $O_4(\mathbb{R})$ is $x \mapsto \alpha \overline{x} \beta$ or $x \mapsto \alpha x \beta$ for some unit $\alpha, \beta \in \mathbb{H}$. $\square$

Note that whenever we end up with a mapping of the form $x \mapsto \alpha x \beta$, this means that we composed an even number of reflections, so these maps make up $SO_4(\mathbb{R})$. On the other hand, mappings of the form $x \mapsto \alpha \overline{x} \beta$ composed an odd number of reflections, so these are the elements in $O_4(\mathbb{R})$ not in $SO_4(\mathbb{R})$.

32.3 SU_2 is a double cover for SO_3

How do we connect this back to $3d$ rotations? Let's look at the imaginary quaternions, i.e., $\mathbb{R}_i \oplus \mathbb{R}_j \oplus \mathbb{R}_k \cong \mathbb{R}^3$.

A $3d$ rotation of $\mathbb{R}_i \oplus \mathbb{R}_j \oplus \mathbb{R}_k$ is equivalent to a $4d$ rotation of $\mathbb{H}$ which fixes 1, i.e., a mapping $x \mapsto \alpha x \beta$ such that $\alpha \cdot 1 \cdot \beta = 1 \implies \beta = \alpha^{-1}$. SO_3 is just conjugation by unit quaternions! Kind of. All this implies is that every rotation can be represented by some unit quaternion, so we have a homomorphism

$$\begin{aligned} \gamma : SU_2 \cong \{\alpha \in \mathbb{H} : |\alpha| = 1\} &\to SO_3 \\ \alpha &\mapsto (x \mapsto \alpha x \alpha^{-1}). \end{aligned}$$

This implies $SO_3 \cong SU_2/\ker\gamma$. Now, $\alpha \in \ker(\gamma)$ if and only if $\alpha x = x\alpha$ for all unit $x \in \mathbb{H}$. If $\alpha = \alpha_1 + \alpha_2\hat{\imath} + \alpha_3\hat{\jmath} + \alpha_4\hat{k}$, then $x = \hat{\imath}$ implies

$$\alpha_1\hat{\imath} + \alpha_2(-1) + \alpha_3(-\hat{k}) + \alpha_4(\hat{\jmath}) = \alpha_1\hat{\imath} + \alpha_2(-1) + \alpha_3(\hat{k}) + \alpha_4(\hat{\jmath}),$$

so $\alpha_3 = \alpha_4 = 0$. Repeating this calculation for the other basis vectors gives $\mathrm{Im}(\alpha) = 0$, so $\alpha = \pm 1$ and $SO_3 \cong SU_2/\{\pm 1\}$.

What this means is that SO_3 is *almost* conjugation by unit quaternions, it just can't distinguish between $\pm\alpha$, where α is the element we're conjugating with. For this reason, we call SU_2 a **double cover** of SO_3, and for most applications this is good enough. Prof. Cohn talks a about a deep connection to electron spin, but I don't know enough about physics to understand.

33 December 5, 2022

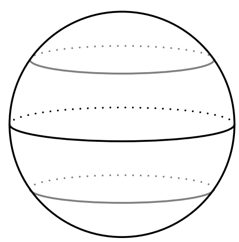
33.1 Spheres of latitude are conjugacy classes in SU_2

Recall the two ways that we can formulate the special unitary group of 2×2 matrices:

Definition 33.1

$$SU_2 = \left\{ \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} : \alpha, \beta \in \mathbb{C}, |\alpha|^2 + |\beta|^2 = 1 \right\} = \{\gamma \in \mathbb{H} : |\gamma| = 1\}.$$

We can represent SU_2 as a 3-sphere embedded in four dimensions.



We attempt to visualize the group in three-dimensions. In the above depiction, the north and south pole represents the identity and minus the identity. If we connect the identity and minus the identity and then cut the sphere with a plane perpendicular to this line, we get a two-sphere. This is represented in the above diagram by the horizontal lines of latitude.

To make things precise, each sphere of latitude encompasses γ with fixed real part. We can imagine that 1 is the basis vector that differentiates each different sphere of latitude (hence the identity elements at the poles). Each sphere of latitude then represents all the quaternions you can get varying the other three dimensions.

It turns out that each sphere of latitude also represents a set of elements with fixed trace, since

$$\langle \gamma, 1\rangle = \operatorname{Re}\gamma\overline{1} = \operatorname{Re}\gamma = \frac{1}{2}\operatorname{Tr}\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}.$$

This is useful due to the following theorem:

Theorem 33.2

The conjugacy classes in SU_2 are

$$\{A \in SU_2, \operatorname{Tr} A = c\}$$

for $-2 \le c \le 2$.

Proof. First, when we conjugate by any $B \in SU_2$, $\operatorname{Tr}(B^{-1}AB) = \operatorname{Tr}(BB^{-1}A) = \operatorname{Tr}(A)$, so trace is an invariant under conjugation.

For the opposite direction, suppose $\operatorname{Tr} A = c$. Note

$$\operatorname{Tr}\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} = 2\operatorname{Re}\alpha.$$

Since $|\alpha|^2 \le 1$, this forces $-2 \le c \le 2$.

The eigenvalues of A are the roots of $x^2 - cx + 1$, since $A \in SU_2$, so $\lambda, \overline{\lambda}$ are uniquely determined by c. By the spectral theorem, there exists $B \in U_2$ such that

$$B^* AB = B^{-1}AB = \begin{pmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{pmatrix}.$$

This shows that A is conjugate to the eigenvalue matrix in U_2, but not necessarily SU_2. To fix this, let $\det B = \delta$, and scale

$$B' = B\begin{pmatrix} \delta^{-1/2} & 0 \\ 0 & \delta^{-1/2} \end{pmatrix}.$$

Now $\det B' = 1$, so $B' \in SU_2$. Then,

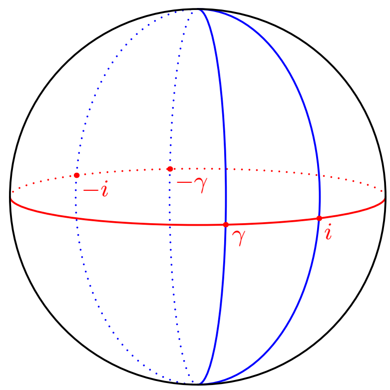
$$(B')^{-1}AB' = \begin{pmatrix} \delta^{1/2} & 0 \\ 0 & \delta^{1/2} \end{pmatrix} B^{-1}AB \begin{pmatrix} \delta^{-1/2} & 0 \\ 0 & \delta^{-1/2} \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{pmatrix},$$

so any A with trace c is conjugate to its uniquely determined eigenvalue matrix in

SU_2, as desired. $\square$

33.2 One-parameter subgroups in SU_2

How about lines of longitude? These turn out to be **one-parameter subgroups**. We'll look at this quaternionically.



The equator of our three-dimensional model of SU_2 contains all purely imaginary $\gamma \in \mathbb{H}$ (i.e., $\operatorname{Re}\gamma = 0$). By the previous theorem, this is a conjugacy class. Note that $\hat{\imath}$ lies on the equator, so consider the circle of longitude passing through $\pm\hat{\imath}, \pm 1$:

$$S = \{\cos\theta + i\sin\theta : \theta \in \mathbb{R}\}.$$

Note that $S \cong S^1 \subseteq \mathbb{C}$, so S is actually just a one-parameter subgroup of the usual complex numbers. If we were to substitute $\hat{\imath}$ with any other basis vector, it doesn't matter, we'll just get some space generated by two basis vectors, which is always the normal complex numbers.

For arbitrary γ with $\operatorname{Re}\gamma = 0$, we'll always get something conjugate to S, which is a subgroup of $\mathbb{C}$. In this case, our "one paramter" generating our line of longitude is γ.

One important faulty diagram moment: S is a 1-sphere (as are all the other lines of longitude), so it is just a normal circle, like how it looks on the diagram. On the other hand, our red equator is a 2-sphere of latitude. Even though they look the same in the diagram, they're not, which is just something we have to deal with given that we can't draw four-dimensional pictures.

33.3 Normal subgroups of SU_2 with more geometry

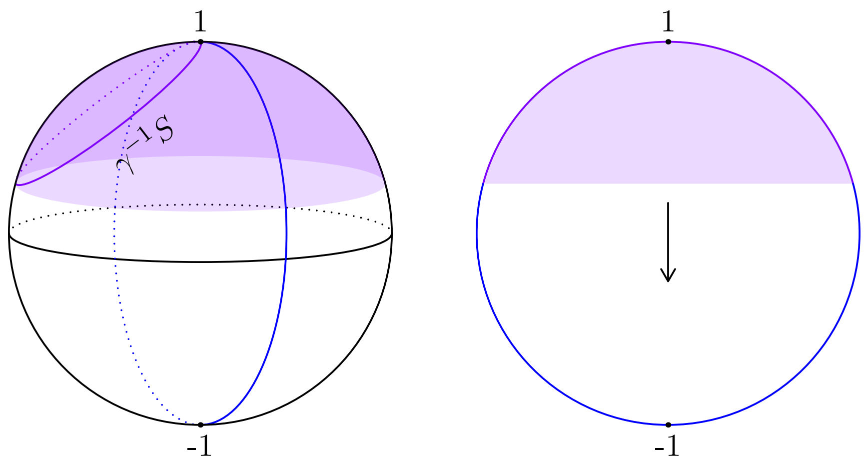
Theorem 33.3
The only normal subgroups of SU_2 are $\{1\}, \{\pm 1\}, SU_2$.

Corollary 33.4
$SU_2/\{\pm 1\} \cong SO_3$ is simple.

The corollary follows by the double-covering that we discussed last lecture. In particular, because there is a surjective homomorphism $\gamma : SU_2 \to SO_3$ with kernel $\{\pm 1\}$, the correspondence theorem gives a bijective mapping between normal subgroups of SU_2 containing $\{\pm 1\}$ and normal subgroups of SO_3. Since the only such normal subgroups of SU_2 are $\{\pm 1\}$ and SU_2, SO_3 is simple.

Now, we prove the main theorem.

Proof. Let G be a normal subgroup in SU_2. Then G is some union of conjugacy classes. The identities are point conjugacy classes, so $\{\pm 1\}$ works on its own. Now, if $G \not\subseteq \{\pm 1\}$, then G contains some other conjugacy class S.



Choose some $\gamma \in S$. Then $\gamma^{-1}S \subseteq G$, since the group is closed under multiplication. Note that $\gamma^{-1}S$ is a rotation of S (i.e., a 2-sphere) which also contains the identity. By continuity, this implies that G contains the conjugacy classes of all elements close to the identity. But now, if we can take any one-parameter subgroup, we can use the elements close to the identity to generate the rest of the subgroup by the normal group law. Since each one-parameter subgroup is a line of longitude, so-to-speak, it contains at least one element from every conjugacy class, so $G = SU_2$, as desired. $\square$

Definition 33.5
$G \subseteq \mathrm{GL}_n(\mathbb{C})$. A **one-parameter subgroup** of G is a homomorphism $\mathbb{R} \to G$ that's differentiable as a function from $\mathbb{R}$ to $\mathbb{C}^{n\times n}$.

$\varphi : \mathbb{R} \to G$ is a homomorphism and its differentiable. $\varphi(s+t) = \varphi(s)\varphi(t)$. Then $\varphi'(s+t) = \varphi'(s)\varphi(t)$, differentiating with respect to s. If we let $s = 0$, then $\varphi'(t) = \varphi'(0)\varphi(t) = A\varphi(t)$. This is a differential equation with solution $\varphi(t) = e^{At} = I + At + A^2t^2/2! + \ldots$. This is the unique solution with $\varphi(0) = I$.

$e^{S+T} = e^S e^T$ if $TS = ST$. Any two multiples of A commute with each other, so this gives a homomorphism.

$$A \longleftrightarrow \varphi'(0) \longleftrightarrow \text{tangent vector to } G \text{ at } 1.$$

Example 33.6
Examples.

$G = GL_n(\mathbb{C}) \to A \in \mathbb{C}^{n\times n}$ $GL_n(\mathbb{R}) \to A \in \mathbb{R}^{n\times n}$ $SL_n(\mathbb{R}) \to A \in \mathbb{R}^{n\times n}, \mathrm{Tr}\,A = 0$. $\det e^A = e^{\mathrm{Tr}\,A}$.

$O_n(\mathbb{R}) \to e^{sA}e^{sA^t} = I$.

This group gives something called the Lie algebra of Lie group G.

34 December 12, 2022

This is the first of two post final-exam lectures just for fun. (Post-lecture note: I got a little lost so the notes here aren't great)

34.1 Tensors

Let V be a finite-dimensional vector space over field F. Then $\mathrm{Hom}_F(V, W)$ is defined as the space of all F-linear transformations from V to W.

Definition 34.1

The **dual** $V^* = \mathrm{Hom}_F(V, F)$.

If $\dim V = n$, then $\dim V^* = n$. $V \cong V^*$.

Given non-degenerate form $\langle \cdot, \cdot \rangle$ on V, gives an isomorphism $V \to V^*$. $\overset{\psi}{v} \mapsto \overset{\psi}{f}$. $f(w) = \langle v, w \rangle$ for $w \in V$. Conversely, given $\varphi : V \to V^*$, $\langle v, w \rangle = \varphi(v)(w)$ is a non-degenerate bilinear form.

$V^{**} \cong V$ is canonically isomorphic to V. $V \to V^{**}$ given by $\overset{\psi}{v} \mapsto \overset{\psi}{ev_v}$, with $ev_v(f) = f(v)$ if $f \in V^*$.

Definition 34.2

The **tensor product** $V \otimes W$ (also sometimes notated $V \otimes_F W$, if you wish to emphasize the field) is the vector space spanned by $v \otimes w$ with $v \in V$, $w \in W$. All such $v \otimes w$ are called **pure tensors**. $\otimes$ is bilinear and obeys the following operations:

- $(v + v') \otimes w = v \otimes w + v' \otimes w$
- $v \otimes (w + w') = v \otimes w + v \otimes w'$
- $(\alpha v) \otimes w = \alpha(v \otimes w) = v \otimes (\alpha w)$.

Tensors is what you get when you impose these relations and no others.

We're interested in: bilinear maps

$$V \times W \to U,$$

which are linear in each coordinate. Intuition: tensors abstract this.

Consider the map $\varphi : V \times W \to V \otimes W$ with $(v, w) \mapsto v \otimes w$. φ is bilinear, because this is how the tensor product is defined. Then, the map $f : V \times W \to U$ is bilinear if and only if there exists $\overline{f} : V \otimes W \to U$ linear such that $\overline{f} \circ \varphi = f$.

$$\begin{tikzcd}
V \times W \arrow[rr] \arrow[dr, "\varphi"'] & & U \\
& V \otimes W \arrow[ur, "\overline{f}"'] &
\end{tikzcd}$$

$\dim V \otimes W = (\dim V)(\dim W)$. The idea is that if $v_1, \dots, v_n$ is the basis of V, and $w_1, \dots, w_m$ is a basis of W, then $v_i \otimes w_j$ is a basis of $V \otimes W$.

$\dim V = n$. An order 2 tensor is an element of $V \otimes V$, $V \otimes V^*$, or $V^* \otimes V^*$ corresponding to type $(2,0),(1,1),(0,2)$, respectively. In general, for order $r+s$, a type (r,s) is given by $V^{\otimes r} \otimes (V^*)^{\otimes s}$. We can think of V^* as inputs, and V as outputs.

35 December 14, 2022

This is the second of two post final-exam lectures just for fun. Last lecture

35.1 Representation Theory

Let G be a group and V a vector space over field F.

Definition 35.1
A **representation** of G on V is an action of G on V such that for all $g \in G$, $v \mapsto gv$ is a linear operator on v.

This is equivalent to a homomorphism $\rho : G \to GL(V) \subseteq \mathrm{Perm}(V)$, where $\rho(g)$ is our linear operator.

Why do we care about group representations? 1) They have lots of structure. 2) They're good for linear things, and non-linear things can often be approximated by linear things, so group representations are useful.

Today assume $F = \mathbb{C}$ and $\dim V < \infty$. $\mathrm{Hom}_G(V,W)$ is the set of linear transformations $TV \to W$ such that $T(gv) = gT(v)$ for all $g \in G, v \in V$.

It's really difficult to know at what point we've characterized all possible representations. This is an endeavor that is sort of pushing the limits of human knowledge right now. Let's look at some examples of representations.

Example 35.2
Trivial representation.

$V = \mathbb{C}$ (one dimensional), and $gv = v$ for all $g \in G, v \in V$.

Example 35.3
S_n acts on $\mathbb{C}^n$ via permutation matrices.
$GL_n(\mathbb{C})$ acts on $\mathbb{C}^n$ via matrix multiplication.

Example 35.4
S_n acts on $\mathbb{C}$ by $\pi v = \mathrm{sgn}(\pi)v$.

Example 35.5
$C_m = \langle g : g^m = 1\rangle$. For each j s.t. $0 \le j < m$, C_m acts on $\mathbb{C}$ by $g \cdot v = e^{2\pi j/m}v \implies g^k v = e^{2\pi ijk/m}v$.

Example 35.6
S_3 acts on $\mathbb{C}^2$.

We can view S_3 as symmetries of an equilateral triangle in $\mathbb{R}^2$. So there exists a homomorphism

$$\rho : S_3 \to GL_2(\mathbb{R}) \subseteq GL_2(\mathbb{C}).$$

In particular, this action also works on $\mathbb{R}^2$.

Definition 35.7
Given V, a G-representation, and subspace $W \subseteq V$, we say W is an **invariant subspace** if it is closed under G by its action. In this case we call W a subrepresentation.

Definition 35.8
V is **irreducible** if $V \neq \{0\}$ and its only subrepresentations are $\{0\}, V$.

Given V, W are G-representations. Then $V \oplus W$ is also a G-representation given by $g(v,w) = (gv, gw)$. The subrepresentations of $V \oplus W$ must at least include $0 \oplus 0$, $0 \oplus W$, $V \oplus 0$, $V \oplus W$.

Is every (fin-diml) representation a direct sum of irreducibles up to isomorphism? In general, the answer is no. For most fields F, it won't work, and it also won't work for some infinite G even if $F = \mathbb{C}$.

Consider $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, which is the usual 2×2 matrix that ruins things because it is not diagonalizable. It turns out that this matrix also helps work as a counterexample to the previous paragraph. In particular, $\mathbb{Z}$ has a 2-diml representation on

$\mathbb{C}^2$:

$$n \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^n \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + ny \\ y \end{pmatrix},$$

so the second coordinate is preserved and the first coordinate is linear in the second coordinate. What are the subrepresentations?

$$\{0\}, \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} : x \in \mathbb{C} \right\}.$$

(Since the first coordinate is linear in the second coordinate, closure only works when the second coordinate is zero). So this representation is not irreducible. On the other hand, it also can't be the direct sum of any two representations, otherwise there would be more subrepresentations.

The fact that such a simple representation breaks things is discouraging, but it turns out that forcing G to be finite and $F = \mathbb{C}$ makes things nice again.

Proposition 35.9

Given finite G and $F = \mathbb{C}$, then V is necessarily a direct sum of irreducibles up to isomorphism, given that $\dim V < \infty$.

Proof. Pick the G-invariant $\langle \cdot, \cdot \rangle$ on V by $\langle gv, gw \rangle = \langle v, w \rangle$ (we proved existence in Problem Set 10). $W \subseteq V$ is a subrepresentation when $GW = W$. Now, W being a subrepresentation implies $W^\perp$ is also a subrepresentation, since for any $v \in W^\perp$, $\langle v, w \rangle = 0 \iff \langle gv, gw \rangle = 0 \iff \langle gv, w \rangle = 0 \iff gv \in W^\perp$, which implies $V = W \oplus W^\perp$ is a direct sum of subrepresentations. We can keep applying this idea until V is the direct sum of irreducibles. $\square$

The G-invariant that we used gives something called a **unitary representation** for G, i.e., a representation V with Hermetian inner product that is G-invariant. This is equivalent to a homomorphism $\rho : G \to U_n$.

Lemma 35.10 (Schur's Lemma)

If V, W are irreducible G-reps, then $\mathrm{Hom}_G(V, W) = \{0\}$ if $V \neq W$ and $\mathrm{Hom}_G(V, V) = \{\lambda I : \lambda \in \mathbb{C}\}$.