18.701: Algebra I

Lecturer: Professor Mike Artin

Notes by: Andrew Lin

Fall 2018

1 September 5, 2018

(Professor Artin wants us to call him Mike.)

First of all, it's good for us to read through the syllabus. On the back page, there is a diagnostic problem that is pretty simple; we should have it done by Monday. It will not count for our grade, but it is required. Apparently the problem sets are hard: we should not expect to finish them quickly.

Fact 1

The two main topics of this semester will be **group theory** and **linear algebra**.

When Professor Artin was young, he wanted to learn the "general axioms." But it's better to use examples and use those to understand the axioms when we're first learning mathematics.

Definition 2

The **general linear group**, denoted GL_n, consists of the invertible $n \times n$ matrices with the operation of matrix multiplication.

(The definition of a **group** was given on the next day.) To make examples easier to write down, we'll take $n = 2$. Matrix multiplication looks like the following:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 5 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 2 \cdot 4 & 1 \cdot 5 + 2 \cdot 1 \\ 3 \cdot 1 + 4 \cdot 4 & 3 \cdot 5 + 4 \cdot 1 \end{bmatrix} = \begin{bmatrix} 9 & 7 \\ 19 & 19 \end{bmatrix}$$

The definition of matrix multiplication seems kind of complicated, but it turns out we can come up with a natural explanation. One way to explain this definition is by looking at column vectors, with matrices "acting from the left side:" if V is the space of 2-dimensional column vectors, we can treat our matrix A as a linear operator on V, where a vector $v \in V$ gets sent to Av.

Fact 3

Given two matrices A and B, it is generally not true that $AB = BA$. (For example, take the matrices $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$: they do not commute.) However, matrix multiplication is associative (that is, $(AB)C = A(BC)$), and we know this because we're just composing three transformations in the same order: C, then B, then A.

In this class, we'll generally deal with invertible matrices (because they make our group operations nicer). By the way, if we don't know this already, the inverse of a 2 by 2 matrix is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Definition 4

An element A of a group has **order** n if A^n is the identity element.

Example 5

Consider the matrix $A = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}$. Since $A^6 = I$, A is an element of GL_2 with order 6.

Elements can have infinite order as well, but it turns out the space of 2×2 matrices is nice:

Theorem 6

If entries of a 2×2 matrix A are rational, and the order is finite, it must be 1, 2, 3, 4, or 6.

(We'll prove this much later on.) Professor Artin likes to use dots instead of zeros in matrices because they look cleaner, but I will not do this in the notes.

Example 7

The following matrix just cycles the indices of a vector, so it has order n if it is n-dimensional.

$$A = \begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \\ 1 & & & & \end{bmatrix} \Bigg\} n$$

There are three kinds of **elementary matrices**, which basically change the identity matrix by a tiny bit. (This is the idea of **row-reducing**.) We have the matrices

$$\begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ a & 1 \end{bmatrix}$$

which add a times the second row to the first row and vice versa, the matrices

$$\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}$$

which multiplies one of the two rows by c, and the matrix

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

which swaps the two rows.

Theorem 8
The elementary matrices generate GL_2. In other words, every A in GL_2 is a product of the above elementary matrices.

Let's say we start with an arbitrary matrix, say

$$M = \begin{bmatrix} 8 & 5 \\ 4 & 6 \end{bmatrix}$$

It's hard to randomly find matrices $E_1, E_2, \cdots$ that multiply to M. Instead, we should work backwards, and try to write

$$E_k \cdots E_2 E_1 A = I.$$

Then we know that $A = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$, since all the elementary matrices have elementary matrix inverses. I'm not going to include how to do this here, but we basically work one column at a time and try to get the matrix to the identity.

2 September 7, 2018

Definition 9
Given two sets S and T, the **product set** or **Cartesian product** $S \times T$ is defined to be the set of all ordered pairs $\{s, t\}$, where $s \in S, t \in T$.

Now we want to define (essentially) a binary operation:

Definition 10
A **law of composition** on S is a law $S \times S \overset{\text{law}}{\to} S$ sending $(a, b) \to c$ (for $a, b, c \in S$). We often refer to c as $a \star b$, ab, $a + b$, or something else of that nature.

Professor Artin forgot to define a group last time, so we'll actually do that now:

Definition 11
A **group** G is a set with a law of composition that also satisfies the following axioms:

- Associativity holds; that is, $(a \star b) \star c = a \star (b \star c)$ for all $a, b, c \in G$.
- There exists an identity $i \in G$ such that for all $a \in G$, $a \star i = i \star a = a$.
- Inverses exist; that is, for all $a \in G$, there exists an $\alpha \in G$ such that $a\alpha = \alpha a = i$.

From now on, we'll refer to the group law as $a \star b = ab$, the identity element as 1, and the inverse of a as a^{-1}.

Example 12
GL_n is a group, with group law (or binary operation) being matrix multiplication.

We should keep in mind that the identity is not always the number 1; for example, the identity for GL_n is the identity matrix I. (This is just symbolic notation.)

Definition 13

A **permutation** of a set S is a **bijective map** $S \xrightarrow{p} S$. (Importantly, S does not need to be finite.)

Recall that a map $S \xrightarrow{p} S$ is called **injective** if $p(s_1) = p(s_2)$ implies $s_1 = s_2$ (it is one-to-one) and **surjective** if for all $t \in S$, there exists $s \in S$ such that $p(s) = t$. Then a **bijective** map is one that is both injective and surjective.

Lemma 14

A map p is bijective if and only if it has an inverse function π (in other words, $p\pi = i$, which happens if p is surjective, and $\pi p = i$, which happens if p is injective).

Lemma 15

If S is a finite set, then a map $p : S \to S$ is injective if and only if it is surjective. Both of these are equivalent to p being bijective.

Basically, all of these just mean we will "fill out" the whole set with p and not leave out any elements.

Definition 16

Let the **set of permutations** of S be $\text{Perm}(S)$, and define a law of composition to be the composition of the two permutations. Then the **symmetric group**, denoted S_n, is the set of permutations $\text{Perm}(1, \cdots, n)$.

The first interesting example happens when we take $n = 3$ (mostly because it's the first time we have a **nonabelian** group, meaning not all elements commute). There are $3! = 6$ total permutations in S_3.

Example 17

Consider the two permutations $p, q \in S_3$ such that p sends 1 to 2, 2 to 3, and 3 to 1, while q keeps 1 fixed but swaps 2 and 3.

Now we want to try composing these together. To do this, just look at where each element goes individually; pq corresponds to doing q, then p, so 1 goes to 1 and then 2. We find in the end that pq sends $(1, 2, 3)$ to $(2, 1, 3)$.

Example 18

Suppose $p \in S_5$ sends $(1, 2, 3, 4, 5)$ to $(3, 5, 4, 1, 2)$. Is there a more concise way to write this permutation?

One good question to ask here: **what does this do visually** if we draw arrows between our numbers $1, 2, 3, 4, 5$? 1 gets sent to 3, which gets sent to 4, which gets sent back to 1. This is a **3-cycle** (because it has three elements). Meanwhile, 2 and 5 are part of their own **2-cycle** or transposition. Thus, we can write this permutation in the **cycle notation** $(134)(25)$.

Fact 19

Any permutation can be written as disjoint cycles; just start from any number and see where it goes.

Example 20

Take $p \in S_5$ as above, and let q have cycle notation $(12)(34)(5)$.

We can then find (pq) by considering one number at a time and follow the cycles. The result is written below:

$$pq = [(134)(25)][(12)(34)(5)] = [(1523)(4)].$$

Essentially, 1 goes to 2 under q and then 5 under p, so 1 goes to 5. Then we try to see where 5 goes, and repeat until we've covered all of the numbers. (It's important to remember that we do the action of q before the action of p.)

There is one problem: (134) and (341) are actually the same cycle. Cycle notation isn't unique! (For now, that really doesn't matter, though.) For convenience and by convention, we will also avoid writing the fixed points from now on. (So if we see an index that doesn't appear, it just goes back to itself in our permutation.)

One last example:

Example 21

Let $r = (12345)$. What is rp? What is r^{-1}? What is rpr^{-1}?

All of these are fairly straightforward calculations or observations. First of all,

$$rp = [(12345)][(134)(25)] = [(142)(35)].$$

Similarly, the inverse of a cycle can be found by traversing it in reverse:

$$r^{-1} = [(12345)]^{-1} = [(15432)].$$

Finally, rpr^{-1} is a bit more important in our study:

$$rpr^{-1} = (rp)r^{-1} = [(142)(35)][(15432)] = [(13)(245)].$$

This last operation is called **conjugation**, and it is important (we will see later in the class that conjugate permutations have the same **cycle type**).

Now, we'll talk a bit about permutation matrices: these are ways to assign matrices to permutations. Specifically, P operates on a column vector (which contains the elements of S) by applying p to it.

Example 22

Let $p = (123) \in S_3$, and let's say our column vector is $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$. Then we define a matrix P associated to the permutation p such that $Px = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix}$.

Notice that this is permuting the **entries**, so x_1 ends up where x_2 is, and so on. In particular, this is the inverse of the actual permutation p.

An important idea from linear algebra is that of a **basis** – let's define one for our matrices. Let the **matrix units** e_{ij} be defined to have an entry 1 in row i, column j, and 0s everywhere else. Then if $A = (a_{ij})$, then we can just write

$A = \sum a_{ij} e_{ij}$ (entry by entry).

Our permutation matrix P can then be written as

$$P = \sum_j e_{p_j, j}$$

(again, notice this is the inverse of the permutation p, since p_j corresponds to j instead of j corresponding to p_j).

What if we want to compose these matrices? Well, notice that $e_{ij} e_{kl}$ is e_{il} if $j = k$ and zero otherwise. So given two permutation matrices P, Q,

$$PQ = \left(\sum_j e_{p_j, j} \right) \left(\sum_k e_{q_k, k} \right)$$

the terms become zero unless $j = q_k$, and we are left with

$$PQ = \sum_k e_{p q_k, q_k} e_{q_k, k} = \sum_k e_{p q_k, k}$$

which is what we want: multiplying matrices gives us our composition.

Let's now return back to the permutation group S_3: we'll try to generate our whole group with $p = (123)$ and $q = (23)$. Then $pq = (12)$, $p^2 = (132) = p^{-1}$, and $p^3 = 1$. Similarly, we can find that $q^2 = 1$ and $p^2 q = qp = (13)$: this is all the elements of S_3. So now we have a way to describe our group law:

Fact 23

S_3 is generated by two elements x, y, such that $x^3 = y^2 = 1$ and $yx = x^2 y$.

Now any element can be written in the form $x^a y^b$, where $0 \leq a \leq 2$ and $0 \leq b \leq 1$. We use the $yx = x^2 y$ part to turn any product of xs and ys to move all the xs to the front and all the ys to the back, and then reduce mod 3 and 2, respectively. This is exactly the set of 6 permutations that we want!

3 September 10, 2018

Recall that a group's law of composition must be associative, have an identity, and have an inverse.

Definition 24

A **subgroup** H of a group G is some subset of the group with the same law of composition. H must be **closed**; that is, if $a, b \in H$, then $ab \in H$. In addition, 1_G, the identity in G, must also be in H, and all inverses a^{-1} of $a \in H$ must also be in H.

Here are a few simple examples of subgroups:

Example 25

The set of positive reals is a subgroup of $\mathbb{R} \setminus \{0\}$ (the set of nonzero real numbers), with multiplication as the law of composition.

Example 26

The **special linear group** SL_n (the set of matrices in GL_n of determinant 1) is a subgroup of GL_n, because determinant is multiplicative.

There are lots and lots of subgroups of GL_n, and that's why it's important! We don't have enough theory to describe them all, though.

Instead, we'll try talking about subgroups of $\mathbb{Z}^+$, the integers under addition. For a subset of the integers to be a subgroup, it must be closed under addition, it must contain 0, the additive inverse, and if $a \in H$, then $-a \in H$.

Fact 27

$n\mathbb{Z} = \{\cdots, -2n, -n, 0, n, 2n, \cdots\}$, that is, the multiples of n, is a subgroup.

Proposition 28

All subgroups of $\mathbb{Z}^+$ are of the form $n\mathbb{Z}$ for some integer n.

Proof. Let H be a subgroup of $\mathbb{Z}^+$. We know $0 \in H$; if H is just that, we're done (and this is $0\mathbb{Z}$). Otherwise, let $a \neq 0$ be some element. Then $-a \in H$, and one of these must be positive, so H contains at least one positive integer. Let n be the smallest such positive integer in H.

We claim now that $H = n\mathbb{Z}$. Since $n \in H$, $2n \in H$, and so on, and same with the negative multiples of n. Thus $n\mathbb{Z} \subset H$, and now we just want to show $H \subset n\mathbb{Z}$. Take some $b \in H$. By the division algorithm, we can write $b = nq + r$ for $q \in \mathbb{Z}, 0 \leq r < n$. Since $b \in H, nq \in H$, $r = b - nq \in H$. But $0 \leq r < n$, and n was defined to be the smallest positive element, so $r = 0$. Thus b must have been a multiple of n, so $H \in n\mathbb{Z}$, concluding the proof. □

Corollary 29 (Bezout's Lemma)

If d is the greatest common divisor of a, b, then it can be written as $d = ra + sb$.

Proof. Let $a, b \in \mathbb{N}$. Then $a\mathbb{Z} + b\mathbb{Z}$ is a subgroup of $\mathbb{Z}$ under addition (it is easy to check closure, identity, and inverses). Thus it is of the form $d\mathbb{Z}$ for some $d \in \mathbb{N}$. This means $a\mathbb{Z} + b\mathbb{Z} \in d\mathbb{Z}$, so $a, b \in d\mathbb{Z}$, so d divides both a and b. But we also know that $d\mathbb{Z} \in a\mathbb{Z} + b\mathbb{Z}$, so we can write $d = ra + sb$ for some $r, s \in \mathbb{Z}$. □

On the other hand, let $a, b \in \mathbb{N}$. Then $a\mathbb{Z} \cap b\mathbb{Z}$ is the set of integers divisible by both a and b. **The intersection of two subgroups is also a subgroup**, so this set is $m\mathbb{Z}$ for some m. So if $a|x, b|x$, then $m|x$, and if $m|y$, then $a|y, b|y$. Taking $y = m$, $a|m$ and $b|m$, and we want m to be the **least common multiple** of a and b.

Theorem 30

Let a, b be positive integers. If d is their greatest common divisor and m is their least common multiple, then $ab = dm$.

Definition 31

A map $\phi : G \to G'$ is an **isomorphism** if ϕ is a bijective map, and for all $a, b \in G$, if $\phi(a) = a'$ and $\phi(b) = b'$, then $\phi(ab) = a'b'$.

(One example is the identity map from G to itself.)

Example 32

S_n is isomorphic to the set of permutation matrices. Also, the real numbers under addition map to the positive reals under multiplication, using the exponential map $x \to e^x$.

Theorem 33

Every group G of order $|G| = n$ is isomorphic to a subgroup of S_n.

Proof. Let $g \in G$. Define a map $m_g : G \to G$ to be multiplication by g; that is, it sends $m_g : x \to gx$. Then m_g is a permutation of the elements in g (it is a bijective map, since inverses exist). Thus $M = \{m_g | g \in G\}$ is a subgroup of all permutations of G, because $m_g m_h(x) = ghx = m_{gh}(x)$. The identity permutation exists in M (it corresponds to the identity element in G), and $m_{g^{-1}} = m_g^{-1}$, so inverses exist too. Since $G \to M$ is a bijective map, it is an isomorphism, and thus G is isomorphic to a subgroup of S_n. □

Example 34

Take $S_3 = \{1, x, x^2, y, xy, xy^2\}$, where $x = (123), y = (23)$.

We can embed this into S_6 as follows: assign the indices 1, 2, 3, 4, 5, 6 to the elements of S_3 above. Then x is the permutation (123)(456), since it sends 1 (with index 1) to x (with index 2), and so on. y is the permutation (14)(26)(35), and now just compose those to get all other permutations. Thus, S_3 is isomorphic to a subgroup of S_6.

4 September 12, 2018

Recall that an isomorphism is a bijective map compatible with group laws. Specifically, it sends $G \to G'$ in such a way that $a \to a'$ and $b \to b'$ means $ab \to a'b'$. There usually do exist isomorphisms from a group to itself that are not the identity though. (For example, negation in $\mathbb{Z}$ works.) We'll generalize this concept a bit:

Definition 35

A **homomorphism** of groups $G \to G'$ that is compatible with group laws, but doesn't necessarily have to be bijective. If we call our homomorphism ϕ, we can write this as

$$\phi(ab) = \phi(a)\phi(b) \forall a, b \in G$$

In other words, homomorphisms are similar to isomorphisms, but two elements of G can be sent to the same element of G'.

Example 36

The determinant map $\phi : GL_n \overset{\det}{\to} \mathbb{R}^\times$ is a homomorphism.

In particular, det sends a matrix A to $\det A$, and indeed $AB \to \det(AB) = \det A \det B$, which means the map ϕ is compatible with group laws.

Lemma 37 (Cancellation law)

Let a, b, c be elements of a group. If $ab = ac$, then $b = c$.

Proof. Multiply by a^{-1} on the left side. □

Lemma 38

Let $\phi : G \to G'$ be a homomorphism. Then $\phi(1_G) = 1_{G'}$, and $\phi(a^{-1}) = \phi(a)^{-1}$.

Proof. For the first part, plug in $a = b = 1$ into the definition of a homomorphism. Since $1 = 1 \cdot 1, \phi(1) = \phi(1)\phi(1)$, which means $\phi(1) = 1_{G'}$ by the cancellation law. For the second part, plug in $b = a^{-1}$. Then $\phi(1) = \phi(a)\phi(a^{-1})$, and the left hand side is the identity (by the previous part). □

Example 39

$\mathbb{C}^+ \to \mathbb{C}^\times$ is a homomorphism under the homomorphism $x \to e^{2\pi i x}$. (This factor of $2\pi i$ in the numerator is nice because 1 gets sent to 1.)

This is not bijective; in particular, if two complex numbers x and y differ by an integer, they are sent to the same element of $\mathbb{C}^\times$.

Example 40

Every permutation has a **sign**, so we can send $S_n \to \{\pm 1\}$, where we send to 1 if the sign is positive (an **even** permutation) and -1 if the sign is negative (an **odd** permutation).

How do we define sign? $p \in S_n$ corresponds to a permutation matrix. Then we just define the sign to be the determinant of P. Since permutations essentially swap rows of the identity matrix, and each swap multiplies the determinant by -1, the sign of a matrix is $(-1)^n$, where n is the number of transpositions!

Example 41

Let G be any group, and pick some $a \in G$. Then there exists a map $\mathbb{Z}^+ \to G$ that sends $k \to a^k$.

Of course, $a^{k+l} = a^k a^l$, so this is a homomorphism.

Definition 42

Let $\phi : G \to G'$ be a homomorphism. Then the **image** of ϕ, denoted $\operatorname{Im}\phi$, is everything in G' that is hit; that is, it is the set

$$\operatorname{Im}\phi = \{a' \in G' : \exists x \in G \text{ s.t. } \phi(x) = a'\}.$$

Proposition 43

If ϕ is a homomorphism from G to G', then $\operatorname{Im}\phi$ is a subgroup of G'.

Proof. If $a', b' \in \operatorname{Im}\phi$, then there exist x, y such that $\phi(x) = a', \phi(y) = b'$. Then $\phi(xy) = \phi(x)\phi(y) = a'b'$, so $a'b'$ is also in the image of ϕ. Thus, $\operatorname{Im}\phi$ has closure. The identity is clearly in $\operatorname{Im}\phi$, since $\phi(1_G) = 1_{G'}$. Finally, if $a' \in \operatorname{Im}(\phi)$, $\phi(x) = a' \implies \phi(x^{-1}) = a'^{-1}$. Thus all group axioms are indeed satisfied. □

Definition 44

Let $\phi : G \to G'$ be a homomorphism. Then the **kernel** of ϕ, denoted $\ker\phi$, is the elements in G that go to the identity; that is,

$$\ker\phi = \{x \in G : \phi(x) = 1\}.$$

Proposition 45

If ϕ is a homomorphism from G to G', then $\ker\phi$ is a subgroup of G.

Proof. If $x, y \in \ker\phi$, then $\phi(xy) = \phi(x)\phi(y) = 1_{G'}1_{G'} = 1_{G'}$. Like before, $\phi(1_G) = 1_{G'}$, and if $\phi(a) = 1_{G'}$, then $\phi(a^{-1}) = 1_{G'}^{-1} = 1_{G'}$. Thus the kernel $\ker\phi$ satisfies all group axioms. □

Proposition 46 (Extra property of the kernel)

If $x \in \ker\phi$ and we have an arbitrary element $g \in G$, then gxg^{-1} (called the **conjugate** of x by g) $\in \ker\phi$. (Conjugate elements of x are in the kernel.)

Proof. This is just a direct computation

$$\phi(gxg^{-1}) = \phi(g)\phi(x)\phi(g^{-1}) = \phi(g)\phi(g^{-1}) = 1_{G'},$$

so $gxg^{-1} \in \ker\phi$. □

Definition 47

A subgroup H of G is called **normal** if for all $x \in H, g \in G$, we have that $gxg^{-1} \in H$ as well.

For example, we just showed that the kernel is a normal subgroup.

Example 48

We continue some of the above examples:

- In the determinant map $GL_n \xrightarrow{\det} \mathbb{R}^\times$, the kernel is SL_n.
- In the exponential map $\mathbb{C}^+ \to \mathbb{C}^\times$, the kernel is $\mathbb{Z}$.
- The kernel of the sign map for S_n is called the **alternating group** A_n.
- The map that sends $k \to a^k$ for an arbitrary $a \in G$ has kernel 0 or $n\mathbb{Z}$, depending on the order of a in G.

The last map in the example above has image $\{\cdots, a^{-1}, 1, a, a^2, \cdots\}$, where this set may or may not be infinite depending on whether the kernel is 0 or $n\mathbb{Z}$. This is denoted $\langle a\rangle$, and it is called the **subgroup generated by** a. If a has order n, that means the group is $\{1, a, \cdots, a^{n-1}\}$, and $a^n = 1$.

But if we have more than one element, things are harder. Sometimes we don't even know how to write down the group generated by two elements a, b with some relations.

Example 49

We can construct a homomorphism from $S_4 \to S_3$. There are three ways to partition $\{1, 2, 3, 4\}$ into sets of 2 (it depends what 1 goes with). Call π_1 the partition that puts $\{1, 2\}$ together, π_2 the one that puts $\{1, 3\}$ together, and π_3 the one that puts $\{1, 4\}$ together. Then any permutation of the elements of S_4 permutes π_1, π_2, π_3, which corresponds to an element of S_3!

For example, let $p = (123)(4) \in S_4$. Then $\pi_1 \to \pi_3$, $\pi_2 \to \pi_1$, $\pi_3 \to \pi_2$. We can write this as $\tilde{p} = (\pi_1\pi_3\pi_2)$. On the other hand, $p = (1234)$ sends $\pi_1 \to \pi_3, \pi_2 \to \pi_2, \pi_3 \to \pi_1$, so $\tilde{p} = (\pi_1\pi_3)(\pi_2)$.

We can have fun checking that $pq \to \tilde{p}\tilde{q}$, so this is indeed a homomorphism! It is also a surjective one, since p and q generate the whole group S_3. One exercise for us: what is the kernel of this map?

5 September 14, 2018

Definition 50

Let H be a subgroup of G, and let $a \in G$. Then a **left coset** $C = aH$ is the subset of G defined via

$$C = \{ah : h \in H\}.$$

There are many different cosets for a single group H.

Example 51

Let G be the symmetric group $S_3 = \{1, x, x^2, y, xy, x^2y\}$. Let $H = \{1, xy\}$ (notice that $(xy)^2 = xyxy = x^3y^2 = 1$).

Then we have

$$\begin{cases} 1H = xyH = \{1, xy\} \\ xH = x^2yH = \{x, x^2y\} \\ x^2H = yH = \{x^2, y\} \end{cases}$$

In particular, there are six ways to get a coset, but there are only three cosets.

Fact 52

All cosets have the same order as H.

This is because multiplying by some group element a is reversible; we can get backwards from aH to H by multiplying every element by a^{-1}. This is a bijective correspondence, so H and aH have the same order.

Proposition 53

The cosets partition the group and are disjoint (unless they refer to the same coset).

First, we prove a lemma:

Lemma 54

Let C be a coset. If $b \in C$, then $C = bH$ (in particular, all elements of G are in a coset of H).

Proof. We know that $C = aH$ for some a, so $b = ah$ for some $h \in H$. Thus $bH = ahH = a(hH)$, and hH is just H. Thus we have $aH = C$.

Except we used some symbolic notation above, so let's write this out more rigorously. We wish to show that if $b = ah$ for some $h \in H$, then $bH = aH$. Take an element of bH; it looks like bh' for some $h \in H$. So $bh' = ahh'$, and

since $hh' \in H$ (since H is closed), $bh' \in aH$. Thus $bH \subset aH$. On the other hand, we also have $bh^{-1} = a$, and we can do the same reasoning to show that $aH \subset bH$; thus, $aH = bH$, and thus $C = bH$. The central idea here is just that we can tack on an h or h^{-1} to convert between bH and aH. □

To finish proving the above proposition, we wish to show the cosets partition G.

Proof. (First of all, all cosets are nonempty, since $1 \in H$, so $a \in aH$. In addition, $a \in aH$ means every element is in a coset, so the cosets cover G.)

Let C_1, C_2 be cosets that have nonempty intersection; we wish to show that $C_1 = C_2$. We can write $C_1 = a_1H, C_2 = a_2H$ for some a_1, a_2 in our group. Then if $b \in C_1 \cap C_2$, then $b \in a_1H \implies bH = a_1H$ by the lemma. In addition, $b \in a_2H \implies bH = a_2H$. Thus $a_1H = a_2H$, so two different cosets cannot have nonempty intersection. □

Definition 55

Let H be a subgroup of G. Then the **index** of H in G, denoted $[G : H]$, is the number of (disjoint) cosets for H.

We know that the size of each coset is the size of H. This yields the following result:

Fact 56 (Counting)

For any subgroup H of a group G, we have $|G| = [G : H]|H|$.

Corollary 57 (Lagrange's theorem)

$|H|$ divides $|G|$ for any subgroup H.

This has a nice corollary if $|G| = p$, where p is prime:

Corollary 58

For groups of prime order, the only subgroups are the whole group and the trivial subgroup (containing only the identity). In addition, the order of any non-identity element of G is p, since the subgroup generated by that element has more than 1 element.

This means that every group of order p is a cyclic group! Just take some nonzero element a, and the group can be written as $\langle 1, a, \cdots, a^{p-1} \rangle$.

Corollary 59

The order of any element a of a group divides $|G|$.

Proof. The element a generates a subgroup whose order is equal to the order of that element, and then we can use Lagrange's theorem to get the desired result. □

Next, let's try to look at groups with non-prime order.

Example 60

What are all the groups of order 4?

Elements must have order 1, 2, or 4. The only element with order 1 is the identity. If G contains an element g of order 4, then G is the cyclic group of order 4 $\langle 1, a, a^2, a^3\rangle$. This can be denoted $\mathbb{Z}/4\mathbb{Z}$, as the integers mod 4 with addition. Otherwise, all other elements have order 2. So $G = \{1, x, y, z \mid x^2 = y^2 = z^2 = 1\}$.

Now either $xy = 1, x, y$,or z. But by cancellation, it can't be x or y. If $xy = 1$, then $xy = x^2$, so $x = y$, which is bad too (we assumed x, y, z are distinct). So $xy = z$, and similarly, $xz = y, yz = x$.

This gives the **Klein four group**, which is actually just $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$; that is, all ordered pairs of integers mod 2, with group operation addition. **Thus, there are two groups of order 4**.

Now, let's apply this idea to homomorphisms $\phi : G \to G'$. We know that the kernel $K = \{x \in G \mid \phi(x) = 1'\}$ is all sent to $1'$.

Fact 61

Let a be an element of G, and let ϕ be a homomorphism from G to G' with kernel K. Then the coset aK is sent to $\phi(a)$.

(This can be verified by writing out $\phi(ak) = \phi(a)$ for any $k \in K$.) In fact, this goes both ways:

Lemma 62

Let $a, b \in G$. Then $\phi(b) = \phi(a)$ if and only if $b \in aK$, where K is the kernel of ϕ.

Proof. $b \in aK$ means that $b = ak$ for some $k \in K$. Then $\phi(b) = \phi(ak) = \phi(a)\phi(k) = \phi(a)$. On the other hand, $\phi(b) = \phi(a) \implies \phi(ba^{-1}) = \phi(b)\phi(a^{-1}) = \phi(a)\phi(a^{-1}) = \phi(aa^{-1}) = \phi(1) = 1'$. Thus ba^{-1} is in the kernel, so $b \in aK$. □

Corollary 63

$|G| = |\operatorname{Im}\phi||\ker\phi|$ for any homomorphism $\phi : G \to G'$.

For example, in the example above $\phi : S_4 \to S_3$, the image was the entire set S_3, but there were 4 elements in the kernel (it is the identity, (12)(34), (13)(24), and (14)(23)). $24 = 6 \cdot 4$, and mathematics works.

6 September 17, 2018

Recall that last time, we had the "counting formula" $|G| = |\operatorname{Im}\phi||\ker\phi|$. We'll see some quick applications of this here:

Example 64

The kernel of the sign permutation that sends $S_n \to \{\pm 1\}$ is the alternating group A_n. It has order $\frac{n!}{2}$ for $n \geq 2$, since the image has order 2.

Theorem 65 (Correspondence Theorem)

Let $G \to G'$ be a (surjective) homomorphism. (Surjectivity means we don't have to mention the image of G' all the time.) Then there is a **bijective correspondence** between subgroups of G containing $\ker \phi$ and subgroups of G'.

Specifically, if H is a subgroup of G, then H corresponds to $H' = \phi H$ (symbolic notation; it's everything we get when we apply ϕ to elements in H). On the other hand, if H' is a subgroup of G, it corresponds to $H = \phi^{-1}H'$ (again symbolic notation; the set of all x such that $\phi(x) \in H'$). In addition, we have $|H| = |H'||\ker \phi|$.

Sketch. First, let H be a subgroup of G containing K. We can show $\phi H = H'$ is a subgroup of G' by verifying the group axioms; this is not too hard.

Now, let H' be a subgroup of G'. We show that $\phi^{-1}H'$ is a subgroup of G containing the kernel. We know that the inverse image of the identity, which is contained in H', is the kernel, so the kernel is contained in $\phi^{-1}H'$. The identity is in the kernel, so we get that for free. Closure is not too hard either: if $x, y \in \phi^{-1}H'$, then $\phi(x), \phi(y) \in H'$, so $\phi(x)\phi(y) = \phi(xy) \in Y' \implies xy \in \phi^{-1}H'$. Inverses can be checked pretty easily too.

Finally, we show the correspondence: $\phi^{-1}\phi H = H$ if H contains the kernel $\ker \phi$. We know that $\phi^{-1}\phi H \supset H$ for every map, so we just need to show that $\phi^{-1}\phi H \subset H$. Let $x \in \phi^{-1}\phi H$; this means that $\phi(x) \in \phi H$. Therefore, $\phi(x) = \phi(y)$ for some $y \in H$, so $\phi(xy^{-1}) = 1$, so $xy^{-1} \in \ker \phi$, so $x \in y \ker \phi \in \phi H$, since H contains the kernel.

Lastly, $\phi\phi^{-1}H' = H'$ (which is true for any surjective map). □

This is pretty interesting when we apply it to a few examples.

Example 66

Let $G = G' = \mathbb{C}^\times$. Consider the map ϕ sending $z \to z^2$. (In general, this map is a homomorphism if and only if the group is abelian, since $(xy)^2 = xyxy \neq x^2y^2$ unless $xy = yx$.) Then the kernel is the set of all $z \in \mathbb{C}^\times$ such that $\phi^2 = 1$; thus, $\ker \phi = \{-1, 1\}$.

Now we can pick a subgroup and apply the correspondence theorem. The subgroup $H_1 = \mathbb{R}^\times \in G$ corresponds to the image H_1' of $\mathbb{R}^\times$ under ϕ, which is the group of positive real numbers under multiplication. On the other hand, $H_2' = \{\pm 1, \pm i\} \in G'$, the set of fourth roots of unity, is a subgroup of G', so it corresponds to the subgroup H_2 of complex numbers with square equal to the fourth roots of unity: that is, the eighth roots of unity. Notice that we indeed have $|H_2| = 8 = |H_2'||\ker \phi| = 4 \cdot 2$.

Example 67

Let $G = S_4$, $G' = S_3$, and let ϕ be the map defined on September 12. Let's write out the correspondence explicitly.

(Recall that π_1, π_2, π_3 are the three partitions of $\{1, 2, 3, 4\}$ into three pairs, and G' tracks the permutation of $\{\pi_1, \pi_2, \pi_3\}$ when G tracks the permutation of $\{1, 2, 3, 4\}$. In particular, the kernel of ϕ is $\{1, (12)(34), (13)(24), (14)(23)\}$.)

What are the subgroups of S_3? They have to have order 1 (trivial), 6 (the whole group), or 2 or 3. The latter two cases are prime order, so they must be cyclic; we can just look at each element and find its order. Write

$$S_3 = \{1, x, x^2, y, xy, x^2y\}; x = (\pi_1\pi_2\pi_3), y = (\pi_2\pi_3)$$

Then the four other subgroups are generated by y, xy and x^2y, each with order 2, and x with order 3. Thus, by the correspondence theorem, there are exactly six subgroups of S_4 that contain the kernel $\ker \phi$. They will have order

4, 8, 8, 8, 12, and 24. The smallest of these is $\ker\phi$, the largest is S_4, and the one with order 12 is likely A_4, as long as $\ker\phi$ is contained inside A_4 (the answer turns out to be yes, since the orders of the nontrivial elements of the kernel are 2).

How would we find the subgroups of order 8? We know that the permutation $y \in S_3$ has order 2. We've defined y to switch π_2 and π_3, so we need to find some permutation not in the kernel that fixes π_1 or switches π_2 and π_3. It turns out (12) swaps π_2 and π_3, so $(12) \to y$. Thus $(12) \in H_1$, and now we can get the rest of H_1 by taking the kernel and multiplying them by (12). In other words, we can write this as

$$H_1 = \langle \ker\phi, (12)\rangle = \{\ker\phi, (12), (34), (1324), (4231)\}.$$

Similarly, we can also find H_2, H_3.

Unfortunately, this correspondence theorem does not tell us about other subgroups that don't contain the kernel. For example, there are many subgroups of S_4 with order 2 or 3 (transpositions and 3-cycles). But we've still managed to gain quite a bit of structure!

7 September 19, 2018

Let's quickly review some important ideas: a **left coset** of a subgroup H of a group G is a subset C consisting of all elements of the form $ah, a \in G, h \in H$. These cosets are important because they are the same size and partition the group: this gives us a counting formula

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

where $[G : H]$ is called the **index** of H and is the number of cosets of H.

Definition 68

Let $\phi : G \to G'$ be a homomorphism with kernel K. The **fibre** of an element in G' is the inverse image of that element: specifically, the fibre over $z \in G'$ is

$$\phi^{-1}(z) = \{g \in G : \phi(g) = z\}.$$

Proposition 69

The nonempty fibres of G' are the left cosets of K.

It's actually not so important that we distinguish between left and right cosets here:

Fact 70

The left and right cosets of the kernel $\ker\phi$ for a group homomorphism are the same.

Proof. If $x \in K$, then $gxg^{-1} \in K$, because$\phi(gxg^{-1}) = \phi(g)\phi(x)\phi(g^{-1}) = \phi(g)\phi(g^{-1}) = \phi(1) = 1$. Thus, $aHa^{-1} = H$, meaning that $aH = Ha$ (we can write this out without symbolic notation if we want; see the proposition below). □

It is important to note that **left and right cosets are not always the same**, though!

Example 71

Take $S_3 = \{1, x, x^2, y, xy, x^2y\}$ and let $H = \{1, y\}$. Then the left cosets are

$$1H = yH = \{1, y\}, xH = xyH = \{x, xy\}, x^2H = x^2y\{x^2, x^2y\},$$

but this is not the set of right cosets:

$$H1 = Hy = \{1, y\}, Hx = Hx^2y = \{x, x^2y\}, Hx^2 = Hxy = \{x^2, xy\}$$

Many of the statements we can make about left cosets are also true for right cosets, but it's just important to notice that they don't coincide exactly! If left and right cosets are the same, though, we are pretty happy.

Definition 72

A subgroup H of G is called a **normal subgroup** if for all $x \in H, g \in G$, $gxg^{-1} \in H$.

An example is $\{1, x, x^2\}$; we know this is normal because it is the kernel of the sign map.

Proposition 73

The following are equivalent:

- H is a normal subgroup.
- Every left coset is a right coset. (Specifically, if a left coset is a right coset, and $C = aH$, then it is also Ha.)

Proof. We'll do this with symbolic notation. H is normal if and only if $gHg^{-1} \subset H$. This means $g^{-1}Hg \subset H \implies H \subset gHg^{-1}$. Thus H is normal if and only if $H = gHg^{-1}$. Now multiply both sides on the right by g, and we have $Hg = gH$.

For completeness, we'll also write this out more concretely. Suppose H is normal. Then for all $h \in H, g \in G$, we have that $ghg^{-1} \in H$ and we wish to show that $gH = Hg$. We know that for every element $h \in H$, $gh \in gH$. We want to then show that there exists $h' in H$ such that $gh = h'g$. We use a trick: write

$$yx = x(x^{-1yx}),$$

so that the order of x and y flip, but y changes to a conjugate. In other words, we can move elements past each other, but then we have to conjguate y. Similarly, we can also move y over to the other side of x:

$$yx = (yxy^{-1})y$$

and this time we conjugate x. So here, $gh = (ghg^{-1})g$, but since H is a normal subgroup, $ghg^{-1} \in H$. So $gh \in Hg$, and we're done! The other direction is basically the same. Finally, if $aH = Hb$, then $a \in Hb$, so $Ha = Hb$ and we actually have $aH = Ha$. □

Next question: we know that kernels of homomorphisms are normal. Are all normal subgroups kernels of a map? The answer is yes! We can make the cosets of a normal subgroup into a group (and remember now that the left and right cosets for a normal subgroup are the same):

Definition 74

Let N be a normal subgroup of G. The **quotient group** $\overline{G}$ or G/N is the set of cosets of N. This is a group with law of composition equal to the product set:

$$C_1C_2 = (aN)(bN) = aNbN = a(bN)N = (ab)N.$$

Formally, the definition of a product set of $S, T \subset G$ is

$$ST = \{g \in G \mid g = st, s \in S, t \in T\}$$

It is extremely important that the subgroup has to be normal:

Example 75

Take $C_1 = xH = \{x, xy\}$, $C_2 = x^2H = \{x^2, x^2y\}$ in S_3. Then C_1C_2 is not a coset: it is $\{x^3, x^3y, xyx^2, xyx^2y\} = \{1, y, x^2, x^2y\}$, which has 4 elements, and $4 \nmid 6$. This is because $H = \{1, y\}$ was not normal.

The identity element of the quotient group G/N is the coset N, and the inverse of aN is $a^{-1}N$. So ultimately, we have a homomorphism $G \xrightarrow{\pi} \overline{G}$ which sends $a \to aN$. The homomorphism of π then has kernel N, so N is always the kernel of some map!

Example 76

We finish with two quick examples of quotient groups:

- $\mathbb{Z}/n\mathbb{Z}$ is the set of (congruence classes of) integers mod n.
- Take $G = SL_2$, $N = \{I, -I\}$ (which is the **center** of SL_2). Then $AN = \{A, -A\}$ for any matrix A, and then $\overline{G} = G/N$ is a set of pairs of matrices (differing by sign). This construction comes up in other areas of mathematics, too.

8 September 24, 2018

We'll switch gears for a few lectures: the main difficulty of linear algebra is the notation.

Definition 77

A field F is a set with $+, -, \times, /$ operations, where 0 (the additive identity) is not included in the $\times$ and $/$ operations. Examples of fields include $\mathbb{R}, \mathbb{C}$, and $\mathbb{F}_p$ (the integers mod p).

Extending this definition, F^n is the space of column vectors

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, x_i \in F.$$

However, it is important to talk about abstract vector spaces, and we will do so:

Definition 78

An ***F*-vector space** V is a set with two operations: addition of two vectors $V \times V \xrightarrow{+} V$ and scalar multiplication $F \times V \xrightarrow{\cdot} V$ with the following axioms:

- $(V, +)$ is an abelian group with identity 0.
- Associativity: $a(bv) = abv$ for all $a, b \in F$ and $v \in V$.
- Distributivity: if $a, b \in F$ and $v, w \in V$, then $(a + b)v = av + bv$ and $a(v + w) = av + aw$.

Using these laws, any computation will result in a **linear combination** of the form

$$a_1 v_1 + a_2 v_2 + \cdots + a_n v_n$$

where $a_i \in F, v_i \in V$. But the important idea is to try working with a finite **ordered** set of vectors, which we will denote $(v_1, v_2, \cdots, v_n)$. (Using braces $\{\}$ from here on out means that we only care about the set elements and not their order.)

Then any linear combination can be written in the matrix multiplication form

$$\begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = v_1 a_1 + v_2 a_2 + \cdots + v_n a_n$$

The scalar is on the wrong side, so let's just define $va = av$ for a scalar a. This implies that F is abelian under multiplication, since $a(bv) = a(vb) = vba \implies ab = ba$.

Fix $S = (v_1, \cdots, v_n)$. Then we can "multiply" S by a column vector X to get $v_1 x_1 + v_2 x_2 + \cdots + v_n x_n$. So using S as a **basis** of V, we now have a map

$$\psi : F^n \xrightarrow{\rightarrow} V$$

that sends $X \to SX$ in a way such that $\psi(X + Y) = \psi(X) + \psi(Y)$ and $\psi(cX) = c\psi(X)$. This means that ψ is a **linear transformation**, and we can therefore talk about the kernel and image of ψ:

Definition 79

These properties define ψ but are really talking about the properties of the basis:

- ψ is **injective** if the kernel is just the zero vector. In other words, $\psi(X) = 0 \implies X = \vec{0}$, or alternatively $X \neq 0 \implies SX \neq 0_V$. Then $(v_1, v_2, \cdots, v_n)$ is called **linearly independent.**
- ψ is **surjective** if the image is all of V. In other words, for all $v \in V$, there exists $X \in F^n$ such that $V = SX$. In other words, every vector can be written as a linear combination of S, so S **spans** V.
- ψ is **bijective** if it is both injective and surjective; in particular, then $(v_1, v_2, \cdots, v_n)$ forms a **basis** for V.

Remark 80. *The idea is that V is not exactly F^n; the two are just isomorphic under the map ψ. We don't need to write a vector space explicitly in terms of basis vectors.*

$(v_1, v_2, \cdots, v_n)$ forming a basis means that for all $v \in V$, there exists **exactly one** $X \in F^n$ such that $v = v_1 x_1 + v_2 x_2 + \cdots + v_n x_n$. Then we can call X the **coordinate vector** with respect to the basis $(v_1, v_2, \cdots, v_n)$.

This means that we can always work with F^n instead of V if we have a basis, but it's generally hard to find the formula $X = \psi^{-1}V$. Also, the basis is not unique; in fact, there are many bases for a vector space V, and there is often no natural basis.

Example 81

Consider an $m \times n$ matrix with entries in F. Let V be the space of solutions to $Ax = 0$, where $x \in F^n$. Then there's a lot of ways to write the nullspace (or kernel) with a basis, and there are many ways to pick one (for example, by row reduction).

Example 82

Let V be the space of functions of t such that $\frac{d^2f}{dt^2} + f = 0$. Then a natural basis is $B = (\cos t, \sin t)$, but there are other natural bases too, such as $B' = (e^{it}, e^{-it})$.

Let's see how we can write vectors as combinations of other vectors with a matrix. Let $S = (v_1, v_2, \cdots, v_m)$ be a basis of V, and let A be an $m \times n$ matrix with entries in F. Then $SA = (SA_1, SA_2, \cdots, SA_n)$, where A_i are the column vectors of A.

Now let $(w_1, \cdots, w_n)$ be any ordered set of vectors in V. Then any w_j can be written as a linear combination of the basis vectors, which we can write as

$$w_j = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \begin{bmatrix} a_{1,j} \\ \vdots \\ a_{m,j} \end{bmatrix}$$

for some $a_{i,j} \in F$. Thus there is a unique matrix A such that

$$\begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} A = \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix}.$$

Theorem 83

If B and B' are two bases for V, then $m = n$. Call this the **dimension** of F.

Proof. We show that if $(v_1, \cdots, v_m)$ spans V and $(w_1, \cdots, w_n)$ is a set of vectors in V, $m < n$ implies that $(w_1, \cdots, w_n)$ is dependent. This would imply that for two bases, $m \geq n$, and similarly $n \geq m$, so $m = n$. From above, we know that we can find a matrix A such that

$$\begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} A = \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix}$$

which we can write as $PA = Q$. Now since $m < n$, the system of equations $AX = 0$ has nonzero solutions for $X \in F^n$. Multiply both sides of the equation by X (the solution); then $PAX = 0 = QX$, so the column vectors of Q are not independent, which is what we wanted to show. □

One final important idea is that of a **basechange** matrix. Basically, we can always write a new basis of our vector space in terms of the old basis vectors as $B' = BP$, where B' and B are vectors and P is a matrix. Then P is a basechange matrix that is also invertible, since we could have reversed the process and written our old basis in terms of the new basis vectors.

Well, let v be any vector in our vector space. Then we can express it in both bases B and B' as $v = BX = B'X'$, so $BX = B'X' = BPX' \implies PX' = X$. Thus, if $B' = BP$ defines the basechange matrix, then $PX' = X$, and the **coordinate vectors transform in the opposite direction as the bases**.

9 September 26, 2018

Definition 84

Let V, W be two vector spaces over a field F. A **linear transformation** is a map $V \xrightarrow{T} W$ such that

- $T(v_1 + v_2) = T(v_1) + T(v_2)$ for any $v_1, v_2 \in V$; that is, T is an additive homomorphism.
- $T(cv) = cT(v)$ for all $c \in F$ and $v \in V$.

We can immediately come up with some (simple but useful) linear transformations:

Example 85

Let A be an $m \times n$ matrix with entries in F. Then the map $F^n \to F^m$ taking a vector $X \to AX$ is a linear transformation.

Example 86

Let $S = (v_1, v_2, \cdots, v_m)$ be a vector in V. The map $\psi : F^m \to V$ sends $X \to v_1x_1 + \cdots + v_mx_m$.

Example 87

Let V be a function of t, and let $a \in \mathbb{R}$. Then the map $V \to \mathbb{R}$ that sends $f \to f(a)$ is a linear transformation.

We can represent any linear transformation T as a matrix **with respect to a basis**. Let $B = (v_1, \cdots, v_m)$ be a basis for V and $C = (w_1, \cdots, w_n)$ be a basis for W. Apply T to each v_j. Then $T(v_j)$ is some vector in W, so it can be written as a linear combination of the basis vectors of W: $T(v_j) = CA_j$, where A_j is a column vector. Then

$$A = \begin{bmatrix} | & | & \vdots & | \\ A_1 & A_2 & \cdots & A_n \\ | & | & \vdots & | \end{bmatrix}$$

describes the linear transformation, and we can say that $T(B) = CA$ as symbolic notation.

Well, we have a bijective map from F^m to V and a bijective map from F^n to W. So essentially, the matrix from F^m to F^n, denoted by A, is basically the same as the linear transformation $T : V \to W$. We'll draw a diagram to explain why $T(B) = CA$. Here's the general form:

$$\begin{array}{ccc} F^m & \xrightarrow{A} & F^n \\ \downarrow{\scriptstyle B} & & \downarrow{\scriptstyle C} \\ V & \xrightarrow{T} & W \end{array}$$

and here's the form with a specific vector X:

$$\begin{array}{ccc} X & \xrightarrow{A} & AX \\ \downarrow{\scriptstyle B} & & \downarrow{\scriptstyle C} \\ BX & \xrightarrow{T} & CAX \end{array}$$

Definition 88

The **kernel** or nullspace of a linear transformation T is the set of vectors $v \in V$ such that $T(v) = 0$. The **image** of T is a subspace of W, defined to be the set of all $w \in W$ such that there exists some v with $T(v) = w$. Basically, we're using the same definitions as before.

There are alternative ways to describe these spaces: in the language of F^m and F^n and describing our linear transformation in matrix form, the kernel is all $X \in F^m$ such that $AX = 0$. In other words, this is the set of solutions to a homogeneous linear equation. Meanwhile, the image is all $Y \in F^n$ such that we can solve $AX = Y$; this is also the span of the column vectors of A.

Theorem 89

Let $\phi : V \xrightarrow{T} W$ be a linear transformation. Then

$$\dim V = \dim \ker \phi + \dim \operatorname{Im} \phi.$$

Proof. Let $(u_1, \cdots, u_k)$ be a basis for the kernel, and let $(w_1, \cdots, w_r)$ be a basis for the image. Let $v_j \in V$ be the vector such that $T(v_j) = w_j$ (this exists because w_js are in the image). We will show that $(u_1, \cdots, u_k, v_1, \cdots, v_r)$ form a basis for V, which implies the result.

First of all, we show that this is a linearly independent set. Say that we have

$$a_1 u_1 + \cdots + a_k u_k + b_1 v_1 + \cdots + b_r v_r = 0.$$

Apply T to this set; then the first k go to zero, and we are left with

$$b_1 w_1 + \cdots + b_r w_r = 0$$

and since w_j formed a basis for the image, all b_i are equal to 0. Now going back to the original set, a_i must all be zero since u_j form a basis for the kernel.

Now we show this spans V. Pick some arbitrary vector $v \in V$, and let $w = T(v)$. We can write w as a linear combination of the vectors in $(w_1, \cdots, w_r)$ because it is in the image:

$$T(v) = w = b_1 w_1 + \cdots + b_r w_r.$$

Now take the corresponding vectors in V: let $v' = b_1 v_1 + b_2 v_2 + \cdots + b_r w_r$, so that $T(v') = T(v)$. Thus $T(v - v') = 0$, so $v - v'$ is in the kernel, so we can write

$$v - v' = a_1 u_1 + \cdots + a_k u_k \implies v = a_1 u_1 + \cdots + a_k u_k + b_1 w_1 + \cdots + b_r w_r$$

and now we have found a way to write v as a linear combination of the set we wanted, so the set spans V.

Since $(u_1, \cdots, u_k, w_1, \cdots, w_r)$ is linearly independent and spanning, it forms a basis. □

Notice that this is really similar to the counting formula: $|G| = |\ker \phi||\operatorname{Im} \phi|$. In particular, it's the same formula in some cases! Consider $F = \mathbb{F}_p$, the integers mod p, and consider some homomorphism ϕ. Then $\dim V = n, \ker \phi = k, \operatorname{Im} \phi = n - k$, and $p^n = p^k p^{n-k}$.

We'll shift gears now and consider a **change of basis**. Let B and B', C and C', and A and A' be old and new bases of V, W, and matrices representing T, respectively. Then recall that we can write $B' = BP, C' = CQ$ for invertbiel

matrices P, Q. Since A is defined by $T(B)X = CAX$, substituting in $B = B'P^{-1}, C = C'Q^{-1}$, we are left with

$$T(B'P^{-1}) = (C'Q^{-1}A) \implies T(B')P^{-1} = C'Q^{-1}A$$

so we know that $T(B') = C'(Q^{-1}AP)$. But we also have that $T(B') = C'A'$, so we actually have the following result:

Fact 90

Let A represent a linear transformation from V to W. Then $A' = Q^{-1}AP$ is the change-of-basis matrix for A if P and Q are the change of basis matrices for V and W, respectively.

Recall that the elementary matrices generate the general linear group: this means that we can do arbitrary row operations on A, as well as arbitrary column operations. (doing either row or column first). This means that we can make A' very simple – In fact, we can make it almost an identity matrix with some zeros on the outside. We'll come back to this later.

10 September 28, 2018

Consider some linear transformation $V \xrightarrow{T} V$ with a matrix $A : B \to B$. (This is also known as a **linear operator.**) This transformation T has the property that if X is the coordinate vector of $v \in V$ with respect to the basis B, then $Y = AX$ is the coordinate vector of $T(v)$.

Example 91

Rotate the plane by an angle θ. Then the corresponding matrix is of the form

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Notice that the first column is the final location of the first basis vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and the second column is the final location of the second basis vector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

Example 92

Let P be the space of polynomials in t with degree at most 3. Then the map of differentiation $P \xrightarrow{\frac{d}{dt}} P$ is a linear operator. Taking our basis to be $(1, t, t^2, t^3)$, the matrix is of the form

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Remember that in general, we can change our basis to be more convenient. If we change both the new and old basis from $B \to B' = BP$, then our new matrix will be of the form

$$A' = P^{-1}AP \text{ (conjugation by } P^{-1})$$

In other words, we have to do a row operation, as well as a column operation, to get to our final matrix.

Example 93

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, and let E be the elementary matrix $\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}$.

Then $E^{-1}AE$ will be

$$\begin{bmatrix} 1 & -x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a - xc & xa - xcx + b - xd \\ c & cx + d \end{bmatrix}$$

This is not super clean. However, we can pick x appropriately so that $a - xc$ or $cx + d$ is zero, as long as $c \neq 0$. Thus, we can write the matrix as $\begin{bmatrix} 0 & b \\ c & d \end{bmatrix}$, and then conjugating by $\begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}$, we end up with a matrix of the form $\begin{bmatrix} 0 & b \\ 1 & d \end{bmatrix}$ for new b, d.

Fact 94

This is called the **rational canonical form.** This is the best we can do: we can fix where one of our basis vectors go, but not both.

Fixing the first column is actually related to the concept of an eigenvector:

Definition 95

Let T be a linear transformation. An **eigenvector** of T is a nonzero vector $v \in V$ such that $Tv = \lambda v$ for some $\lambda \in \mathbb{R}$. λ is the corresponding **eigenvalue** of v.

In particular, if we pick $(v_1, \cdots, v_n)$ to be a basis of eigenvectors, then the matrix A is diagonal with entries equal to the eigenvalues. (This is really nice!)

Example 96

Let $V = \mathbb{R}^2$, and let T be the transformation represented as multiplication by the matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$.

This sends the standard basis vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to the columns of A, $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$. Thus, the first quadrant gets sent to a more narrow slice of the first quadrant. Apply the transformation again: $A^2 = \begin{bmatrix} 5 & 5 \\ 5 & 10 \end{bmatrix}$, and we have an even more narrow slice. Keep doing this, and we get a line - that's a line of eigenvectors! We can state this as a more general result:

Theorem 97

A real matrix with positive entries has a ray of eigenvectors with all positive entries.

The "proof" is basically the same for more dimensions: we apply our transformation repeatedly.

Example 98

The eigenvectors of the differentiation map are of the form $(c, 0, 0, 0, \cdots)$ with eigenvalue 0. (Yes, this is an allowed eigenvector.)

In general, how do we find eigenvectors? **First, find the eigenvalues!** Assume x is a nonzero eigenvector with eigenvalue λ: then

$$Ax = \lambda x \implies (A - \lambda I)x = 0 \implies \det(A - \lambda I) = 0$$

($A - \lambda I$ must be singular to have a nonzero solution, so its determinant is 0).

Example 99

The eigenvalues of a 2 by 2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ must satisfy $(a-\lambda)(d-\lambda)-bc = 0 \implies \lambda^2-(a+d)\lambda+(ad-bc) = 0$. So there are at most two eigenvalues for a 2 by 2 matrix.

$a + d$ is the **trace** of the matrix; it is the sum of the diagonal entries but also the sum of the eigenvalues.

Example 100

Let R be the rotation matrix $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

Then the characteristic polynomial is $\lambda^2 - (2\cos\theta)\lambda + 1$, which has no real roots unless $\cos\theta = \pm 1$. Then we get eigenvalues $\cos\theta \pm \sqrt{\cos^2\theta - 1} = \cos\theta \pm i\sin\theta = e^{\pm i\theta}$. This results in corresponding eigenvectors are $\begin{bmatrix} 1 \\ \pm i \end{bmatrix}$.

In general, how would we find the determinant of $tI - A$ (flipping the sign so the characteristic polynomial is nicer)? We'll call this polynomial $p(t)$. The leading coefficient is t^n (from the diagonal), and the t^{n-1} coefficient is $-\sum a_{ii}$, which is the negative of the trace of the matrix. (All the coefficients in the middle are a huge mess by Vieta's formula.) Finally, the constant term is easy: it's just $(-1)^n \det A$, since the ts cannot contribute. Thus, the characteristic polynomial $p(t)$ of an $n \times n$ matrix is a degree n polynomial in t with some additional nice properties.

In particular, consider the case $F = \mathbb{C}$. In general, the characteristic polynomial $p(t)$ has n distinct roots, and this gives us a nice result:

Lemma 101

Let $v_1, \cdots, v_k$ be eigenvectors with eigenvalues $\lambda_1, \cdots, \lambda_k$. If all λs are distinct, then $(v_1, \cdots, v_k)$ are independent. Thus, we can form a basis with our eigenvectors.

Proof. Suppose $\boxed{a_1 v_1 + a_2 v_2 + \cdots + a_k v_k = 0}$. We wish to show that $a_i = 0$ for all $1 \le i \le k$. We induct on k; this is clearly true for $k = 0, 1$.

For the inductive step, assume this is true for all values less than k. Thus, $v_1, \cdots, v_{k-1}$ are independent. Apply the linear transformation T; now $T(0) = 0$, so

$$\sum A a_i v_i = \boxed{\sum \lambda_i a_i v_i = 0}.$$

Now multiplying the first boxed expression by λ_k, $\boxed{\sum \lambda_k a_i v_i = 0}$. Subtracting these, the $\lambda_k a_k v_k$ terms cancel, so we now have $\sum a_i(\lambda_i - \lambda_k)v_i = 0$ for $1 \le i \le k-1$; $\lambda_i - \lambda_k$ can't be zero, since all eigenvalues are distinct. Since

$v_1, \cdots, v_{k-1}$ are independent by inductive assumption, $a_i = 0$ for all $1 \le i \le k-1$, and therefore $a_k v_k = 0 \implies a_k = 0$ as well, showing that the v_i are independent. □

11 October 1, 2018

Let $V \xrightarrow{T} V$ be a linear transformation, and let A be a matrix with respect to some basis. Recall that the **characteristic polynomial** $p(t) = \det(tI - A)$ is a degree n polynomial in t, such that the coefficients of $t^n, t^{n-1}, 1$ are $1, -\operatorname{tr} A, (-1)^n \det A$, respectively.

Notice that the polynomial is independent of basis, because the roots are always the same (they are the eigenvalues of the linear transformation).

Corollary 102

Under a change of basis (so $A' = PAP^{-1}$), the trace of the matrices A and $P^{-1}AP$ are always the same. This also means that the trace of AB and BA are always the same: set $A = BA, P = B$.

Example 103

Recall that we can always get a matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ into the form $A' = \begin{bmatrix} 0 & b' \\ 1 & d' \end{bmatrix}$. We can now see that we are stuck (and can't simplify further) for sure: d' must be the trace of A, and $-b$ must be the determinant of A.

Let's assume that our vector space V is over the field $F = \mathbb{C}$. If $p(t)$ has all distinct roots $\lambda_1, \cdots, \lambda_n$, and v_i are their corresponding eigenvectors, then $(v_1, \cdots, v_n)$ form a basis, and A is the diagonal matrix (in that basis) with entries λ_i. Most polynomials have all distinct roots, so this is usually fine.

But suppose our characteristic polynomial has a double root; for example, let $\dim V = 2$. Then the characteristic polynomial is $(t - \lambda)^2$, and we can pick an eigenvector v_2. If (v_1, v_2) is a basis for V, in that new basis, we have $A' = \begin{bmatrix} \lambda & 0 \\ c & \lambda \end{bmatrix}$ for some constant c, and we can't always **diagonalize** our matrix. In particular, if $c \neq 0$, then v_2 and all of its multiples are the only eigenvectors, and we have a degenerate case, since

$$\begin{bmatrix} \lambda & 0 \\ c & \lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \lambda x \\ cx + \lambda y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$$

only if $c = 0$ or $x = 0$; if $c \neq 0$ then all eigenvectors are $\begin{bmatrix} 0 \\ y \end{bmatrix}$, which in this basis means they are multiples of v_2.

So how do we deal with these degenerate cases? We need the notion of decomposing a matrix into blocks:

Definition 104

Let W_1 and W_2 be subspaces of V. V is called the **direct sum** of W_1 and W_2, denoted $V = W_1 \oplus W_2$, if all $v \in V$ can be written as $v = w_1 + w_2$ in a unique way (where $w_1 \in W_1$ and $w_2 \in W_2$).

It is equivalent to say that $W_1 \cap W_2 = \{0\}$ and that $W_1 + W_2 = V$.

Definition 105

Let $W \subset V$ be a subspace. Then W is **T-invariant** if $TW \subset W$.

Suppose $V = W_1 \oplus W_2$, and W_1, W_2 are T-invariant. Then let $(v_1, \cdots, v_c)$ be a basis for W_1 and $(v_{c+1}, \cdots, v_n)$ be a basis for W_2. Then we can write the matrix of T with respect to this new basis as

$$M = \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}$$

where A and D are square matrices. This is because $v_1, \cdots, v_c$ all go to some linear combination of the first c basis vectors, and similar for $v_{c+1}, \cdots, v_n$.

Definition 106

A **Jordan block** is any of the following matrices for a fixed λ:

$$\begin{bmatrix} \lambda \end{bmatrix}; \begin{bmatrix} \lambda & 0 \\ 1 & \lambda \end{bmatrix}; \begin{bmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 1 & \lambda \end{bmatrix} \cdots$$

Theorem 107 (Jordan normal form)

Let $V \xrightarrow{T} V$ be a linear operator on a complex vector space. Then there exists a basis such that the matrix is made up of Jordan block matrices:

$$\begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{bmatrix}$$

where the λ_i for each J_i may be equal to each other or not.

This has a lot of zeros, and in some ways it's the "simplest" way we can represent our matrix.

Proof. Choose an eigenvalue λ of T. Replace T by $T - \lambda I$; now we can assume T has eigenvalue 0 because T is not invertible. We will construct a Jordan block for λ.

Let $N_1 = \ker T$; this is not the zero space (for the reason above). Let $N_2 = \ker T^2$, and so on. Anything "killed" (sent to zero) by T is certainly killed by T^2, which is certainly killed by T^3, and so on, so we have a chain of subspaces $\{0\} \subset N_1 \subset N_2 \subset \cdots$. Similarly, let $W_1 = \operatorname{Im} T, W_2 = \operatorname{Im} T^2, \cdots$. For a similar reason, if something can be written as $T(T(v))$, it can be written as $T(v)$. Thus, we also have the chain of subspaces $V \supset W_1 \supset W_2 \supset \cdots$.

These are infinite chains, but we have a finite dimensional vector space. Thus, there exists some k such that $N_k = N_{k+1} = \cdots = N$, and $W_k = W_{k+1} = W_{k+2} = \cdots = W$. By the dimension formula, $\dim N_i + \dim W_i = \dim V$, so it is also true that $\dim V = \dim N + \dim W$. Now we need to prove a quick result:

Lemma 108

N and W are T-invariant, and $V = N \oplus W$.

Proof of lemma. N and W are T-invariant by definition (we got to the constant part of the sequence). In particular, everything in the kernel of T^k is still sent to something in the kernel of T^{k+1}, and the same idea is true for the image. By the dimension formula, if we can show that the intersection of N and W is trivial, we will have shown that $V = N \oplus W$, since the basis for N and basis for W gives everything for V.

Assume there is some $x \in N \cap W$. Then $x = T^k v$ for some v, but also $T^k x = 0$. So $T^k(T^k v) = 0 \implies T^{2k} v = 0$, so $v \in N_{2k}$. Thus $v \in N_k$ as well, so $T^k v = 0$, so $x = 0$, completing the proof. □

So by the work we did above, we can choose a basis for V such that the first part is for N and the second part is for W. By block decomposition, we can now write our matrix as

$$M = \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}$$

Since N is not trivial, A has size at least 1, so now we induct on D: repeat the process again with our new eigenvalue. So all we need to do is show that A is a Jordan block.

Since A is the matrix for T on the space N, there exists some k such that $T^k = 0$; this means T is nilpotent. To be continued... □

12 October 3, 2018

Today, we'll take the field $F = \mathbb{R}$.

Definition 109

Define the **dot product** of two column vectors as $x \cdot y = \sum x_i y_i = x^T y$.

This has the following important properties:

- It is linear in x and y.
- It is symmetric.
- $x \cdot x = x_1^2 + \cdots + x_n^2 = |x|^2$.
- $x \cdot y = |x||y| \cos\theta$, where θ is the angle between the vectors x and y. This is true in $\mathbb{R}^n$, but we have to be more careful with defining an "angle."

The proof of this last statement is the law of cosines: make a triangle with vectors $\vec{x}, \vec{y}, \vec{x} - \vec{y}$. Then

$$|x - y|^2 = |x|^2 + |y|^2 - 2|x||y| \cos\theta = (x - y) \cdot (x - y) = |x|^2 + |y|^2 - 2x \cdot y$$

and the result follows.

Definition 110

Let $(v_1, \cdots, v_n)$ be a basis of $\mathbb{R}^n$. Such a basis is **orthogonal** if $v_i \cdot v_j = 0$ for all $i \neq j$, and it is **orthonormal** if

$$v_i \cdot v_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Notice that the standard basis is orthonormal (all vectors are orthogonal and all have length 1).

Lemma 111

If $v_1, \cdots, v_k$ are orthogonal and nonzero, then they are independent.

Proof. Assume we have a linear combination $c_1v_1 + \cdots + c_kv_k = 0$. Then for all i,

$$0 = v_i \cdot (c_1v_1 + \cdots + c_kv_k) = v_i \cdot (c_iv_i) = c_i,$$

so $c_i = 0$. Thus the only linear combination is the trivial one, and thus the v_is are independent. □

Definition 112
A linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is **orthogonal** if $T(x) \cdot T(y) = x \cdot y$ for all x, y.

Let A be the corresponding matrix for T. Then we can rewrite this as

$$(Ax) \cdot Ay = x \cdot y \implies (Ax)^T(Ay) = x \cdot y \implies x^TA^TAy = x^Ty,$$

so $A^TA = I$ is a **sufficient** condition for A to be orthogonal, and it turns out to be **necessary** as well:

Lemma 113
$x^TMy = x^Ty$ for all x, y if and only if $M = I$.

Proof. It's pretty clear that the equation is true if $M = I$. For the other direction, let $x = e_i, y = e_j$ (the vectors with 1 in that entry and 0 everywhere else). Then x^TMy is entry M_{ij}, but it is also equal to $e_i^Te_j$, which is also the Kronecker delta (1 if $i = j$ and 0 otherwise). Thus M must be the identity. □

Definition 114
A square matrix A is **orthogonal** if $A^TA = I$, or equivalently that $A^T = A^{-1}$.

This is equivalent to having an orthogonal operator, and it is also equivalent to having the columns of A forming an **orthonormal basis**.

It turns out that the orthogonal matrices form a group $O_n \subset GL_n$! This is pretty easy to check: if $A^TA = I, B^TB = I$, then $(AB)^TAB = B^TA^TAB = B^TB = I$, so $A, B \in O_n \implies AB \in O_n$. I is clearly in the group, and $A^T = A^{-1} \implies A^{-1^T} = A$ by taking the transpose of both sides.

Fact 115
Notably, $\det A^T = \det A$, so the determinant of an orthogonal matrix is ± 1.

Definition 116
Let SO_n be the **special orthogonal group**, consisting of the elements of O_n with determinant 1. (SO_n creates two cosets in O_n.)

Let's describe these orthogonal matrices in dimension 2 and 3: we want all the column vectors to have length 1 and be orthogonal. In dimension 2, any vector of length one can be written as $\begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}$, so our matrix must be written in the form

$$A = \begin{bmatrix}\cos\theta & b \\ \sin\theta & d\end{bmatrix}.$$

Notice that rotation matrices work: $R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ are good. Then $R^T = R^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$, and by closure, $R^T A$ is an orthogonal matrix as well. But the first column of $R^T A$ is $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, so the second column must be $\begin{bmatrix} 0 \\ \pm 1 \end{bmatrix}$. Plugging this back in, we find that A must be of one of the two forms

$$\begin{bmatrix} c & -s \\ s & c \end{bmatrix}, \begin{bmatrix} c & s \\ s & -c \end{bmatrix}$$

where $c = \cos\theta, s = \sin\theta$ for some θ. Thus O_2 is the set of rotation / reflection matrices, and SO_2 is the set of rotation matrices. Let R be SO_2 and S be $O_2 \setminus SO_2$; these are the two cosets of SO_2.

Fact 117

Consider the characteristic polynomial of a matrix $M \in S$. By direct computation, the polynomial is always $\lambda^2 - 1$, so there is an eigenvector x with eigenvalue 1 and an eigenvector y with eigenvalue -1. Then $Mx = x, My = -y$, so

$$x \cdot y = (Mx) \cdot (My) = x \cdot (-y) \implies x \cdot y = 0$$

So x and y are orthogonal, and therefore we can describe any matrix in S as fixing a line and reflecting everything else over it.

It turns out that the angle α of the fixed line of eigenvectors from the x-axis is $\alpha = \frac{1}{2}\theta$. This is because we can get from a matrix in S to a matrix in R (a rotation matrix) by changing the sign of the second column. This is a multiplication by $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, which flips over the x-axis. So a vector p on the line of reflection at α is sent to $-\alpha$. But then rotating back by the matrix in R gets us back to p. Thus, $\alpha - (-\alpha) = \theta$, and $\alpha = \frac{1}{2}\theta$.

Theorem 118

Any matrix $A \in SO_3$ is a **rotation** of $\mathbb{R}^3$ (where we fix a line as an axis of rotation and rotate the orthogonal plane by some angle).

This is not true in $\mathbb{R}^4$, since we can take two orthogonal planes and rotate them in different angles.

Proof. First, we will find a fixed vector with eigenvalue 1. We do this by showing $\det(A - I) = 0$. We know that $\det(A^T(A - I)) = \det(A - I) = \det(I - A^T)$, since $\det A^T = 1$. So $\det(A - I) = \det((I - A)^T) = \det(I - A)$. But the determinant of the negative of a 3 by 3 matrix adds a factor of -1. Thus $\det(A - I) = 0$, so $\lambda = 1$ is an eigenvalue. □

The rest of the proof is using this eigenvector of eigenvalue 1 as one of our basis vectors in an orthonormal basis – to be continued.

13 October 5, 2018

Definition 119

Let F be a figure in the plane. A **symmetry** of F is a bijective, distance-preserving map from $F \xrightarrow{S} F$.

Example 120

Consider the symmetries of a pentagon. There are 5 rotations, including the identity, and we can reflect over one of the 5 lines of bilateral symmetry. This is a total of 10 symmetries.

Fact 121

The symmetries of a figure F always form a group.

This is not hard to prove: the identity is a symmetry, we can invert a symmetry, and we can compose symmetries. Often, we can think of extending the symmetry to the whole plane.

Definition 122

An **isometry** of $\mathbb{R}^n$ (mostly $n = 2$) is a map $\mathbb{R}^n \xrightarrow{f} \mathbb{R}^n$ that is distance-preserving: we have

$$d(f(x), f(y)) = d(x, y) \forall x, y \in \mathbb{R}^n$$

where $d(x, y) = |x - y|$.

Isometries are bijective maps, though this may not be obvious.

Example 123

Any orthogonal linear operator ϕ is an isometry.

This is because we know that $\phi(x) \cdot \phi(y) = x \cdot y$, so

$$|\phi(x) - \phi(y)|^2 = (\phi(x) - \phi(y)) \cdot (\phi(x) - \phi(y)) = \phi(x - y) \cdot \phi(x - y) = (x - y) \cdot (x - y) = |x - y|^2,$$

and distances are indeed preserved.

Example 124

Translations t_a by a vector a of the form $t_a(x) = x + a$ are also isometries.

(It is pretty obvious that these preserve distance: $t_a(y) - t_a(x) = y - x$.)

Theorem 125

Let f be an isometry, and suppose $f(0) = 0$. Then f is an orthogonal linear operator.

Proof by Sharon Hollander. Orthogonality is easy: since $|f(x) - f(y)| = |x - y|$, setting $y = 0$,

$$(f(x) - f(0)) \cdot (f(x) - f(0)) = (x - 0) \cdot (x - 0) \implies f(x) \cdot f(x) = x \cdot x$$

So now, we can expand $(f(x) - f(y)) \cdot (f(x) - f(y)) = (x - y) \cdot (x - y)$:

$$\phi(x) \cdot \phi(x) - 2\phi(x) \cdot \phi(y) + \phi(y) \cdot \phi(y) = x \cdot x - 2x \cdot y + y \cdot y$$

and canceling, $\phi(x) \cdot \phi(y) = x \cdot y$.

Showing that the operator is compatible with $+$ is a bit harder. We want to show $\phi(x + y) = \phi(x) + \phi(y)$, which is the same as saying $\phi(x + y) - \phi(x) - \phi(y) = 0$, so the length must be 0. Thus, we want to show

$$(\phi(x + y) - \phi(x) - \phi(y)) \cdot (\phi(x + y) - \phi(x) - \phi(y)) = 0$$

Expanding this out,

$$\phi(x + y) \cdot \phi(x + y) + \phi(x) \cdot \phi(x) + \phi(y) \cdot \phi(y) - 2\phi(x + y)\phi(x) - 2\phi(x + y)\phi(y) + 2\phi(x)\phi(y) = 0$$

Since we know ϕ is orthogonal, we can basically drop all the ϕs: this becomes

$$(x + y) \cdot (x + y) + x \cdot x + y \cdot y - 2(x + y) \cdot x - 2(x + y) \cdot y + 2x \cdot y$$

and factoring,

$$((x + y) - x - y) \cdot ((x + y) - x - y) = 0$$

which is true. So now follow all of the steps in reverse! □

Corollary 126

All isometries are a composition of an orthogonal linear operator and (then) a translation. Specifically, $f(0) = a \implies f = t_a \circ \phi$ for some orthogonal operator ϕ.

Proof. Define $\phi = t^{-a} \circ f$. Then $\phi(0) = 0$ and Theorem 125 tells us that ϕ is an orthogonal operaotr. □

In particular, since translations and orthogonal operators are both bijective, we've shown that all isometries are bijective.

So now, let's look at the special case $\mathbb{R}^2$. We know that the orthogonal operators are either of the form

$$\rho_\theta = \begin{bmatrix} c & -s \\ s & c \end{bmatrix},$$

where $c = \cos\theta, s = \sin\theta$, or

$$\rho_\theta r = \begin{bmatrix} c & s \\ -s & c \end{bmatrix},$$

taking r to be the standard reflection matrix $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

So all isometries of $\mathbb{R}^2$ are of the form $t_a\rho_\theta$ or $t_a\rho_\theta r$. There are some rules that help with composition:

- $t_a t_b = t_{a+b}$.
- $\rho_\alpha \rho_\beta = \rho_{\alpha+\beta}$.
- $r^2 = 1$, the identity isometry.
- We know that $\rho_\theta t_a(x) = \rho_\theta(x + a) = \rho_\theta(x) + \rho_\theta(a)$. Thus, $\rho_\theta t_a = t_{\rho_\theta(a)}\rho_\theta$.

- Similarly, $rt_a = t_{r(a)}r$.
- Finally, $r\rho_\theta = \rho_{-\theta}r$.

It's more interesting to think of geometric descriptions of these isometries. If there is no r, then we can either have t_a, a translation, or $t_a\rho_\theta$, which turns out to be a rotation of the same angle θ, just about a different point. These are the **orientation-preserving** isometries.

Then we have the **orientation-reversing** isometries: $\rho_\theta r$ is a reflection over angle $\frac{1}{2}\theta$, and finally, $t_a\rho_\theta r$ is a **glide reflection**:

Definition 127

A **glide reflection** is defined by a glide line and a glide vector parallel to it. The glide reflection first reflects about the glide line and then translates by the glide vector.

Let's try to write this out. What does it mean to rotate an angle θ around some point $\mathbf{p}$? We first apply a translation so $\mathbf{p}$ is at the origin, then rotate, then translate back. Thus, we have

$$\rho_{\theta,\mathbf{p}} = t_{\mathbf{p}}\rho_\theta t_{-\mathbf{p}} = t_{\mathbf{p}}t_{\rho_\theta(-\mathbf{p})}\rho_\theta = t_{\mathbf{p}}t_{-\rho(\mathbf{p})}\rho_\theta = t_{\mathbf{p}-\rho(\mathbf{p})}\rho_\theta.$$

So to show that $t_a\rho_\theta$ is a rotation about some point for $\theta \neq 0$, we just need to show that $a = \mathbf{p} - \rho(\mathbf{p})$ for some p. Well, let R be the matrix corresponding to ρ. Then

$$a = \mathbf{p} - \rho(\mathbf{p}) = (I - R)\mathbf{p}$$

and as long as $I - R$ has nonzero determinant, it is invertible and we can find the point $\mathbf{p}$. Well, let's write it out explicitly:

$$\det \begin{bmatrix} 1-c & s \\ -s & 1-c \end{bmatrix} = 1 - 2c + c^2 + s^2 = 2 - 2c$$

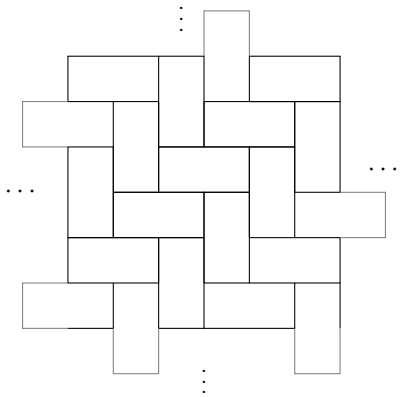
so this is not zero unless $\cos\theta = 1$, which means $\theta = 0$. But we assumed $\theta \neq 0$ (otherwise we just have a translation), so $I - R$ is invertible where relevant.

By the way, there is also a nice geometric construction for finding $\mathbf{p}$ such that $\mathbf{p} - \rho(\mathbf{p}) = a$. Basically, place the angle θ on the perpendicular line to a.

14 October 12, 2018

The quiz grades will be posted by tomorrow. Also, there is a student who has not taken the quiz yet, so the solutions will be posted after that.

Here's a picture of a plane figure F (it is infinite in all directions):



There are no immediate lines of reflection for this shape, unlike the normal brick pattern, which has two lines of reflection. Let's think about the group of symmetries: what transformations take this back to itself?

Fact 128

We can pick the center of any brick and rotate around its center by 180 degrees. Also, we can do a translation by integer combinations of two different vectors a and b: denote this as $a\mathbb{Z} + b\mathbb{Z}$.

So the symmetries of our figure F are isometries of the plane P such that $F \xrightarrow{f} F$ carries the figure to itself. In other words, the group of symmetries of F are a **subgroup** G of the group M of all isometries of P.

Recall that we have elements of M of the form $m = t_a\rho_\theta$ or $m = t_a\rho_\theta r$, where ρ_θ is the rotation by θ around the origin, and r is the standard reflection over the x-axis. In the first case, if $\theta = 0$, this is a translation, and otherwise, it's a rotation obout some point $p \in P$. The second case consists of glide reflections: reflect over some line and then glide by a vector in the parallel direction. In particular, if we decompose a into parallel and perpendicular components to the line of reflection of $\rho_\theta r$, the actual glide line goes through the midpoint of the perpendicular component.

Notice that changing the origin that we're rotating around adds a conjugation factor to our isometries.

Fact 129

There is an important homomorphism $M \xrightarrow{\pi} O_2$, where O_2 denotes the set of orthogonal operators. Basically, take any $t_a\rho_\theta \to \rho_\theta$, and take $t_a\rho_\theta r \to \rho_\theta r$.

In particular, we send $t_a\phi \to \phi$ for any orthogonal operator ϕ. Remember that the group of isometries is not abelian!

It turns out this is a **homomorphism**, and we'll show that now. Recall the rule for moving translations past operators: $\phi t_a = t_{\phi(a)}\phi$. So

$$\pi((t_a\phi)(t_b\psi)) = \pi(t_a t_{\phi(b)}\phi\psi)) = \phi\psi = \pi(t_a\phi)\pi(t_b\psi)$$

since $\phi\psi$ is an orthogonal operator. This proves that π is always a homomorphism, and the kernel of π is all translations T of the form $\{t_a\}$ where $a \in V = \mathbb{R}^2$.

Remark 130. *Here, we should make a distinction between V, the plane of vectors, and P, the plane of points. t acts on P via $t_v(p) = p + v$; notice that we pick a specific zero vector or origin for V, but we do not need to do so for the plane of points P.*

Let's make a small assumption about the set of symmetries for our brick figure: we want it to be discrete.

Definition 131

A subgroup G of M is **discrete** if it does not contain arbitrarily small rotations (about any point) or translations.

In other words, there exists some ε_1 such that $|v| > \varepsilon_1$ if t_v is in our subgroup of symmetries and $v \neq 0$. Similarly, there exists some ε_2 such that $|\theta| > \varepsilon_2$ if $\rho_{\theta,p} = t_a \rho_\theta \in G$ and $\theta \neq 0$.

Example 132

The symmetries of a circle are not discrete: we can rotate by arbitrarily small angles. Similarly, the symmetries of a line can be translated by any arbitrarily small length in the direction of the line.

So if G is a discrete subgroup of M, and $M \xrightarrow{\pi} O_2$ is a homomorphism with kernel T, we can think about the map $G \xrightarrow{\pi'} \overline{G}$, which is the image of G under π in O_2. Then the kernel of π' is exactly the translations $T_G = T \cap G$ in our discrete subgroup.

Definition 133

The image of G under π, denoted $\overline{G}$, is called the **point group** of the group of symmetries.

What does the point group "remember" about G? Any rotation in G becomes just information about the **angle** of rotation; the point which the rotation was done about is lost. Similarly, in a glide reflection, we remember the angle or **slope of the glide line** $\alpha = \frac{1}{2}\theta$. (The translation is completely lost.) So now, let's call the corresponding rotations and reflections $\overline{\rho}_\theta$ and $\overline{\rho}_\theta \overline{r}$.

So now the point group is easy to find. We know that in a **standard brick pattern**, we can rotate by π (either from the center of the brick or an edge, but those are equivalent). Thus for the standard brick pattern, the point group

$$\overline{G} = \{\overline{1}, \overline{\rho}_\pi, \overline{r}, \overline{\rho}_\pi \overline{r}\} :$$

we can either reflect over a line or rotate by 180 degrees. This is D_2, the dihedral group, which is equivalent to the Klein four group.

What about the plane figure F? It's a bit more confusing to visualize, but we have a 180 degree rotation about any center of a brick, which can be denoted $\overline{\rho}_\pi$, and we actually have a glide reflection at a 45 degree angle. So it turns out that this brick pattern also has the same point group $\overline{G}$! However, the actual groups of symmetries G are different for the two symmetry groups.

What more can we say about the kernel T? If G is a discrete subgroup of M, the corresponding point groups $\overline{G}$ are discrete subgroups of O_2. In addition, $T_G = \{t_a \in G\}$ is also a discrete subgroup of M. Specifically, $t_a \iff a$ gives a correspondence between the set of translations T_G and the set of points $L_G \subset V = \mathbb{R}^2$. In other words, since T_G is discrete, L **is a discrete subgroup of** V^+! In particular, G being discrete is equivalent to $\overline{G}$ and L both being discrete.

So what are the discrete subgroups of O_2 and of V^+? We can describe them pretty easily here:

Proposition 134

If L is a discrete subgroup of the vectors $V^+ = \mathbb{R}^2$, then L is either $\{0\}$, $a\mathbb{Z}$ for some vector a, or $a\mathbb{Z} + b\mathbb{Z}$ for a, b linearly independent.

If G is discrete and $L_G = a\mathbb{Z} + b\mathbb{Z}$, then we call G a **plane crystallographic group**. In discrete groups where the translations are $a\mathbb{Z} + b\mathbb{Z}$ in two different directions, then there are 17 possible groups! In three dimensions, there are

more than 200 of them. If we want, we can search up crystallographic groups, but they're pretty annoying and not that useful.

Proposition 135
If $\overline{G}$ is a discrete subgroup of O_2, then it is the cyclic group C_n generated by $\overline{\rho}_\theta$ where $\theta = \frac{2\pi}{n}$, or D_n, the dihedral group, which can be represented as $\{\overline{\rho}_n, \overline{\rho}_n\overline{r}'\}$, where r' is a reflection.

D_n is the symmetry group of a regular n-gon. For example, D_5 gives 10 symmetries of a pentagon: rotate in one of 5 ways, or reflect in one of 5 ways.

15 October 15, 2018

We will continue talking about plane figures. Like last time, let G be a discrete subgroup of M, which is an isometry of the plane P. Recall that we have a homomorphism $\pi : M \to O_2$, which sends an isometry $t_a\rho_\theta \to \overline{\rho}_\theta$ and $t_a\rho_\theta r \to \overline{\rho}_\theta\overline{r}$; basically, it drops the translation. Then $G \subset M$, so the **point group** $\overline{G} \subset O_2$ is an important group to study.

Recall that the kernel of π is the set of translations in G, and we can define $L = T \cap G$ to be the set of vectors $\{v \in \mathbb{R}^2 \mid t_v \in G\}$. We were only dealing with discrete subgroups, so L is discrete. We also know that $\overline{G}$, the point group, is either the cyclic group of n elements C_n or the dihedral group of $2n$ elements D_n.

Fact 136 (Simple examples)
C_1 is the trivial group, D_1 is the group generated by $\overline{r}$, and C_2 is the group generated by $\overline{\rho}_\pi$.

Also recall that in a plane, L is either trivial, $a\mathbb{Z}$ for some vector $a \neq 0$, or $a\mathbb{Z} + b\mathbb{Z}$ for linearly independent vectors a, b. In this last case, we have a lattice. Let's consider some special cases:

Example 137
What if $L = 0$? In other words, what if the kernel of $G \xrightarrow{\pi} \overline{G}$ is trivial?

Then the only translation in the group of isometries is the identity. (An example of this is a pie.) Here, $G \to \overline{G}$ is bijective, so G is also the cyclic or dihedral group. As a fun fact, dihedral groups D_n are the symmetries of an n-gon in the plane.

Lemma 138
If G is a finite group of isometries of the plane, and $L = 0$, then there exists a fixed point $p \in P$ such that $gp = p$ for all g in G.

This is not obvious; how do we know that all rotations, reflections, and so on all keep a point fixed?

Proof. Start with an arbitrary point $q \in P$, and define its **orbit** to be $\{q_1, \cdots, q_N\} = \{gq \mid g \in G\}$, where $N = |G|$. Basically, multiply q by all possible elements in G (which means that we look at all possible images of q, counting multiplicity). Then the fixed point will be the center of mass or **centroid**

$$p = \frac{1}{N}(q_1 + \cdots + q_N).$$

To show that p is fixed, we'll show that if g is any isometry, and $q_1, \cdots, q_N$ are any points, the centroid of $\{gq_1, \cdots, gq_N\}$ is gp, where p is the centroid of $q_1, \cdots, q_N$. This is enough to show the result, because multiplying $q_1, \cdots, q_N$ by an element $g \in G$ will result in those same points in some other order, so their average p remains constant.

It is enough to show this for the two cases where g is a translation or an orthogonal operator. If $g = t_v$, then $t_v = q_i + v$, and

$$\frac{1}{N}\sum t_v q_i = \frac{1}{N}\sum q_i + v = t_v \frac{1}{N}\sum q_i = t_v p.$$

Meanwhile, if g is an orthogonal operator, we just use the fact that $g(x+y) = g(x) + g(y)$. □

Note that the point group $\overline{G}$ operates on L:

Lemma 139

If $v \in L$ and $\overline{g} \in \overline{G}$, then $\overline{g}v \in L$.

Proof. If $v \in L$, then there is a translation $t_v \in G$. If $\overline{g} \in G$, then there is an element $t_a\overline{g} \in G$ (where $\overline{g}$ is orthogonal). Conjugating t_v by $t_a\overline{g}$,

$$(t_a\overline{g})t_v(t_a\overline{g})^{-1} = t_a(\overline{g}t_v)\overline{g}^{-1}t_{-a} = t_a t_{\overline{g}v}\overline{g}\overline{g}^{-1}t_{-a} = t_{\overline{g}v},$$

so $\overline{g}v \in L$. □

However, G does not necessarily operate on L. For example, it's possible that G contains only glide reflections but not pure reflections.

Example 140

Now let's look at the other extreme. What if $L = a\mathbb{Z} + b\mathbb{Z}$, and we have a **crystallographic restriction** on our isometries?

Theorem 141

If G is a discrete subgroup of M and L is a lattice, then $\overline{G}$ is C_n or D_n, with the additional restriction that we must have one of $n = 1, 2, 3, 4, 6$.

In other words, there are no crystals with five-fold rotational symmetry! There does exist a quasi-crystal, but it does not have translational symmetry.

Proof. Choose $a \in L$ to be (one of) the shortest vector(s) that is not the zero vector. This exists because we have a discrete set of symmetries. Let's say that $\overline{\rho}_\theta \in \overline{G}$: by the previous lemma, $\overline{G}$ operates on L, so $\overline{\rho}_\theta a \in L$.

But L is a lattice, and it is closed under addition, so $a - \rho_\theta a \in L$ as well. Since a was the shortest vector, the length of $a - \rho_\theta a$ must be at least as large as the length of a, and if $\theta < \frac{\pi}{3}$, this is not true. Since $\theta = \frac{2\pi}{n}$, we must have $n \leq 6$.

Now, for $n = 5$, there is a separate argument. Again, pick a to be a shortest vector, and also consider $\rho a, \rho^2 a$ (where ρ is rotation by $\frac{2\pi}{5}$). Now notice that $|a + \rho^2 a| < |\rho a| = |a|$, which is a contradiction, so five-fold symmetry is not possible either. □

So we have already found 10 groups for $\overline{G}$, and it turns out there are 17 groups G in total that result from this. Let's do an example of a computation in which we find G from $\overline{G}$:

Example 142
Say $\overline{G}$ contains $\overline{\rho} = \overline{\rho}_{\pi/2}$, so $\overline{G} = C_4$ or D_4. What are the possible groups G?

Let's say $L \neq 0$ (or the group is obviously C_4 or D_4). Then picking a to be the shortest nonzero vector in L, $b = \overline{\rho}a \in L$ as well. And now any integer linear combination of a and b is also in L.

Claim 143. *$L = a\mathbb{Z} + b\mathbb{Z}$, which forms a square grid.*

Proof of claim. We know that L contains $a\mathbb{Z} + b\mathbb{Z}$, and if there were other points, there would be some v inside our square grid. Then the distance from v to the closest vertex is smaller than that of the side length, contradicting the minimality of $|a|$. □

Now, the elements in G that map to $\overline{\rho}$ in $\overline{G}$ are rotations of $\frac{\pi}{2}$ about points in the plane. Choose our coordinates so that 0 is the rotation point and the side length is 1. So now $a = \begin{bmatrix}1\\0\end{bmatrix}$, $b = \begin{bmatrix}0\\1\end{bmatrix}$, and $L = \mathbb{Z}^2$.

We can now let $H = \{t_v \mid v \in L\}$ be the group of translations. **If $\overline{G} = C_4$,** then the index of H in G is 4, and G will be the union of the cosets $1H \cup \rho H \cup \rho^2 H \cup \rho^3 H$ (that is, combining a 0, 90, 180, or 270 degree rotation with a translation in $\mathbb{Z}^2$). We also know how to multiply using the isometry rules, and the group is determined. For the same reason in general, such a group of isometries is defined for all cyclic groups C_1, C_2, C_3, C_4, C_6.

Meanwhile, **if $\overline{G} = D_4$,** things are more complicated. Then $\overline{r} \in \overline{G}$, so there exists some vector n such that $t_n r \in G$. Ideally, we want to show that r in g, but we only know that for all $v \in L \implies t_v \in H$, we have $t_v t_u r \in G$. Thus, we can only move our vector n to some vector u in the square generated by a and b (it might be the origin, but it might not be).

Now $t_u r \in G$, so $(t_u r)^2 = t_u t_{ru} r^2 = t_{u+ru} \in G$. Let $u = \begin{bmatrix}u_1\\u_2\end{bmatrix}$. We can calculate what u must be: $u + ru$ must be in L, so $\begin{bmatrix}2u_1\\0\end{bmatrix} \in L$: this means $u_1 = 0$ or $\frac{1}{2}$. We can do some more work and find that this only corresponds to two possibilities: u is the origin, or $u = \begin{bmatrix}1/2\\1/2\end{bmatrix}$. u being the origin means that $r \in G$, which is good, but the other case is more confusing. $t_u r$ is a glide reflection along the line $\frac{1}{4}$ of the way up the square grid (horizontally), and then translated to the right by $\frac{1}{2}$. To visualize a group with this symmetry, around all points $(\mathbb{Z}, \mathbb{Z})$, put a counterclockwise-oriented symbol, and around all points $\left(\mathbb{Z} + \frac{1}{2}, \mathbb{Z} + \frac{1}{2}\right)$, put a clockwise-oriented symbol. Indeed, the point group here is D_4.

16 October 17, 2018

Today, we'll start talking about group operations. We'll start with some examples: let G be a group of isometries of the plane P. G **operates** on the set of lines in P: if $f \in M$ is an isometry, and L is a line in P, then L is sent to fL. The isometries also operate on triangles: Δ gets sent to $f\Delta$.

Example 144
Let $G = D_5$, the symmetries of a regular pentagon. Then D_5 operates on the vertices (and also on the edges).

However, group operations don't have to be geometric.

Example 145

Let V be the Klein Four group $\{1, a, b, c\}$, defined such that $a^2 = b^2 = c^2 = 1, ab = ba = c, bc = cb = a, ac = ca = b$. Switching any of a, b, c around doesn't change the properties of the group, so S_3, the symmetric group, operates on V, the Klein Four group!

Fact 146

A game was played in the Kennedy days where the answer would be given and you had to come up with a question. The answer was "9W." Well, the question was, "Do you spell your name with a V, Mr. Wagner?"

Time to actually give a formal definition of what's going on:

Definition 147

Let G be a group and S be a set. An **operation** of G on S is a map $G \times S \to S$ that sends $g, s \to s' = gs$, with the following properties:

- $1s = s$ for all $s \in S$; that is, the identity does nothing to S.
- $g_1(g_2 s) = (g_1 g_2)s$; the associative law holds. As a corollary, $g^{-1}gs = s$, so inverses also do inverse operations on S.
- $gs = gs'$ if and only if $s = s'$.

Notice that this means that operations permute the set S (since they have "inverses")! Also, notice that $gs = g's$ does not tell us that $g = g'$.

Definition 148

Let G operate on S, and let $s \in S$. Then the **orbit** of s is the points

$$\text{Orbit}(s) = \{s' \in S : \exists g \in G, gs = s'\}$$

Example 149

There is always an isometry that sends any line in the plane to any other line. This means that **the orbit of any line is the set of all lines**.

Example 150

On the other hand, the orbit of any triangle in the plane is the set of all triangles congruent to it. (We have to preserve lengths and angles under isometries.)

If $\text{Orbit}(s)$ is the whole set S for all s, then we call the operation **transitive**.

Proposition 151

Orbits of a group operation are basically "cosets" in the sense that they partition S.

Proof. For any element s, $s = 1s \in \text{Orbit}(s)$. So orbits are not empty, and they cover the group, since every element is in an orbit. So, to complete the proof, we want to show that if two orbits $\text{Orbit}(s)$ and $\text{Orbit}(s')$ have nonzero intersection, they are the same set.

Equivalently, we can show that $s' \in \text{Orbit}(s) \implies \text{Orbit}(s') = \text{Orbit}(s)$. This is really not hard; if s' is in the orbit of S, then $s' = gs$ for some g. But then the orbit of s' is $\{xs' \mid x \in G\} = \{xgs | x \in G\}$ and $xg \in G$, so any element in the orbit of s' is in the orbit of s. To get the other inclusion, repeat the argument but using $s = g^{-1}s'$. □

Proposition 152 (First Counting Formula)
Let S be a finite set, and let G operate on S. If the orbits are denoted $O_1, \cdots, O_k$, then

$$|S| = \sum_{i=1}^{k} |O_i|.$$

Definition 153
Let G operate on S, and let $s \in S$. Then the **stabilizer** of s, denoted $\text{Stab}(s)$, is the set of group elements $g \in G$ such that $gs = s$.

$\text{Stab}(s)$ form a subgroup of G. This is not very hard to show - identity, closure, inverses are all satisfied.

Example 154
What's the stabilizer of a fixed triangle in the plane?

It might seem like the only isometry is the identity, since three points fix a plane. But recall that we're acting on the set of triangles, so an equilateral triangle (for example) can have nontrivial rotations or reflections that fix the triangle, **even if they don't fix the vertices or edges themselves**. So the group is the identity $\{1\}$ only, $\{1, r\}$, or D_3, depending on whether the triangle is scalene, isosceles, or equilateral (respectively).

Example 155
Let H be the stabilizer $\text{Stab}(s)$, and let $s' \in \text{Orbit}(s)$; this is the set of points where there exists $g \in G$ with $gs = s'$. What are all elements $x \in G$ that send s to s'?

This is the coset gH. Indeed, if $xs = s' = gs$, then $g^{-1}xs = s \implies g^{-1}x \in \text{Stab}(s)$. Thus, $g^{-1}x \in H \implies x \in gH$. In particular, this means that there is one coset of H for every element in the orbit of s, and that leads us to the following result:

Proposition 156 (Second Counting Formula)
Let G be a group acting on a set S. Then for any element $s \in S$, $|G| = |\text{Orbit}(s)||\text{Stab}(s)|$.

Recall the formula $|G| = [G : H]|H|$ for a subgroup H of G. In this case, if H is $\text{Stab}(s)$, then $\text{Orbit}(s)$ is the index of H in G.

Example 157
SO_3 is the set of rotations of $\mathbb{R}^3$. Take G to be the symmetries of a cube centered at the origin.

We can rotate in the Rubik's cube style by $\frac{\pi}{2}$, around a space diagonal by $\frac{2\pi}{3}$, or around the center of an edge by π. So G is the stabilizer of the cube in SO_3.

A cube has 6 faces. Let's look at the rotations that fix F, the front face. Then the size $|\text{Stab}(F)|$ is 4, since there are four possible rotations of the front square. But the face can be sent to any of the 6 faces in general, so the counting formula tells us that there are $6 \cdot 4 = 24$ symmetries of a cube. Similarly, there are 8 vertices, and the order of the stabilizer of any vertex is 3. Also, there are 12 edges, and the order of the stabilizer of any edge is 2. Notice that

$$6 \cdot 4 = 8 \cdot 3 = 12 \cdot 2 = 24.$$

As a final note, let H be a subgroup of the group G, and let $\mathcal{C}$ be the set of left cosets of H. Then G operates on $\mathcal{C}$ by the law $g, C \to gC$ (in other words, if $C = aH$, then $gC = gaH$). This is a transitive operation, and the stabilizer of the coset $1H$ is just H, so this formula is equivalent to the counting formula from earlier in class.

17 October 19, 2018

Today's topic is **finite rotation groups.** Recall that if a group G operates on a set S, then we define the **orbit** of s to be $\text{Orbit}(s) = \{s' \mid s' = gs\}$ and the **stabilizer** of s to be $\{g | gs = s\}$. Orbits partition S, much like cosets do, and the order of G is always the products of the orders of the orbit and stabilizer of s for some element s.

Corollary 158

If s, s' are in the same orbit, then the order of $\text{Stab}(s)$ and $\text{Stab}(s')$ are the same.

As a sidenote, let's say H is the stabilizer of s, which is $s' = gs$ for some g. Then the stabilizer of s' is the conjugate subgroup gHg^{-1}.

Theorem 159

Let G be a finite subgroup of SO_3, which are the rotations in $\mathbb{R}^3$. Then G must be C_n, a cyclic group, D_n, a dihedral group, or T, O, I, the rotation groups of the tetrahedron, octahedron, and icosahedron, respectively.

(We may ask how to get reflections in the dihedral group using rotations. Well, we can rotate in the third dimension, which gives us an extra degree of freedom.)

Note that the cube and octahedron are **duals**: connecting centers of the cube faces gives an octahedron, and vice versa. This means that they have the same rotation groups! Similarly, the icosahedron and dodecahedron are also duals.

To analyze these, we will look at the group operations on two sets.

Definition 160

The **pole** of $g \neq 1 \in G$ is a unit vector along the axis of rotation.

Every $g \in G$ that is not the identity has two poles (a unit vector and its negative). One nice way to say this is that p is a pole of $g \neq 1$ if $|p| = 1$ and $gp = p$.

Lemma 161

G operates on the set of poles P. In other words, $g \in G, p \in P \implies gp \in P$.

Proof. If $\mathbf{p}$ is a pole, then there exists some ρ in G such that $\mathbf{p}$ is the pole of ρ. In other words, $\rho\mathbf{p} = \mathbf{p}$. Now, if $\mathbf{p}' = g\mathbf{p}$, then

$$g\rho g^{-1}\mathbf{p}' = g\rho(g^{-1}\mathbf{p}') = g\rho\mathbf{p} = g\mathbf{p} = \mathbf{p}',$$

so $g\rho g^{-1}$ fixes $\mathbf{p}'$. And $g\rho g^{-1}$ is not the identity, because that would imply $g\rho g^{-1} = 1 \implies g\rho = g \implies \rho = 1$. Thus $\mathbf{p}'$ is indeed a pole of some element of G. □

With this, we can now prove Theorem 159:

Proof. Since G operates on P, we can decompose P into orbits

$$P = O_1 \cup \cdots \cup O_k.$$

Let's say $|O_i| = n_i$, and let $|G| = N > 1$ (otherwise it's the trivial group C_1). If a pole $p \in O_i$, then its stabilizer $\text{Stab}(p)$ has order $r_i = \frac{N}{n_i}$, depending only on i (which orbit it is in). But we also know that the number of poles is $|P| = |O_1| + \cdots + |O_k|$. (Notice that the stabilizer of a pole p is the set of rotations about that axis, which forms a **cyclic group** of order r_i.)

So now define the set $S = \{(g, p) \mid g \neq 1 \in G, p \text{ pole of } g\}$. (Such ordered pairs (g, p) are called **spins** for some reason.) Since every element differing from 1 has 2 poles, there are a total of $2(N-1)$ poles. But there's another way to count S: consider a fixed pole p. Then there are $|\text{Stab}(p)| - 1 = r_i - 1$ such (g, p) that work, since we need g to be a nonidentity element in the stabilizer of p. Summing this over all p (which can be done by summing over all orbits)

$$2N - 2 = |S| = \sum_p (r_i - 1) = \sum_i (n_i(r_i - 1)) = \sum_i (N - n_i)$$

where the last equality comes from $n_i r_i = N$. Now dividing by N,

$$2 - \frac{2}{N} = \sum_i \left(1 - \frac{n_i}{N}\right) = \sum_i \left(1 - \frac{1}{r_i}\right).$$

Notice that $2 - \frac{2}{N} < 2$, while $1 - \frac{1}{r_i} \geq \frac{1}{2}$ if $r_i > 1$ (and we know that $r_i \neq 1$ because any pole p must have a **nontrivial** rotation that stabilizes it by definition). So this means that $i \leq 3$. Time for casework!

- If $i = 1$, the left hand side is ≥ 1, while the right hand side is < 1. This can't happen.
- If $i = 2$, then $2 - \frac{2}{N} = 1 - \frac{1}{r_1} + 1 - \frac{1}{r_2} \implies \frac{1}{r_1} + \frac{1}{r_2} = \frac{2}{N}$. But r_i is the order of a stabilizer, so $r_i | N$, and we must therefore have $r_1 = r_2 = N$. So there are two poles, fixed by all $g \in G$ (since the orbits are trivial). This means they are opposite poles on a specific, and this leads to $G = C_n$.
- Now, if $i = 3$. we have

  $$2 - \frac{2}{N} = 1 - \frac{1}{r_1} + 1 - \frac{1}{r_2} + 1 - \frac{1}{r_3} \implies \frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3} = 1 + \frac{2}{N}$$

  This is actually really hard to satisfy; if none of r_i are 2, then the left hand side is less than 1. So there are only a few possibilities: the edge cases are

  $$\frac{1}{3} + \frac{1}{3} + \frac{1}{3} = \frac{1}{2} + \frac{1}{4} + \frac{1}{4} = \frac{1}{2} + \frac{1}{3} + \frac{1}{6} = 1.$$

  WLOG let $r_1 \leq r_2 \leq r_3$; we must have $r_1 = 2$ and $r_2 = 2$ or 3. So we have $(2, 2, r), (2, 3, 3), (2, 3, 4), (2, 3, 5)$, which correspond to $N = 2r, 12, 24, 60$ and $n_i = (r, r, 2), (6, 4, 4), (12, 8, 6), (30, 20, 12)$. These correspond to D_r, T, O, I respectively. In fact, for the tetrahedron, octahedron, and icosahedron, these actually correspond to edges, vertices, and faces!

□

Example 162
Professor Artin brought a power line insulator. Its group of symmetries is D_4, but the element of order 4 is not a pure rotation; it is orientation-reversing.

18 October 22, 2018

We'll now consider **operations of a group on itself**: what are some simple examples? First of all, the trivial operation sends $(g, x) \in G \times G \to gx$. This is not very interesting.

On the other hand, **conjugation** sends (g, x) to $g * x = gxg^{-1}$. It's good to check that this is indeed a valid group operation: we want $1 * x = x$ and $g_1 * (g_2 * x) = (g_1 g_2) * x$ for all $x \in G$. The first is easy, and

$$g_1 * (g_2 * x) = g_1 * (g_2 x g_2^{-1}) = g_1 g_2 x g_2^{-1} g_1^{-1} = (g_1 g_2) x (g_1 g_2)^{-1}.$$

So now that conjugation is a group operation, we can talk about orbits and stabilizers of this action.

Definition 163
The orbit of $x \in G$ is $\{x' \mid x' = gxg^{-1}, g \in G\}$, which is the set of elements conjugate to x. This is called the **conjugacy class** of x, and it's often denoted $C(x)$.

Definition 164
The stabilizer of x is all g such that $gxg^{-1} = x$; that is, $gx = xg$. Thus, this is the set of g that commute with x, which is the **centralizer** $Z(x)$.

Thus, the counting formula (that is, the orbit-stabilizer theorem) tells us that

$$|G| = |C(x)||Z(x)|$$

for any $x \in G$. We also know that the orbits partition our group G:

Theorem 165 (Class Equation)
Let $C_1, C_2, \cdots, C_k$ be the conjugacy classes of a group G. Then $|G| = |C_1| + \cdots + |C_k|$.

For example, the only element conjugate to the identity is itself, and thus one of the $|C_i|$s must be 1.

Let's discuss a few more important things about each conjugacy class $C(x)$:

- $x \in C(x)$.
- $|C(x)|$ divides the order of the group $|G|$. Notice that this means $|G| = |C_1| + \cdots + |C_k|$ is a sum of divisors of $|G|$, one of which is 1.

Similarly, we can talk about the centralizer:

- $x \in Z(x)$.

- $|Z(x)|$ divides the order of the group $|G|$.

When is the order of a conjugacy class 1? This means that everything in the group commutes with it, and the order of the centralizer is $|G|$.

Definition 166
The **center** of a group G is the set $Z = \{x \in G : gx = xg \ \forall g \in G\}$.

Notice that in an abelian group, the class equation is useless, since the conjugation operation does nothing. In addition, the elements of Z are each in their own conjugacy class, so we get additional terms of 1 in the class equation. Finally, $Z \subset Z(x)$ for any x.

It is important here to state a quick fact:

Lemma 167
Conjugate elements in a group have the same order.

Proof. If $x^r = 1$, then $(gxg^{-1})^r = gx^r g^{-1} = gg^{-1} = 1$. □

Example 168
Consider D_5, the symmetries of a pentagon. D_5 has order 10; what is its class equation?

Let $x = \rho_{2\pi/5}$ and $y = r$ (the standard reflection). Then the group is described by

$$G = \{x^a y^b \mid x^5 = 1, y^2 = 1, xy = yx^{-1}\}.$$

Notice that x, x^2, x^3, x^4 all have order 5. The orders of the other nonidentity elements have order 2, since they are just reflections. By the above lemma, the conjugacy class of x must be a subset of $\{x, x^2, x^3, x^4\}$. It doesn't have to be the whole set - in fact, it can't be, because 4 is not a divisor of 10.

We know that $C(1) = \{1\}$ (the identity is conjugate to only itself). Notice that $yx = x^{-1}y \implies yxy^{-1} = x^4$. Thus, the conjugacy class of x contains x^4, and therefore $C(x) = \{x, x^4\}$ (we can't add anything else or the order of $C(x)$ wouldn't divide 10). Similarly, $C(x^2) = \{x^2, x^3\}$.

Now let's look at the other elements. In general, **it's easier to find the centralizer of an element than the conjugacy class**, because it is a subgroup (and that gives it more structure). So let's find the centralizer of y. $Z(y)$ contains y, and since it's a group, $|\langle y \rangle| = 2$ divides $|Z(y)|$, which divides $|G| = 10$. But this means $|Z(y)|$ must be 2 or 10, and it's not 10, since y is not in the center of G (for instance, $yx \neq xy$). Thus $|Z(y)| = 2$ and $|C(y)| = 5$.

This gives our class equation

$$10 = 1 + 2 + 2 + 5.$$

Since there's only one 1, this proves that the only element of the center in D_5 is the identity.

This kind of class equation calculation can be insightful in different ways:

Definition 169
A ***p*-group** is a group whose order is p^r for some $r \geq 1$ and prime p.

Proposition 170

The center of a p-group is not just the trivial group $\{1\}$.

Proof. The class equation of G is of the form

$$p^r = 1 + \sum C_i = 1 + \sum p^{c_i},$$

where the c_is are nonnegative integers. Remember that terms of 1 in the class equation correspond to elements of the center. If there is nothing else in the center besides the identity element, then all terms in the sum are multiples of p. This is a contradiction since $0 \neq 1 \bmod p$. Thus, we must have some additional p^0 terms. □

Example 171

$D_4 = \{x^a y^b \mid x^4 = y^2 = 1, yx = x^{-1}y\}$ has 8 elements, and 8 is a prime power. Thus, its center is nontrivial. It turns out the center is $\{1, x^2\}$, since $yx^2 = x^{-2}y = x^2y$. We can check that the class equation turns out to be

$$8 = 1 + 1 + 2 + 2 + 2.$$

Example 172

Let G be the set of 3×3 matrices of the form

$$A = \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$$

where a, b, c are integers mod p.

The order of the group is $|G| = p^3$. What is the center of G – that is, when is it true that

$$\begin{bmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{bmatrix} ?$$

Doing out the matrix multiplication, this happens when

$$\begin{cases} x + a = a + x \\ b + xc + y = y + az + b \\ z + c = c + z. \end{cases}$$

Most of this is trivially satisfied: the only problem is that we need to have $xc = az$. Well, this can only hold for all x, y, z if $a = c = 0$. **So the center has order** p: it's all matrices of the form $\begin{bmatrix} 1 & 0 & b \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Theorem 173

If $|G| = p^2$, then G is abelian.

Proof. We know that Z is not trivial by Proposition 170, and the order of the center $|Z|$ divides $|G| = p^2$, meaning it is either p or p^2.

If it is the latter, we are done: everything commutes with everything. Assume the former for contradiction. Then there exists some element $x \in G$ that is not in the center. Let H be the subgroup generated by $|Z|$ and x. Since Z is a subgroup of H, $|Z| = p$ divides $|H|$, which divides $|G| = p^2$. In addition, $|H|$ contains x, which is not in $|Z|$. Thus $|H| = p^2$, and everything in the group G can be generated by Z and x.

But $x, Z \subset Z(x)$, so $Z(x) = G$ is the whole group. Thus, x commutes with all $g \in G$. But this means x is in the center, contradiction. □

Corollary 174

If $|G| = p^2$, $|G|$ is either cyclic of order p^2 or the direct product of two cyclic groups. In other words, $G = C_{p^2}$ or $C_p \times C_p$.

Proof. All elements of G that are not the identity have order p or p^2. If some element x has order p^2, the subgroup generated by x gives a cyclic group. Otherwise, all elements have order p. Take two of them, x, y, such that y is not a power of x. Then these generate the whole group, and we can check that this is isomorphic to $C_p \times C_p$. □

19 October 24, 2018

Today we're going to calculate some more class equations. First, recall the definition: for any $x \in G$, let $C(x)$ be the **conjugacy class** of x, which is the set of all elements x' that can be written as gxg^{-1} for some $g \in G$. We also let $Z(x)$ be the **centralizer** of x: the set of all elements g such that $gx = xg$. Recall that $|G| = |C(x)||Z(x)|$ and that $x \in C(x), Z(x)$. Also, the center of the group Z is a subset of $Z(x)$.

Then the class equation says that $|G| = \sum C_i$, where C_i are the distinct conjugacy classes. Since 1 is by itself in a conjugacy class, $C_1 = 1$ is one of the terms in the sum. Also, elements of the center Z contribute 1s in the class equation.

Example 175

Let $G = SL_2(\mathbb{F}_3)$, the set of 2×2 matrices with entries $\in \{-1, 0, 1\}$ and determinant 1.

First of all, what's the order of $SL_2(\mathbb{F}_3)$? We did this for homework; it's $(p-1)p(p+1) = 24$ in this case. So we'll write 24 as a sum of conjugacy classes.

First of all, $\pm I$ are in the center, so two of the terms are 1s. Only 22 elements to go!

Fact 176

The characteristic polynomial of a matrix x is the same as the characteristic polynomial of gxg^{-1}. (This is because gxg^{-1} is the same operator in a different basis, so the eigenvalues should stay the same.)

Thus, all matrices in a conjugacy class will have the same characteristic polynomial. What are the possible characteristic polynomials? They will be of the form $p(t) = t^2 - (\text{tr } x)t + (\det x) = t^2 - trx + 1$. Then tr x, the trace, can be $-1, 0$, or 1, and this limits us to three posssible characteristic polynomials: $t^2 + 1, t^2 + t + 1 = (t-1)^2$, and $t^2 - t + 1 = (t+1)^2$.

First of all, we find a matrix A with characteristic polynomial t^2+1. One example that works is

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

To find the order of the conjugacy class of A, we wish to find its centralizer: that is, the set of matrices P such that $PA = AP$. Let $P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$; then we're asking for

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

which is true if $a = d, b = -c$. So matrices in the centralizer are of the form $P = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$, and to have determinant 1, we need $a^2+b^2 = 1 \pmod 3$. So one of a^2 and b^2 is 1 and the other is 0: the possibilities are $\{0, \pm 1\}$ and $\{\pm 1, 0\}$. This means there are 4 elements in $Z(A)$, so $|C(A)| = 6$.

Next, we look at $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = I + e_{12}$. (This has a different trace, so it is in a different conjugacy class.) If P is in the centralizer of B, which means $PB = BP$, we need $e_{12}B = Be_{12}$, which means $c = 0, a = d$. For this to have determinant 1, $a = \pm 1$ and then there are 3 choices for b, so $|Z(B)| = 6$ and $|C(B)| = 4$.

Similarly, we can take $B' = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix} = -I + e_{12}$; a similar calculation yields another $|Z(B')| = 6$ and $|C(B')| = 4$.

Now we try $B^T = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. This has the same characteristic polynomial as B; are B and B^T in the same conjugacy class? In other words, we have to solve $PBP^{-1} = B^T$. Since $B = I + e_{12}$, we can equivalently ask whether $Pe_{12} = e_{21}P$. This happens if $a = 0, b = c$. But this leads to a matrix of the form $\begin{bmatrix} 0 & b \\ b & d \end{bmatrix}$, which can never have determinant 1. So B and B^T are **not conjugates**, and we do have a different conjugacy class. An identical calculation also gives $|C(B^T)| = 4$, and similarly $|C(B'^T)| = 4$. Finally, we've arrived at our final class equation:

$$24 = 1 + 1 + 6 + 4 + 4 + 4 + 4.$$

Example 177

Let $G = S_5$. This group has order $5! = 120$.

To understand conjugation in the symmetric, group let's think about different ways to label indices. For example, let's say we have a permutation $p = (123)(45)$ that we want to write in terms of the indices a, b, c, d, e. Then we can represent this as having a function f that sends $1 \to a, 2 \to b$ and so on. Then our permutation is $\tilde{p} = fpf^{-1}$. This is a weird way to think about conjugation: f is just a way to think between languages; it's sort of like a translation:

$$\tilde{p}(b) = [fpf^{-1}](b) = [fp](2) = f(3) = c.$$

But now, what if f doesn't send $1, 2, 3, 4, 5$ to a, b, c, d, e – **it instead sends us to another permutation of the numbers**? For example, what if f sends them to $5, 3, 2, 1, 4$ respectively? Then f is a permutation as well, and $\tilde{p} = fpf^{-1}$.

Corollary 178

The conjugacy class of a permutation p is all permutations that have the same cycle type.

For example, p in the example above is conjugate with all permutations with a 3-cycle and a 2-cycle. But we have to be a bit careful. The permutation (123)(45) is conjugate with both (532)(14) and (325)(14), which correspond to different fs. However, those are the same permutation represented in different ways, so we don't have a one-to-one correspondence.

Example 179

Let's first find the class equation for S_4.

The different types of cycles are the identity, 2-cycles, pairs of 2-cycles, 3-cycles, and 4-cycles. Well, there's 1 identity, $\binom{4}{2} = 6$ 2-cycles, 3 pairs of 2-cycles, $\binom{4}{3} \cdot 2 = 8$ 3-cycles, and $3! = 6$ 4-cycles, and we easily have our class equation

$$24 = 1 + 6 + 3 + 8 + 6.$$

(The order in which we write the class equation doesn't matter.)

Example 180

Now we do S_5 – by the way, S_6 is for homework.

We have 1-cycles, 2-cycles, pairs of 2-cycles, 3-cycles, a 3 and 2-cycle, a 4-cycle, and a 5-cycle. Call these C_1 through C_7. $|C_1| = 1$; this is just the identity. There are $|C_2| = \binom{5}{2} = 10$ 2-cycles. We can get pairs of 2-cycles in $|C_3| = \frac{1}{2}\binom{5}{2}\binom{3}{2} = 15$ ways. We can just keep going at this point: $|C_4| = \binom{5}{3} \cdot 2 = 20$ (pick three elements and then pick the order of the cycle), $|C_5| = 20$ as well, since we just put the remaining two indices in a 2-cycle, $|C_6| = \binom{5}{4} \cdot 3! = 30$, and $|C_7| = \frac{5!}{5} = 24$. This yields

$$120 = 1 + 10 + 15 + 20 + 20 + 30 + 24.$$

Example 181

Finally, let's find the class equation for A_4.

Notice that each cycle type is either even or odd. In S_4, the identity, 3-cycles and pairs of transpositions are even permutations, so $12 = 1 + 3 + 8$. However, just because permutations are conjugates in the symmetric group doesn't mean they are conjugates in the alternating group! In fact, 8 doesn't even divide 12. It turns out that the class equation for A_4 is

$$12 = 1 + 3 + 4 + 4.$$

How would we find something like this in general? We know that $S_n \supset A_n$. If p is an even permutation $\in A_n$, then it has conjugacy class in both groups: let it be C_s in S_n and C_A in A_n. But if we can conjugate by something in A_n, we can do the same in S_n, so $C_S \supset C_A$. In other words, $|C_A| \leq |C_S|$.

Similarly, if we have centralizers Z_A and Z_S respectively, if an element commutes with p in A_n, it'll commute in S_n. So $Z_A \subset Z_S$, and since centralizers are groups, we actually have that $|Z_A|$ divides $|Z_S|$. This is actually powerful: we know that $|C_S||Z_S| = |S_n| = n!$ and $|C_A||Z_A| = |A_n| = \frac{n!}{2}$. We only have a factor of 2 to work with, and C_A, Z_A are subsets of C_S, Z_S respectively

Fact 182

For any permutation in S_n, there are only two possibilities: either $|C_A| = |C_S|$ or $\frac{|C_S|}{2}$. These correspond to a centralizer of $|Z_A| = \frac{|Z_S|}{2}$ and $|Z_S|$, respectively.

The case where $|C_A| = \frac{1}{2}|C_S|$ occurs when all $q \in S_n$ such that $qp = pq$ is even. On the other hand, $|C_A| = |C_S|$ occurs if there is an odd permutation in Z_s for p.

So now, we return to S_5. The alternating group contains the identity, pairs of 2-cycles, 3-cycles, and 5-cycles. There are 15 pairs of 2-cycles, so we must have $|C_A| = 15$. (Alternatively, (12), an odd permutation, commutes with (12)(34).) For the 3-cycles, (45) commutes with (123), so this stays at 20. Finally, there are 24 5-cycles, and 24 does not divide 60, so this must split in half. This gives a class equation for A_5:

$$60 = 1 + 15 + 20 + 12 + 12.$$

20 October 26, 2018

Today we will be talking about the **icosahedral group** I, the set of rotational symmetries of an icosahedron or dodecahedron (which are dual).

Fact 183

The dodecahedron has 12 faces, 20 vertices, and 30 edges. All of the faces are regular pentagons.

We're going to return to the idea of a **spin**: given a rotation by θ around some axis, we can draw a unit vector p along the axis. That ordered pair (g, p) is called a spin. By the way, recall that

$$\rho_{\theta,p} = \rho_{-\theta,-p}.$$

Basically, a rotation that rotates a face by $+\theta$ also rotates the opposite face by $-\theta$. Call faces f, vertices v, edges e.

Let's try to count the elements of our group. As always, we have 1 identity, and every other symmetry can be described in one of several ways.

Each face has 4 nontrivial face rotations, and there are 12 faces, but each one is counted twice, so we have $\frac{4\cdot 12}{2} = 24$ face rotations. Each vertex is connected to 3 faces, so here we have $\frac{2\cdot 20}{2} = 20$ vertex rotations. Finally, we have edge rotations: each edge can be rotated by π, so we have $\frac{1\cdot 30}{2} = 15$ edge rotations. The total order is $1 + 24 + 20 + 15 = 60$.

We can also look at orbits and stabilizers to count the order of I. All the faces form an orbit, since we can rotate around vertices. There are 12 faces and 5 elements that fix each face. Similarly, there are 20 vertices and 3 elements that fix each vertex, and 30 edges and 2 elements that fix each edge, so

$$60 = 12 \cdot 5 = 20 \cdot 3 = 30 \cdot 2.$$

But now compute the class equation of I, and to do that, let's try to visualize conjugation. Recall that if we want to rotate around q by an angle θ,

$$\rho_{\theta,q} = g\rho_{\theta,p}g^{-1}$$

where g is a rotation that sends p to q. In other words, **rotations of the same angle** are always in the same conjugacy class!

So rotations by $\frac{2\pi}{5}$ are conjugate to each other, and also to those rotations of $-\frac{2\pi}{5}$. In total, this gives 12 rotations by $\frac{2\pi}{5}$ (avoiding double counting). There's another conjugacy class of 12 rotations of $\frac{4\pi}{5}$, and a conjugacy class of 20 rotations about vertices. Finally, there are 15 rotations by π around edges, for a class equation of

$$60 = 1 + 12 + 12 + 20 + 15.$$

What's the use of a class equation? We can use them to analyze normal subgroups! Recall that a subgroup N of a group G is **normal** if for all $x \in N, g \in G$, we have $gxg^{-1} \in N$.

Then N is a union of conjugacy classes, including the class of 1 (the identity). Furthermore, $|N|$ divides $|G|$. This is pretty powerful: let's say we want a normal subgroup of I, the icoshedral group. We need to add some of $12, 12, 20, 15$ to 1 to get a divisor of 60: the only way to do this is to get the trivial subgroup or the whole group!

Corollary 184

The icosahedral group I has no proper normal subgroups except for the trivial subgroup (only containing the identity).

This means I is a **simple group**, meaning there are no nontrivial proper normal subgroups. For example, groups of prime order are simple, but those are not interesting.

Recall the class equation for S_4:

$$24 = 1 + 6 + 3 + 8 + 6.$$

The normal subgroups can potentially be formed from $1, 1+3, 1+3+8, 1+6+3+8+6$. These correspond to the identity, pairs of transpositions, alternating group, and the whole group. In this case, these all happen to be normal subgroups.

Theorem 185

The icosahedral group I is isomorphic to A_5.

Proof. To prove this, we find a way to have I operate on 5 things. It turns out there are 5 cubes that are inscribed in the dodecahedron! Then this gives us a homomorphism $I \xrightarrow{\phi} S_5$, because each symmetry sends the 5 cubes to a different permutation. The kernel is a normal subgroup of I, and there are no proper normal subgroups, so it must be trivial or the whole group. If the kernel was the whole group, then everything does nothing to the five cubes, which is not true. So ϕ is injective, and I maps isomorphically to some subgroup H of S_5.

Now consider the sign map that sends $S_5 \to \{\pm 1\}$. Where does H go? The image is either $\{1\}$ or $\{\pm 1\}$. If $H \xrightarrow{S} \{\pm 1\}$, then the kernel of S has order $\frac{1}{2}|H| = 30$. But this means there is a corresponding normal subgroup of order 30 in I (by the correspondence theorem), which is a contradiction. Thus, this sign map must only map H to $\{1\}$, so H is in the kernel of the sign map, which is A_5. Indeed, both have order 60, and we've shown our isomorphism. □

Corollary 186

The alternating group A_5 is simple.

Theorem 187

A_n is simple for all $n \geq 5$.

Notice that A_4 is not simple: it has the normal subgroup of order 4.

Outline of a proof. Let N be a nontrivial normal subgroup of A_n: we must show that $N = A_n$. If N is normal in G, then it is closed under multiplication, inversion, and conjugation. In particular, the commutator $gxg^{-1}x^{-1} \in N$ is in N. The idea is that if the group is noncommutative enough, this gives us a lot of elements in N.

First, show that the 3-cycles generate A_n. Then, we show that for $n \geq 5$, the 3-cycles form one conjugacy class. Now, to show $N = A_n$, we can take $x \in N$ and $g \in A_n$ and find a 3-cycle $gxg^{-1}x^{-1} \in N$. Then if one 3-cycle is in N, then all of them are, since the 3-cycles are in one conjugacy class. Then those generate the whole group. □

Let's do an example: suppose that N contains the element $x = (12345)\cdots$. Letting $g = (123)$,

$$gxg^{-1}x^{-1} = (123)[(12345)\cdots][(132)][\cdots(15432)] = (124)(3)(5)(\cdots)$$

and the $\cdots$ cancel and we can ignore them; we've found a 3-cycle in N.

21 October 29, 2018

Today, we're going to talk about the three Sylow theorems, which are probably the first important general structure theorems (except for maybe some results that Galois proved). We'll state them and do some applications today, and we'll do the proofs and some other things on Wednesday.

Theorem 188 (Sylow Theorem 1)
Let G be a finite group of order n, and let p be a prime dividing n such that $n = p^e m$ and $p \nmid m$. (In other words, p^e is the largest power of p dividing n.) Then there always exists a **Sylow p-subgroup**, which is a subgroup of G that has order p^e.

As a corollary, there is always an element of order p, since we can take a nontrivial element of a Sylow p-group and look at the cyclic group generated by it.

Let's see what we can do with this.

Example 189
Let $|G| = 6$. What can we say about the group from the first Sylow theorem?

Solution. By above, there exists an element x of order 3, and there also exists an element y of order 2. So we have the six elements

$$G = \{1, x, x^2, y, xy, x^2y\}$$

We can see that xy is not equal to $1, x, x^2, y$ by cancellation, and similarly that x^2y is also distinct from the others. Thus all six of the elements in this set are distinct, so this is actually our group G.

Let K be the Sylow 3-subgroup generated by x. It has index 2 in G, since $G = K \cup yK$.

Fact 190
Any subgroup of order 2 is normal, so K is normal.

(This is because $G = K \cup yK = K \cup Ky$. Therefore $yK = Ky$, and all left and right cosets are equal.) As a result, yxy^{-1} must be an element of K. It's not 1, or we would get that $x = 1$, so either $yxy^{-1} = x$ or x^2. These two

possibilities must then correspond to the two non-isomorphic groups of order 6 that we know: the first case gives us the group law for C_6, and the second case gives us the group law for $S_3 = D_3$. Every group of order 6 is therefore isomorphic to either C_6 or S_3. □

Notice that this isn't the only way to represent C_6. We actually did it as a product group $C_3 \times C_2$ here:

Definition 191

Given two groups H, K, the **product group** $H \times K$ has group elements that are the product set (h, k), $h \in H, k \in K$, with multiplication componentwise:

$$(h, k) \cdot (h', k') = (hh', kk')$$

Lemma 192 (Chinese Remainder Theorem)

For any $n = ab$, where $\gcd(a, b) = 1$, C_n is isomorphic to $C_a \times C_b$.

Proof. Since a and b are relatively prime, there exist $r, s \in \mathbb{Z}$ such that $ra + sb = 1$. Let's say $x \in C_n$ generates the group: that is, it has order n. Let $u = x^a$ and let $v = x^b$; then u has order b and v has order a.

Let y generate C_a and z generate C_b. Now consider the map

$$C_a \times C_b \to C_n : (y, 1) \to v, (1, z) \to u$$

This sends $(y^i, z^j) \to v^i u^j = x^{ai+bj}$, and picking the relevant i and j to make $ai + bj = 1$, we can get x^1 out of this map, so we can generate all of C_n. Therefore the map is surjective, and since C_n and $C_a \times C_b$ have the same finite order, the map must be injective as well (and we have an isomorphism). □

Theorem 193 (Sylow Theorem 2)

All Sylow p-subgroups are **conjugate**: if H, H' are two Sylow p-subgroups, then $h = gHg^{-1}$ for some g.

Theorem 194 (Sylow Theorem 3)

The number of Sylow p-subgroups divides $m = \frac{n}{p^e}$ and is congruent to 1 mod p.

This last result is really useful for looking at groups of some fixed order.

Fact 195

If there is only one Sylow p-subgroup H, then it is normal.

This is because it is equal to all of its conjugates, so the left and right cosets are indeed equal.

Example 196

What can we say about the groups of order $10 = 2 \cdot 5$?

Solution. The number of Sylow 5-groups divides 2 and is 1 mod 5, so it is 1. Thus K, the subgroup of order 5, is normal, and it is generated by some x with order 5.

By the way, we know that the number of Sylow 5-groups divides 5 and is 1 mod 2, so it is either 1 or 5, but we won't use this. Either way, there exists an element y of order 2 that generates one of the Sylow 2-groups. So we can write the elements out as

$$G = \{1, x, x^2, x^3, x^4, y, xy, x^2y, x^3y, x^4y\}.$$

Since $K = \langle x \rangle$ is normal, $yxy^{-1} = x^r \implies yx = x^r y$ for some r (since we must stay within the normal subgroup K after conjugating by y). We also know that $x^5 = y^2 = 1$.

Does such a group exist for all r? Technically yes, but if we put bad relations on a group, it might collapse. For example, let's try $yx = x^2y$. After some manipulation,

$$x = y^2x = yx^2y = x^4y^2 = x^4 \implies x^3 = 1 = x^5$$

so $x = 1$, which means we don't have a group of order 10 at all. Moral of the story: relations are always fine, but we might collapse the group. Well, let's try this calculation with other r:

$$x = y^2x = yx^ry = x^{r^2}y^2 = x^{r^2}$$

so if we want x to be nontrivial, $r^2 \equiv 1 \bmod 5$, which implies $r = 1, 4$. Thus, there are two isomorphic classes of groups of order 10. And we know what they are: $yx = xy$ corresponds to $C_{10} = C_5 \times C_2$ and $yx = x^4y$ corresponds to D_5. □

Example 197

What are the groups of order 15?

Proof. The number of Sylow 5-groups divides 3 and is congruent to 1 mod 5, so it is again 1. The number of Sylow 5-groups divides 3 and is congruent to 1 mod 3, so this is also 1. So both subgroups are normal; call them K and H respectively.

What's the intersection of H and K? $H \cap K$ is a subgroup of both H and K, so the size of the intersection must divide both 3 and 5, so it must be 1.

We claim that G is isomorphic to $C_3 \times C_5 = C_{15}$. In other words, there is only one isomorphism class of groups of order 15. To justify this, we need the following result:

Lemma 198

Let H and K be subgroups of G.

- If K is normal, then $HK = KH$ is a subgroup of G.
- If both are normal and $H \cap K = \{1\}$, the elements of H commute with the elements of K, so $hk = kh$.
- Finally, if both are normal, $H \cap K = \{1\}$, and $HK = G$, then G is isomorphic to $H \times K$.

Proof. Let's do these one by one. If K is normal, then $hk = (hkh^{-1})h \in KH$. So $HK \subset KH$, and we can analogously show that $KH \subset HK$, so $HK = KH$. This is a subgroup of G because (checking closure) $HKHK = H(KH)K = H(HK)K = HHKK = HK$.

For the second one, we show that $hkh^{-1}k^{-1} = 1$. It's equal to $(hkh^{-1})k \in K$, and it's equal to $h(kh^{-1}k^{-1}) \in H$. Therefore, the element $hkh^{-1}k^{-1}$ in the intersection $H \cap K$, so this commutator must be the identity, and therefore $hk = kh$.

Finally, we know that H and K still commute in the third case, and the map $H \times K \to G$ which sends $h, k \to hk$ is a homomorphism. The kernel of this map is 1, since $hk = 1 \implies h = k = 1$ (we can't have them be non-identity inverses if H and K have only trivial intersection). Therefore, the image is $HK = G$, so this is an isomorphism and G is isomorphic to $H \times K$. □

And using this with our groups C_3 and C_5, we have now showed that $G = C_5 \times C_3 = C_{15}$ is the only group of order 15. □

22 October 31, 2018

We're going to prove the Sylow theorems today. Here they are together:

Theorem 199 (All three Sylow Theorems)
Let G be a finite group with order $n = p^e m$ for a prime p (where $p \nmid m$).

- There exists a subgroup of order p^e, called a Sylow p-subgroup.
- All Sylow p-subgroups are conjugates.
- The number of Sylow p-subgroups is 1 mod p, and it always divides m.

Proof of the first Sylow theorem. Let S be the set of all subsets of G of order p^e. The size of S is

$$|S| = \binom{n}{p^e} = \frac{n!}{p^e!(n-p^e)!} = \frac{n(n-1)\cdots(n-p^e+1)}{1 \cdot 2 \cdots p^e}.$$

This is **not divisible by** p: we'll accept this for now, but we can verify by checking that the powers of p match up in the numerator and denominator.

Now G operates on S by sending any $U \in S$ to $g * U = gU$. Since $|S|$ is not divisible by p and is partitioned into orbits, **some orbit has order not divisible by** p: say that it is $\text{Orbit}(U)$. We claim that this orbit covers G evenly. More specifically, say that the element 1 is in all of $U_1, \cdots, U_k$, where the $U_i \in S$ are all in the orbit of U. Then $g \in gU_1, \cdots, gU_k$, which are all also in the orbits, so 1 and g are in the same number of orbits of U.

Thus, every element is in the same number of orbits: let's say the group G is covered k times. So

$$|U||\text{Orbit}(U)| = k|G| \implies p^e|\text{Orbit}(U)| = kp^e m,$$

so the order of the orbit of U is km. But the order of the orbit must divide $|G| = p^e m$, meaning k must be a power of p. On the other hand, we chose the orbit's order to **not be divisible by** p, so we must have $k = 1$. So the orbit actually only covers G once, and the orbit of U has order m. Then the orbit-stabilizer counting formula tells us the stabilizer of U has order p^e.

But the stabilizer is a group! So we have indeed found a Sylow p-subgroup, as desired. □

We'll take a break now to study some more structure of Sylow p-subgroups:

Example 200
We found last time that groups of order 10 are either C_{10} or D_5, and groups of order 15 must be isomorphic to C_{15}. What about groups of order 30?

Solution. The number of Sylow 5-subgroups divides 6 and is 1 mod 5, so it is either 1 or 6. The number of Sylow 3-subgroups divides 10 and is 1 mod 3, so it must be 1 or 10. If there are 6 Sylow 5-subgroups, they are all cyclic and don't intersect, each contributes 4 elements of order 5, so we have 24 elements of order 5. On the other hand, if there are 10 Sylow 3-subgroups, we have 20 elements of order 3. So we either have a normal Sylow 5-subgroup H or a normal Sylow 3-subgroup K.

Lemma 201

If H and K are subgroups, and at least one is normal, then HK is a group.

(This is just verification of closure and the other axioms.) H has order 5 and K has order 3, and they (again) must have no intersection, so HK has order 15. Thus, G contains a subgroup C_{15}; let x be one of its elements with order 15.

We know there exists a Sylow 2-subgroup, so there is an element of order 2; let it be y. Then $\langle x \rangle$ has order 15 and index 2 in G, and index 2 subgroups are normal. So yxy^{-1} is still in $\langle x \rangle$, which means $yx = x^r y$ for some $1 \leq r < 15$, and also

$$x = y^2 1 y^{-2} = y x^r y^{-1} = x^{r^2}$$

so $r^2 \equiv 1 \bmod 15$, which gives $r = \pm 1, \pm 4$.

Each of these corresponds to a different group $\{x^{15} = y^2 = 1, yx = x^r y\}$, and it is unique because we can write out a multiplication table just based on these properties of x and y. We just need to check that these actually give a group of order 30 without collapsing, and it turns out there are indeed 4 groups of order 30: C_{30}, D_{15}, $C_3 \times D_5$, and $C_5 \times D_3$. The last two groups have centers of C_3 and C_5 respectively, so they are distinct. So we've found all groups of order 30! □

Proof of the second Sylow theorem. Let H be a Sylow p-subgroup, and let C be the set of cosets of H. The group G operates on the cosets via $g * aH = gaH$. There is only one orbit – the orbit of $1 \cdot H$ – and the stabilizer of $1 \cdot H$ has order $|H| = p^e$, so by the orbit-stabilizer theorem, the orbit of $1H$ has order m, which is not divisible by p.

Now let H' be another Sylow p-subgroup; we wish to show it is conjugate to H. Restrict the operation of G on C to H', so we can only multiply cosets by elements of H'.

Decompose C into H'-orbits. p does not divide m, the order of Orbit$(1 \cdot H)$.

Lemma 202 (Fixed Point Theorem)

Let H be a p-group that operates on a set S. If p does not divide $|S|$, there exists a fixed point s such that $hs = s$ for all $h \in H$.

Proof. Partition S into orbits. Since $|H|$ is a power of p and the orbits' orders divide $|H|$, $|S|$ is a sum of powers of p. But $|S|$ does not divide p, so we must have a 1 term in there somewhere: thus there is an element with orbit of 1, making it a fixed point. □

Thus, the operation of H' on cosets of H has a fixed point; let's say gH is fixed by all elements of H'. So H' is contained in the stabilizer of gH, which is gHg^{-1}.

This means that $H' \subset gHg^{-1}$, but both have order p^e. Thus $H' = gHg^{-1}$, and the two subgroups are indeed conjugate. □

Remark 203. *We didn't get to the proof of the third Sylow theorem, but the main ideas are as follows: since all p-subgroups are conjugate, we can write them as gHg^{-1} for g in some subgroup K of G. Now H is a subgroup of K, so $[G : H] = m$ is a multiple of $[G : K]$.*

To show that the number of groups is 1 *mod* p*, we decompose into orbits under conjugation by* H*. Then* H *is its own orbit, but all other orbits have order dividing* $|H| = p^e$*, so the total sum is* 1 *mod* p *as desired.*

23 November 2, 2018

Recall that S_3, the symmetric group on 3 elements, can be represented in the form

$$\{\langle x, y\rangle \mid x^3 = 1, y^2 = 1, yxyx = 1\}$$

Things like $x^3, y^2, yxyx$ are called **relations.** Today's class is about how we can use the **Todd-Coxeter algorithm** to see how the group operates on cosets of a subgroup!

First, we choose a subgroup H of G and write it by using **words** in x, y that are generators for H. For example, we can take $H = \langle y\rangle$. This has order 2, so there will be 3 cosets.

Theorem 204 (Todd-Coxeter Algorithm)

Here are some rules for operating on cosets of H: this generates a unique correct table.

- The relations operate trivially, since they're equal to 1.
- The operation of a generator (in this case, x, or y) is a permutation of the cosets.
- The generators of H fix the cosets of H.

We're going to work with right cosets so that we don't have to read backwards.

We can start off by writing the relations out like this:

x	x	x		y	y		y	x	y	x

Cosets aren't uniquely represented by elements, so we'll denote them by indices instead. In this example, we'll let "1" denote the coset $H \cdot 1$.

We don't know what x does to H, but we know x^3 sends it to 1, and so do the other relations:

x	x	x		y	y		y	x	y	x
1		1	1		1	1				1

But we also know that y generates H, so it sends 1 to 1 as well:

x	x	x		y	y		y	x	y	x
1		1	1	1	1	1	1			1

We don't know what x does to 1, so we'll say 1 goes to 2, and then we'll say 2 goes to 3, so we know 3 goes to 1:

x	x	x		y	y		y	x	y	x
1	2	3	1	1	1	1	1	1	2	3	1

Well, now let's try to figure out what happens to the cosets represented by 2 and 3: we can already fill out the remainder of the table just given what we know.

	x	x	x		y	y		y	x	y	x
1	2	3	1	1	1	1	1	1	2	3	1
2	3	1	2	2	3	2	2	3	1	1	2
3	1	2	3	3	2	3	3	2	3	2	3

And now we notice $\langle y \rangle$ has index 3 (because there are 3 cosets which are only sent to each other under x and y). Now the order of x is 3 (since it operates nontrivially on the cosets) and the order of y is 2 (it operates nontrivially on 3), so $|G| = 6$, as we probably already knew.

And we can write out the operations on the cosets: x sends (123) and y sends (23). And this is exactly the standard representation for S_3 that we're used to!

Example 205

Let's apply the Todd-Coxeter algorithm to the tetrahedral group.

In the tetrahedral group, we can rotate around a face by $\frac{2\pi}{3}$, about a vertex by $\frac{2\pi}{3}$, or around an edge by π. Call these x, y, z respectively. If we pick the vertex to be one of the vertices of the face, and the edge to be the counterclockwise edge from that vertex along that face, we have the following relations:

$$x^3 = y^3 = z^2 = xyz = 1$$

(By the way, this tells us $xy = z^{-1} = z$.)

Are these relations enough to determine the group? Let's use Todd-Coxeter! We'll make the calculation easier by using generators x, y and having relations $x^3, y^3, xyz = xyxy$.

We'll use the subgroup $H = \langle x \rangle$, so 1 stands for the coset of H. In particular, this means x does nothing to 1.

	x	x	x		y	y	y		x	y	x	y
1	1	1	1	1				1	1			

We don't know what y does to 1, so we'll say that 1 sends it to 2 and then to 3.

	x	x	x		y	y	y		x	y	x	y
1	1	1	1	1	2	3	1	1	1	2	3	1

Now what happens to 2? We know y sends 2 to 3 and then 3 to 1 from above:

	x	x	x		y	y	y		x	y	x	y
1	1	1	1	1	2	3	1	1	1	2	3	1
2			2	2	3	1	2	2				2

and reading off the top row, we also know that x sends 2 to 3.

	x	x	x		y	y	y		x	y	x	y
1	1	1	1	1	2	3	1	1	1	2	3	1
2	3	$\boxed{4}$	2	2	3	1	2	2	3	1	1	2

(It's also possible right now that x is one of $1, 2, 3$; we don't know. For now, though let's put a 4 there. If we find $1 = 4$ or something, that's fine, and we can correct it later)

Now we can fill in the action for the cosets 3 and 4, since we know that x cycles $2, 3, 4$:

	x	x	x		y	y	y		x	y	x	y
1	1	1	1	1	2	3	1	1	1	2	3	1
2	3	4	2	2	3	1	2	2	3	1	1	2
3	4	2	3	3	1	2	3	3	4	4	2	3
4	2	3	4	4	4	4	4	4	2	3	4	4

So there are 4 cosets, so $\langle x \rangle$ has order 4. We know that $x^3 = 1$ and it operates on a 3-cycle, so x has order 3, so the order of the group is $4 \cdot 3 = 12$. This is indeed the order of the tetrahedral group, so the representation above with x and y is a valid description.

Remark 206. *Alternatively, we could have written out all possibilities for x and y. But this isn't commutative, so it's not enough to write any element as $x^a y^b$.*

So what do we know about what x and y do to the cosets? $x = (234)$ and $y = (123)$; these generate A_4, so the tetrahedral group is isomorphic to A_4.

As an interesting exercise, let's see what happens to $xyxy$. Remember that we're operating on **right** cosets, so $xyxy$ means to first apply x, then y, then x, then y: composing the permutations $(234), (123), (234), (123)$ in that order indeed yields the identity permutation, as we expect.

Time to use a "bad set of relations" to see what happens with Todd-Coxeter:

Example 207

Take $x^3, y^2, yxyxy$ as relations this time: What happens to the Todd-Coxeter algorithm?

	x	x	x		y	y		y	x	y	x	y
1			1	1		1	1					1

Let's take H to be the subgroup generated by x. Let's say y sends 1 to 2, so y sends 2 to 1.

	x	x	x		y	y		y	x	y	x	y
$\boxed{1}$	1	1	1	1	2	1	1	2			$\boxed{2}$	1

But x sends 2 and 1 to 1, so $1 = 2$. So $\langle x \rangle$ has index 1 (since we could have used this argument for any coset). This means $\langle x \rangle = G$, so $y = x^r$ is in the group generated by x. Then $yxyxy = x^{3r+2} = 1$, so $x^2 = 1 \implies x = 1$ and the whole group is trivial.

This algorithm is a deterministic way to determine our group, as long as the order is finite!

Remark 208. *The only problem is the name: we should list names alphabetically in mathematics.*

Let's turn our attention away from those relations for a second and speak more generally:

Definition 209

A **word** is any string of letters in an alphabet, along with their inverses.

For example, $yxy^{-1}x^2$ is a word. The free group is basically just the set of words, but if we have x and x^{-1} next to each other, they should cancel.

Definition 210

A **reduced word** is a word with no possible cancellation:

There's two problems now: we could end up with the empty word, which we can call 1. Also, we can cancel adjacent terms in different orders: in particular, if we had something like $xyy^{-1}x^{-1}x$, the end result could either be the first x or the last x. Luckily, the reduced word obtained by cancellation is **unique**, and it's not hard to prove this.

Definition 211

The **free group** is the set of reduced words, where the law of composition is juxtaposition and then cancellation. For instance,

$$(xyx^{-1})(xy^{-1}x) = xyx^{-1}xy^{-1}x = x^2.$$

Any subgroup of a free group aside from the trivial group is infinite, but we can mod out by a normal subgroup using relations.

There is a hard problem called the **word problem**. Given a bunch of generators and relations, we can ask whether a word is equal to 1. But it turns out there is no way to predict this with a computer program (it is undecidable)!

24 November 5, 2018

Recall that we can define a group G by specifying a set of **generators** $x, y, \cdots$ and relations $r_1, \cdots, r_k$, which are words consisting of x, x^{-1}, y, y^{-1}, and so on.

We basically started with the free group F on generators: the elements are all reduced words using those generators (and their inverses), and then the group G is defined to be F/N, where N is the **smallest normal subgroup** that contains those relations.

Fact 212

Ifa relation $r = 1$ is true in G, then grg^{-1} should also be true in G.

So the normal subgroup N is generated by all conjugates of relations

$$\{grg^{-1} \mid g \in F, r \text{ relation}\}.$$

This is unfortunately very hard to use, even though it's easy to define.

Example 213

Let the generators be x, y, and let's say $xy = yx$. Then $xyx^{-1}y^{-1}$ is the only relation.

So G is generated by x and y, and they commute, so it's just going to be isomorphic to $\mathbb{Z}^2$, since all elements can be written as $x^a y^b$. G is called the **free abelian group.**

Notice that $x^2y = yx^2$, so something like $x^2yx^{-2}y$ is in the normal subgroup N that we're modding out by. Specifically, we want the normal subgroup N generated by $g(xyx^{-1}y^{-1}x^{-1})g^{-1}$, where g is any element in the free group.

Well, it's pretty annoying to try to find $x^2yx^{-2}y$ explicitly in N: one way is to do

$$x(xyx^{-1}y^{-1})x^{-1}(xyx^{-1}y^{-1}) = x^2yx^{-2}y,$$

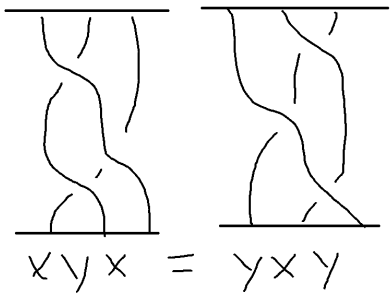
but it's a mess and not very important.

Example 214

Let's do another example of the Todd-Coxeter algorithm, this time with $xyx = yxy$. This is known as the **braid relation** (we'll add other relations in a second).

The idea here is to imagine having three strands that we try to braid together. Then x swaps the first and second (with the left one on top) and y swaps the second and third. (For example, x^2 twists the braid twice.) The inverses x^{-1} and y^{-1} also exist (we just swap so that the right is on top instead).

Here's a picture to explain why $xyx = yxy$: it can be checked that these two braids are equivalent.



To make it a finite group, let's also say that $x^3 = y^2 = 1$. Then the braid relation becomes $xyxyx^{-1}y = 1$:

x	x	x	y	y	x	y	x	y	x^{-1}	y
1		1	1	1	1					1

Take $H = \langle x \rangle$. Recall the rules for operating on cosets: generators operate as permutations, relations must be the identity permutation (operate trivially), and generators for H also fix the coset H. We're working with **right cosets**! So this means applying x sends 1 to 1:

x	x	x	y	y	x	y	x	y	x^{-1}	y
1	1	1	1	1	1	1				1

We don't know what y does to 1: let's say it sends 1 to 2. Then 2 must go back to 1.

x	x	x	y	y	x	y	x	y	x^{-1}	y
1	1	1	1	2	1	1	2			1
2					2		2	1	1	2

We don't know what happens to 2 under x, so let's say x cycles $2, 3, 4$.

x	x	x	y	y	x	y	x	y	x^{-1}	y
1	1	1	1	2	1	1	2	3	4	2	1
2	3	4	2	1	2	2	3	4	2	1	2

But now x^{-1} sends 4 to 2, so x sends 2 to 4. Thus $3 = 4$, and that means 2 and 3 are both sent to 3, so $2 = 3$ as well. So now our relation just looks like

x	x	x	y	y	x	y	x	y	x^{-1}	y
1	1	1	1	2	1	1	2	2	2	2	1
2	2	2	2	1	2	2	2	2	1	1	2

and now y sends both 1 and 2 to 2, so $1 = 2$. So there's only one coset, meaning that $\langle x \rangle = G$! But now a little more analysis collapses the group further: we see that y must be x^k and has order 2, so $y = 1$. This means $x^2 = x$, so $x = 1$ and we have the trivial group.

Proposition 215

If the Todd-Coxeter table is consistent, we have a correct operation on cosets.

Proof. We will show that we have a bijective map ϕ from I, the set of coset indices, to C, the set of cosets, which is compatible with the operations of the group. To do this, we first write out what the algorithm actually makes us do.

Let's say we're at some stage, and we have some indices I^*, as well as some operations of generators on I^*. When we fill in the table, one of two things always happens: we equate two indices if the rules tell us it is necessary, or we add new indices j such that $ig = j$ for some generator g and some existing index i.

We start with $I^* = \{1 = H1\}$, and we stop when we have a consistent table. So our map ϕ will send 1 to $H \cdot 1$. Whenever we equate new indices, we're saying they act the same under all rules, so this is consistent. On the other hand, adding new indices is definitely fine: if $i \to Ha$ in our map ϕ, then we just send $j \to Hag$. So we will always have a map $I \xrightarrow{\phi} C$ which is compatible with the operations.

To finish, we need to show that this map is bijective. Every index is in the orbit of 1, and the operation is surjective, so this map is surjective. Now, if $\phi i = \phi j$, then we know that $Ha = Hb \implies H = Hba^{-1}$, which means that ba^{-1} is an element of H. But now if ϕ takes j to Hb, then $ja^{-1} \to Hba^{-1} = H$, meaning that $ja^{-1} = 1$. On the other hand, $i \to Ha$, so $ia^{-1} = 1$. This means $i = j$, and we've shown injectivity. □

25 November 9, 2018

Quiz scores will probably be posted by tomorrow. There are many quizzes, so we won't get them back until Wednesday.

Definition 216

A **symmetric form** on a real vector space V is a function

$$V \times V \to \mathbb{R} : v, w \to \langle v, w \rangle$$

which satisfies the following properties:

- Symmetry: $\langle v, w \rangle = \langle w, v \rangle$
- Linearity in the second variable: $\langle v, wc \rangle = \langle v, w \rangle c$ and $\langle v, w_1 + w_2 \rangle = \langle v, w_1 \rangle + \langle v, w_2 \rangle$. Notice this also implies linearity in the first variable by symmetry.

(A motivating example of this is the standard dot product.) We can also define the matrix of a symmetric form with respect to a basis:

Definition 217

If we have a basis $\mathcal{B} = (v_1, \cdots, v_n)$, then the **matrix of the symmetric form** is the matrix $A = (a_{ij})$, where

$$a_{ij} = \langle v_i, v_j \rangle.$$

Then if X and Y are coordinate vectors of v and w, then

$$v = \mathcal{B}X, w = \mathcal{B}Y,$$

and we can compute the symmetric form by linearity:

$$\begin{aligned}\langle v, w\rangle &= \left\langle \sum v_i x_i, \sum v_j y_j \right\rangle \\ &= \sum_{i,j} x_i \langle v_i, v_j\rangle y_j \\ &= \sum_{i,j} x_i a_{ij} y_j \\ &= X^T A Y.\end{aligned}$$

Notice that if A is the identity matrix, this gives the standard dot product with respect to $\mathcal{B}$. It's interesting to think about what the standard dot product looks like with respect to other bases, but we'll answer that a bit later.

Recall that the standard dot product can be rewritten as

$$X^T X = \sum x_i^2 = |X|^2,$$

and it is always nonnegative (and only zero when $X = 0$). This prompts the following definition:

Definition 218

A symmetric real matrix (and corresponding form) is **positive definite** if $\langle v, v\rangle > 0$ for all $v \neq 0$.

It turns out that this doesn't depend on the basis, and if a form is positive definite, it is the dot product for some basis! Now, let's extend the idea of a symmetric form to complex numbers.

Definition 219

The **standard Hermitian form** on $\mathbb{C}^n$ is defined via

$$\langle X, Y\rangle = \overline{X}^T Y = \overline{x_1} y_1 + \cdots + \overline{x_n} y_n$$

where X, Y are complex vectors (and $\overline{x_1}$ denotes the complex conjugate of x_1).

In particular, for any vector,

$$\langle X, X\rangle = \overline{x}_1 x_1 + \cdots + \overline{x}_n x_n = |X|^2,$$

since for any complex number $x = a + bi$, $\overline{x}x = a^2 + b^2 = |x|^2$. This means that standard Hermitian forms are positive definite.

Definition 220

The **adjoint** of a matrix A is the (conjugate transpose) matrix $A^* = \overline{A}^T$.

For example,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^* = \begin{bmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{bmatrix}.$$

This has the following rules: $\boxed{A^{**} = A}$, and since $\overline{A}$, complex conjugation, is an automorphism of $\mathbb{C}$, $\overline{AB} = \overline{A} \cdot \overline{B}$. So we also have $\boxed{(AB)^* = B^* A^*}$, analogous to how $(AB)^T = B^T A^T$.

Definition 221

A **Hermitian form** on a complex vector space V is a function

$$V \times V \to \mathbb{C} : v, w \to \langle v, w \rangle$$

which satisfies the following properties:

- Hermitian symmetry: $\langle w, v \rangle = \overline{\langle v, w \rangle}$ for all $v, w \in V$.
- Linearity in the second variable: $\langle v_1, w_1 + w_2 \rangle = \langle v_1, w_1 \rangle + \langle v_2, w_2 \rangle$ and $\langle v, wc \rangle = \langle v, w \rangle c$. Hermitian symmetry gives us something a little more complicated than the real case: addition still holds as normal, but this time we have $\langle cv, w \rangle = \overline{\langle w, cv \rangle} = \overline{\langle w, v \rangle}\overline{c} = \langle v, w \rangle \overline{c}$.

As a check, for the standard Hermitian form,

$$\langle Y, X \rangle = Y^* X = (Y^* X)^{**} = (X^* Y)^* = \overline{X^* Y} = \overline{\langle X, Y \rangle}$$

where the second-to-last step comes from $X^* Y$ being a 1 by 1 matrix and therefore being equal to its own transpose.

The **matrix of a Hermitian form** has the same definition as before: we let $A_{ij} = \langle v_i, v_j \rangle$. Again, how do we compute the form explicitly? Let's say X and Y are coordinate vectors for two vectors v, w in some basis $\mathcal{B}$: then $v = \mathcal{B}X, w = \mathcal{B}Y$ just like before, and now

$$\begin{aligned} \langle v, w \rangle &= \left\langle \sum v_i x_i, \sum v_j y_j \right\rangle \\ &= \sum_{i,j} \overline{x_i} \langle v_i, v_j \rangle y_j \\ &= X^* A Y. \end{aligned}$$

As stated before, if a form is positive definite, it's a dot product in some basis. Say we have two bases $B = (v_1, \cdots, v_n)$ and $B' = (v_1', \cdots, v_n')$. Then we can relate these via a basechange matrix $B' = BP$: with this definition, if $v = BX, v' = B'X'$, then $PX' = X$. In our old basis, $\langle v, w \rangle = X^* A Y$, and in our new basis, $\langle v, w \rangle = \boxed{X'^* A' Y'}$, so we have

$$X^* A Y = (PX')^* A (PY') = \boxed{X'^* (P^* A P) Y'} \implies \boxed{A' = P^* A P}.$$

Remember that a linear operator can be associated with a matrix (in a given basis). **In a Hermitian form, we also have a matrix, but these two matrices are not the same**! It may seem like a symmetric form is somehow like a linear operator: this is not true, because linear transformations change to $P^{-1}AP$, while Hermitian forms change to $P^* A P$. Notice that this definition for forms preserves the Hermitian property: $(P^* A P)^* = P^* A^* P = P^* A P$ since $A^* = A$.

Lemma 222

A form is symmetric/Hermitian (for the real/complex case) if and only if $A^* = A$. (In the real case, $A^* = A^T$.)

Proof. For any Hermitian form A,

$$Y^* A X = (X^* A Y)^* = Y^* A^* X^{**} = Y^* A^* X,$$

so for this to be true for all X, Y, we must have $A^* = A$. □

Theorem 223

Given any Hermitian form, there exists a basis such that the matrix for the form is diagonal.

We won't do the proof explicitly here, but we know that A must have real entries on the diagonal, and then we'll have conjugates α_{ij} and $\overline{\alpha}_{ij}$ off the diagonal. In the 2×2 case, if the top left entry is nonzero, we can do a column operation to clear out the top right entry (and get the bottom left for free). A similar argument of row and column operations helps us prove this in general.

Corollary 224

A Hermitian form is **positive definite** if and only if there exists a basis so that the matrix is diagonal with positive entries.

This is because we can write $X^*AX = r_1\overline{x_1}x_1 + \cdots + r_n\overline{x_n}x_n$, and this quantity is always positive for nonzero x **if and only if** $r_i > 0$ for all i.

Remark 225. *We can normalize this diagonal matrix by dividing the ith row by $\sqrt{r_i}$. But we can only do this if r_is are all positive, so this isn't recommended in general.*

26 November 14, 2018

We'll start with a bit of review. Recall that a **Hermitian form** is a function $\langle v, w\rangle \to \mathbb{C}$ which is linear in w and satisfies $\langle w, v\rangle = \overline{\langle v, w\rangle}$. This tells us that $\langle cv, w\rangle = \overline{c}\langle v, w\rangle$.

We can write a matrix with respect to a basis $\mathcal{B} = (v_1, \cdots, v_n)$. Then the matrix satisfies $A = (a_{ij})$, $a_{ij} = \langle v_i, v_j\rangle$. To compute a form, we write $v = BX, w = BY$ for some coordinate vectors X, Y. Then $\langle v, w\rangle = X^*AY$.

Notice that A is a self-adjoint matrix, which means that $A^* = \overline{A}^T = A$. The proof is that

$$Y^*AX = \overline{X^*AY} = (X^*AY)^*$$

(because these are all 1 by 1 matrices), and therefore

$$Y^*AX = Y^*A^*X^* = Y^*A^*X$$

which can only be always true if $A = A^*$.

Theorem 226

Eigenvalues of a Hermitian matrix are real.

In the 1×1 case, a Hermitian matrix is always just a real entry, so its eigenvalue is that real entry. In the 2×2 case, Hermitian matrices look like

$$\begin{bmatrix} r_1 & \alpha \\ \overline{\alpha} & r_2 \end{bmatrix}$$

The characteristic polynomial for this matrix is $t^2 - (r_1 + r_2)t + (r_1r_2 - \overline{\alpha}\alpha)$. Then the discriminant is

$$(r_1 + r_2)^2 - 4(r_1r_2 - \overline{\alpha}\alpha) = (r_1 - r_2)^2 + 4\overline{\alpha}\alpha = (r_1 - r_2)^2 + 4|\alpha|^2 \geq 0.$$

Thus, applying the quadratic formula shows that all eigenvalues are real here, too.

Here's the proof in general:

Proof. Suppose λ is an eigenvalue of the Hermitian matrix A. Then there exists a vector $X \neq 0$ such that $AX = \lambda X$. Now

$$X^*AX = X^*(AX) = X^*\lambda X = \boxed{\lambda X^*X}$$

and also

$$X^*AX = (X^*A)X = (A^*X)^*X = (AX)^*X = (\lambda X)^*X = X^*\overline{\lambda}X = \boxed{\overline{\lambda}X^*X}$$

since $A^* = A$ (as A is Hermitian). But now $X^*X = |X|^2 > 0$ since we defined X to not be the zero vector, so $\lambda = \overline{\lambda}$, which means that λ is real. □

Corollary 227

A real symmetric matrix has real eigenvalues.

(This is because real symmetric matrices are Hermitian.)

Next, let's think a bit more about some properties of the inner product. In $\mathbb{R}^2$, X and Y are orthogonal if $X^TY = 0$ – more generally, $X^TY = |X||Y|\cos\theta$.

Definition 228

For a real symmetric or Hermitian form on a vector space V, we say that v and w are **orthogonal** if $\langle v, w\rangle = 0$.

This is not worth thinking about geometrically, especially since we need to fix a basis and think of many dimensions. (Note that v and w are orthogonal if and only if w and v are orthogonal.)

Definition 229

Let W be a subspace of V. Then the **orthogonal space to W**, denoted $W^\perp$, is

$$W^\perp = \{v \in V \mid \langle v, w\rangle = 0 \ \forall w \in W\}.$$

There is one unfortunate case: the form could be identically zero, which we don't want to think about.

Definition 230

The **nullspace** N of a form is the set of all vectors $v \in V$ such that $\langle v, w\rangle = 0$ for all $w \in V$. (In other words, $N = V^\perp$.) If $N = \{0\}$, the form is called **nondegenerate**.

Here's a way to restate the definition: in a nondegenerate form, for any $v \neq 0 \in V$, there is a $w \in V$ such that $\langle v, w\rangle \neq 0$.

Lemma 231

A form is nondegenerate if and only if its matrix with respect to any basis is invertible.

Proof. Consider the representation of $\langle v, w\rangle = X^*AY$. If A is singular (not invertible), there exists a vector $Y \neq 0$ such that $AY = 0$. Then X^*AY must be 0 for all X, so $Y \in N$, and the form is degenerate.

On the other hand, say that A is invertible. Then for any coordinate vector $Y \neq 0$, $AY \neq 0$. Therefore, we can always find X such that $X^*AY \neq 0$; in fact, we can just take $X = AY$. So Y is not in the nullspace, and since Y was an arbitrary nonzero vector, the nullspace must be trivial. □

If we have a subspace W of V, we can always look at the **restricted form** on W. Then the definition is exactly the same, except that the domain is just W instead of V. Then nondegeneracy comes up again in a more subtle way: it's possible a form is nondegenerate in one of the two spaces but not the other.

Example 232

Say that $\dim V$ is 2, and we have the form defined by $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. A is degenerate, but if W is the span of v_1, the form is nondegenerate on W.

On the other hand, if we take $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, this is a nondegenerate form, but if we look at only on the span of v_1, it becomes degenerate.

This means that we should be careful. **What does it mean for the form to be nondegenerate on W?** It could mean that the elements of w aren't in the nullspace: for every $w \in W$, we can find a $v \in V$ such that $\langle w, v \rangle \neq 0$. But it can also mean that when we restrict the form to W, it's nondegenerate: for every $w \in W$, there is another $w' \in W$ such that $\langle w, w' \rangle = 0$. So these are different things – usually we like the second definition more.

Lemma 233

The form on V is nondegenerate on W if and only if $W \cap W^\perp = \{0\}$.

In other words, if $w \neq 0 \in W$, then $w \in W^\perp$, so there exists w' such that $\langle w, w' \rangle \neq 0$. The proof basically follows from the definition of degeneracy.

Theorem 234

Given a form on V and a subspace W of V, such that the form is nondegenerate on W, we can write

$$V = W \oplus W^\perp.$$

In other words, $W \cap W^\perp = 0$ and $W + W^\perp = V$.

Proof. Choose a basis for V: $\mathcal{B} = (w_1, \cdots, w_k, u_{k+1}, \cdots, u_n)$ such that the w_i form a basis for W. Then the matrix of the form with respect to $\mathcal{B}$ is

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A is k by k and D is $n-k$ by $n-k$. We want B to be 0, because the u_i will then be orthogonal to w_is, and so every vector in V can be written as a direct sum of the w_is and the u_is.

So we'll change our basis with a matrix of the form $M' = P^* A P$. Letting $P = \begin{bmatrix} I & Q \\ 0 & I \end{bmatrix}$ for some undetermined Q,

$$P^* A P = \begin{bmatrix} I & 0 \\ Q^* & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & Q \\ 0 & I \end{bmatrix}$$

which results in $AQ + B$ where B originally was. But A is invertible, so we can just take $Q = A^{-1}B$ and we've diagonalized the block matrix! Notice that a change of basis does not change the fact that the matrix is Hermitian, so we don't need to verify that the bottom left block also disappears. □

27 November 16, 2018

Let V be a vector space with an inner product $\langle\ ,\ \rangle$, and let W be a subspace of V. Recall that v_1, v_2 are **orthogonal** if $\langle v_1, v_2\rangle = 0$: let $W^\perp$, defined to be $\{u \in V \mid u \perp W\}$, be the **orthogonal** space of W.

We say that a form is **nondegenerate** on W if for all $w \in W$, there exists a $w' \in W$ so that $\langle w, w'\rangle \neq 0$. (In other words, w doesn't collapse everything.) We found last time that a form is nondegenerate on W if and only if the matrix of the form restricted to W is invertible. (This means every element in V can be written as $w + u$, where $w \in W, u \in W^\perp$, in exactly one unique way. And this is equivalent to saying that $W + W^\perp = V$, while $W \cap W^\perp = \{0\}$.)

Now, we'll talk about the **orthogonal projection operator**. Consider a map $\pi : V \to W \subset V$. If $v = w + u$, with $w \in W, u \in W^\perp$, then v is sent to w. This is a linear transformation satisfying the two defining properties:

- If $w \in W$, then $\pi(w) = w$.
- If $u \in W^\perp$, then $\pi(u) = 0$.

There is a nice formula, but first let's discuss the concept of an **orthogonal basis.**

Definition 235

Let $(v_1, \cdots, v_n)$ be a basis of V. If $\langle v_i, v_j\rangle = 0$ for all $i \neq j$, which means that the matrix of the form with respect to V is diagonal, then the basis is **orthogonal**.

We now want to show that our vectors have "positive norm," so that we can form a basis with them:

Lemma 236

If a form $\langle, \rangle$ is not identically zero, then there exists v such that $\langle v, v\rangle \neq 0$.

Proof. We know there exists v_1, v_2 so that $\langle v_1, v_2\rangle = c \neq 0$. Now, change $v_2 = c^{-1}v_2$ (since the form is linear with respect to the second variable), so we have $\langle v_1, v_2\rangle = 1$. Now $\langle v_2, v_1\rangle = \overline{1} = 1$ and

$$\langle v_1 + v_2, v_1 + v_2\rangle = \langle v_1, v_1\rangle + \langle v_2, v_2\rangle + \langle v_2, v_1\rangle + \langle v_1, v_2\rangle = \langle v_1, v_1\rangle + \langle v_2, v_2\rangle + 2$$

so at least one of the $\langle v, v\rangle$ terms in this equation is not zero. □

Proposition 237

There exists an orthogonal basis for any nondegenerate Hermitian form.

Proof. If the form is identically zero, any basis is an orthogonal basis.

Otherwise, we pick v_1 such that $\langle v_1, v_1\rangle \neq 0$. Let W be the span of v_1. Then the matrix of form restricted to W is $(\langle v_1, v_1\rangle)$, which is nondegenerate. Thus $V = W \oplus W^\perp$, but we can induct on the dimension to show that $W^\perp$ also has an orthogonal basis. Tack on v_1 to this basis and we're done. □

Theorem 238 (Projection formula)

Let a form be nondegenerate on W with an orthogonal basis $(w_1, \cdots, w_k)$ (on W). Then if $v \in V$,

$$\pi(v) = w_1 c_1 + \cdots + w_k c_k$$

where $c_i = \frac{\langle v, w_i\rangle}{\langle w_i, w_i\rangle}$.

Notice that $\langle w_i, w_i \rangle$ is one of the diagonal entries of a diagonal matrix that is invertible, so it must be nonzero.

Proof. It is sufficient to show that $\pi(w_j) = w_j$ and $\pi(u) = 0$ for all $u \in W^\perp$. If u is orthogonal to W, then $\langle u, w_i \rangle = 0$ for all i, so $\pi(u) = 0$.

On the other hand, plugging in a basis vector w_j into the above definition, we have

$$\pi(w_j) = \sum_i w_i c_i = \sum w_i \frac{\langle w_j, w_i \rangle}{\langle w_i, w_i \rangle}.$$

But this is an orthogonal basis, so c_i is 0 except when $i = j$, in which case it is 1. So $\pi(w_j) = 0 + \cdots + 0 + w_j \cdot 1 = w_j$. □

Example 239

Let $V = \mathbb{R}^3$ with the standard form (dot product), and let W be the span of (w_1, w_2), where $w_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $w_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$. What is $\pi(v)$, where $v = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$?

A direct computation yields that $\langle w_1, w_1 \rangle = 3$ and $\langle w_2, w_2 \rangle = 6$, while $\langle v, w_1 \rangle = 6$ and $\langle v, w_2 \rangle = -3$. So

$$\pi(v) = \frac{6}{3} w_1 + \frac{-3}{6} w_2 = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} - \begin{bmatrix} 1/2 \\ 1/2 \\ -1 \end{bmatrix} = \boxed{\begin{bmatrix} 3/2 \\ 3/2 \\ 3 \end{bmatrix}}.$$

One special case of this formula is when $W = V$. Then if $(v_1, \cdots, v_n)$ is an orthogonal basis for V, then $v = v_1 c_1 + \cdots + v_n c_n$, where $c_i = \frac{\langle v, v_i \rangle}{\langle v_i, v_i \rangle}$. (Basically, this gives us a way to decompose our vector into components along the basis vectors)

Example 240

Take $w_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $w_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$, $w_3 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, and keep $v = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Then $\langle w_3, w_3 \rangle = 2$, and $\langle v_1, w_3 \rangle = -1$. Therefore, the projection formula tells us that

$$v = 2w_1 - \frac{1}{2} w_2 - \frac{1}{2} w_3.$$

(We can verify this by direct computation as well!)

It's tempting to get rid of the denominators in the projection formula. If $\langle w_i, w_i \rangle$ is positive, we can divide w_i by the square root of $\langle w_i, w_i \rangle$, and this just gives us $c_i = \langle v, w_i \rangle$. On the other hand, if $\langle w_i, w_i \rangle$ is negative, we'll get a denominator of -1. However, the vector will then have a square root term - division is better than having square roots, so (again) it is advised to not do this.

Example 241

Let $V = \mathbb{R}^{2\times 2}$, and let $\langle A, B \rangle$ be the trace of AB, $a_{11}b_{11} + a_{12}b_{12} + a_{21}b_{12} + a_{22}b_{22}$.

Let's choose an orthogonal basis for this vector space. This is not too hard: start with

$$v_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, v_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, v_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, v_4 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Then $\langle v_1, v_1 \rangle = \langle v_2, v_2 \rangle = 1$, $\langle v_3, v_3 \rangle = 3$ and $\langle v_4, v_4 \rangle = -2$. So now we can decompose $v = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$: we find that $\langle v, v_1 \rangle = 1, \langle v, v_2 \rangle = 4, \langle v, v_3 \rangle = 5$, and $\langle v, v_4 \rangle = -1$. So

$$v = 1v_1 + 4v_2 + \frac{5}{2}v_3 - \frac{1}{2}v_4$$

which is true because

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 5/2 \\ 5/2 & 0 \end{bmatrix} + \begin{bmatrix} 0 & -1/2 \\ 1/2 & 0 \end{bmatrix}.$$

Note that v_1, v_2, v_3 form a basis for the **symmetric matrices**, so if we wanted to project v to the space of symmetric matrices, v would just become $v + \frac{1}{2}v_4$. On the other hand, if we project to the space of skew-symmetric matrices (spanned by v_4), we just get $\frac{1}{2}v_4$. So this is a useful way to think about subspaces of our original vector space!

28 November 19, 2018

Today, we will discuss linear operators on a Hermitian space. First, a bit of review: let V be a complex vector space with a Hermitian form $\langle , \rangle$. Let's say there exists an orthogonal basis $(v_1, \cdots, v_n)$; that is, $\langle v_i, v_J \rangle = 0$ for all $i \neq j$.

Definition 242

If $\langle v_i, v_i \rangle > 0$ for all i, then we can scale our vectors so that $\langle v_i, v_i \rangle = 1$. Then this is called an **orthonormal** basis, and the matrix of the form is the identity.

In this basis, the form becomes the standard dot product $\langle v, w \rangle = X^*Y$, where $v = BX, w = BY$ with respect to B. An orthonormal basis exists if and only if the form is **positive definite**: $\langle v, v \rangle > 0$ for all v.

Definition 243

A vector space V with a positive definite form $\langle , \rangle$ is a **Hermitian space**.

We can also change from orthonormal bases to other orthonormal bases: the rule is that if B, B' are two bases, then we can write $B' = BP$ for some specific matrix P. Let M be the matrix of our positive definite form with respect to B, and let M' be the matrix with respect to B'. Then since $M = M' = I$ (we started and ended up with an orthonormal basis),

$$M' = P^*MP \implies I = P^*IP = P^*P.$$

Definition 244

A matrix P is **unitary** if $P^*P = I$.

(This is very similar to being orthogonal in the real-valued case.)

Example 245

What are the unitary matrices for 2×2 matrices?

Let $P = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} X & Y \end{bmatrix}$ for column vectors X, Y. Then

$$P^*P = \begin{bmatrix} a\overline{a} + c\overline{c} & \overline{a}b + \overline{c}d \\ \overline{b}a + \overline{d}c & \overline{b}b + \overline{d}d \end{bmatrix} = \begin{bmatrix} |X|^2 & X^*Y \\ Y^*X & |Y|^2 \end{bmatrix}$$

So if this is the identity, we just want $|X| = |Y| = 1$ and $X^*Y = 0$: our column vectors form an orthonormal basis.

We can also think about unitary operators in the framework of linear operators! Let $V \xrightarrow{T} V$ be a linear operator, and let A be the matrix of T with respect to an orthonormal basis. Then the change of basis matrix from B to B' for a linear operator is $A' = P^{-1}AP$. But if B' and B are orthonormal bases, then P must be unitary, so $P^*P = I \implies P^{-1} = P^*$. Therefore, $A' = P^*AP$ is true as well, so changing the basis for the linear operator and the Hermitian form gives us the **same expression**.

Definition 246

Given an operator $V \xrightarrow{T} V$ with matrix A, the **adjoint operator** is the operator T^* corresponding to the matrix A^*.

Here's a characteristic property for adjoint operators:

Proposition 247

For a linear operator T and its adjoint operator T^*, we have $\langle Tv, w\rangle = \langle v, T^*w\rangle$.

We can also do the same thing the other way around: we have $\langle v, Tw\rangle = \langle T^*v, w\rangle$.

Proof. Let $v = Bx, w = By$ with respect to a basis B. Then the coordinate vector of Tv is AX, so

$$\langle Tv, w\rangle = (AX)^*Y = X^*A^*Y.$$

Similarly, the coordinate vector of T^*w is A^*Y, so

$$\langle v, T^*w\rangle = X^*(A^*Y)$$

and indeed the two are equivalent. □

We can check that this definition still holds when we change to a different orthonormal basis: if we let B be the new matrix for T^*, then

$$B = P^*A^*P$$

but $B^* = P^*AP = A'$, so we still have $B = A'^*$ and everything works.

Lemma 248

If $v_1, v_2 \in V$, and $\langle v_1, w\rangle = \langle v_2, w\rangle$ for all $w \in V$, then $v_1 = v_2$.

This is true because by linearity, we have $\langle v_1 - v_2, w\rangle = 0$ for all w, but the form is positive definite, so it can only be zero when we plug in $w = v_1 - v_2$ if $w = v_1 - v_2 = 0$.

Here's a natural extension of what we've been discussing:

Definition 249
A linear operator T is a **Hermitian operator** if $T^* = T$; that is, for the associated matrices, $A^* = A$.

Looking back at the characteristic property, we have $\langle Tv, w\rangle = \langle v, Tw\rangle$. Indeed, it's true that $(AX)^*Y = X^*(AY)$, because $A^* = A$ for a Hermitian operator.

Definition 250
A linear operator T is **unitary** if $T^* = T^{-1}$; that is, $A^*A = I$.

This one has a characteristic property as well:

Proposition 251
For a unitary operator T, $\langle Tv, Tw\rangle = \langle v, w\rangle$.

Proof. We want (writing the inner products in martix form) the equality $(AX)^*AY = X^*Y$, which is satisfied for all X and Y as long as $A^*A = I$. Alternatively, we could have used the adjoint operator:

$$\langle Tv, Tw\rangle = \langle v, T^*Tw\rangle = \langle v, w\rangle.$$

□

Finally, we talk about a more general category of operators:

Definition 252
A linear operator T is **normal** if $T^*T = TT^*$.

This includes both the unitary and Hermitian operators, since unitary operators have both equal to the identity and Hermitian operators have $T = T^*$.

Proposition 253
$\langle Tv, Tw\rangle = \langle T^*v, T^*w\rangle$ for normal operators T.

Proof. We use the adjoint property: $\langle Tv, Tw\rangle = \langle v, T^*Tw\rangle = \langle v, TT^*w\rangle = \langle T^*v, T^*w\rangle$. □

The normal operators are completely uninteresting in general, but we have an important result (which we will prove next time):

Theorem 254 (Spectral Theorem)
Let $V \xrightarrow{T} V$ be a normal operator on a Hermitian space V. Then there exists an orthonormal basis of eigenvectors of T in V.

Corollary 255
For any Hermitian matrix A, there exists a unitary matrix P such that P^*AP is real diagonal. If A is unitary, there exists a unitary P such that P^*AP is diagonal and unitary.

In particular, let's say A was already diagonal and unitary. Then if those entries are $a_1, \cdots, a_n$, A^*A is the diagonal matrix with diagonal entries $\overline{a_i}a_i$. So this means all entries must have absolute value 1.

Example 256

Take A to be the rotation matrix $\begin{bmatrix} c & -s \\ s & c \end{bmatrix}$, where $c = \cos\theta$ and $s = \sin\theta$. Then $A^*A = I$, so this matrix is unitary and orthogonal.

Then the Spectral Theorem says there exists a unitary matrix P such that P^*AP is diagonal: it turns out (after some computation) to be $P^*AP = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}$.

29 November 21, 2018

We'll start with some review. Let V be a Hermitian space with a positive definite form $\langle , \rangle$. (If $\mathcal{B} = (v_1, \cdots, v_n)$ is an orthonormal basis, then the form is just $\langle v, w \rangle = X^*Y$. And if we want to change between two orthonormal bases via $\mathcal{B}' = \mathcal{B}Q$, then Q is unitary.) Then if T is a linear operator $V \xrightarrow{T} V$ with matrix A with respect to $\mathcal{B}$, then the **adjoint operator** T^* has matrix $A^* = \overline{A}^T$.

Generally, we always have $\langle Tv, w \rangle = \langle v, T^*w \rangle$, or equivalently, $\langle T^*v, w \rangle = \langle v, Tw \rangle$. T is called **Hermitian** if $T^* = T$; then we also have $\langle Tv, w \rangle = \langle v, Tw \rangle$. T is **unitary** if $T^*T = I$; in this case, we have $\langle Tv, Tw \rangle = \langle v, w \rangle$. Finally, T is **normal** (which encompasses the first two categories) if T^* and T commute. In these cases, we always have $\langle Tv, Tw \rangle = \langle T^*v, T^*w \rangle$.

The main result of today is the **spectral theorem**, which says that if T be a normal operator on a Hermitian space, then there exists an orthonormal basis of eigenvectors of T in V. We can also state this result in matrix form:

Theorem 257 (Spectral Theorem, matrix form)

Let A be a matrix such that $AA^* = A^*A$. Then there exists a unitary Q such that Q^*AQ is diagonal (which is the eigenvector basis).

Two special cases: if A is Hermitian ($A^* = A$), there exists Q unitary such that $Q^*AQ = D$ is a real diagonal matrix. Meanwhile, if A is unitary, we can find Q such that Q^*AQ is unitary and diagonal; in particular, the diagonal entries must all have magnitude 1.

To prove this result, we'll start by thinking about invariant subspaces. If $V \xrightarrow{T} V$ is our linear operator, and W is a subspace of V, then a subspace W is **T-invariant** if $TW \subset W$.

Lemma 258

If W is T-invariant, then $W^\perp$ is T^*-invariant.

Proof of lemma. Take a vector $u \in W^\perp$; we want to show $T^*u \in W^\perp$. In other words, we need to show that

$$\langle T^*u, w \rangle = 0 \, \forall w \in W.$$

But by properties of the form, $\langle T^*u, w \rangle = \langle u, Tw \rangle$, and since $Tw \in W$ (by assumption that W is T-invariant) and $u \in W^\perp$, this must be equal to 0. Thus u is still in $W^\perp$ under T^*, so $W^\perp$ is T^*-invariant. □

Lemma 259

If T is a normal operator with eigenvector v such that $Tv = \lambda v$, then v is also an eigenvector of T^* with eigenvalue $\overline{\lambda}$.

Proof of lemma. First, we do a special case. If $\lambda = 0$, then $Tv = 0$. Since T is normal,

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

But the form is positive definite, so this can only be 0 if T^*v is the zero vector, and indeed v is an eigenvector of T^* with eigenvalue $\overline{0}$.

Now let's say λ is arbitrary. Let $S = T - \lambda I$. Then

$$Sv = Tv - \lambda v = 0,$$

so v is an eigenvector of S with eigenvalue 0. Notice that S is normal, because

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

while

$$SS^* = TT^* - \overline{\lambda}T - \lambda T^* + \lambda\overline{\lambda}I$$

which are clearly equivalent since $TT^* = T^*T$. Thus, by the special case above, v is also an eigenvector of S^* with value 0. This means $S^*v = 0 \implies T^*v - \overline{\lambda}v = 0 \implies T^*v = \overline{\lambda}v$, completing the proof. □

With this, we can finally prove our main result:

Proof of the spectral theorem. We are given a normal operator $V \xrightarrow{T} V$. Pick an eigenvector v_1 such that $Tv_1 = \lambda_1 v_1$, and let W be the span of v_1. The form is nondegenerate on W, so $V = W \oplus W^\perp$. Notice that W is T-invariant (T just scales W), so $TW^\perp \subset W^\perp$. In other words, we can restrict the operator T to $W^\perp$, and it will still be normal. Now just induct: take v_1 along with the orthonormal basis of $W^\perp$. □

Proposition 260 (Polar Decomposition)

Every invertible complex-valued matrix can be written uniquely as a product of the form $A = UP$, where U is unitary and P is positive definite Hermitian.

In other words, $GL_n(\mathbb{C})$ is bijective to $U_n \times \mathcal{P}$.

Lemma 261

If A is invertible, then A^*A is positive definite and Hermitian.

Proof. This matrix is Hermitian because $(A^*A)^* = A^*(A^**) = A^*A$. Meanwhile, for all $x \neq 0$, we have that

$$0 < \langle Ax, Ax\rangle = x^*A^*Ax$$

so A^*A is indeed positive definite. □

Proof of Proposition 260. Using the previous lemma, we know by the Spectral Theorem that there exists a unitary Q such that $Q^*(A^*A)Q = D$ is diagonal with positive real entries. If D has diagonal entries $d_1, \cdots, d_n$, let R have only entries $r_i = \sqrt{d_i}$ on the diagonal. Then

$$Q^*A^*AQ = R^2 \implies A^*A = (QRQ^*)^2.$$

Since Q is unitary, QRQ^* is just a change of basis. But this means that $P = QRQ^*$ is also positive definite and Hermitian, so

$$A^*A = P^2 = P^*P.$$

Therefore, $(P^{*^{-1}}A^*)(AP^{-1}) = I$, so $(AP^{-1})^*(AP^{-1}) = I$. So $U = AP^{-1}$ is unitary, and $A = UP$ – we've indeed written A in the desired form.

Finally, showing uniqueness comes from the fact that any matrix that is both positive definite Hermitian and unitary is the identity matrix (because all eigenvalues must be 1). □

30 November 26, 2018

Definition 262

A **quadric** Q is the locus of points that satisfy $f(x) = 0$ for some quadratic equation in $\mathbb{R}^n$.

Let's start studying these objects with some linear algebra:

Definition 263

A real vector space V with a positive definite symmetric form $\langle , \rangle$ is called a **Euclidean space.**

We use orthonormal bases (just like we did in Hermitian spaces), and a change of orthonormal bases is given by $PX' = X$ for some orthogonal matrix P. As before, a linear operator $V \xrightarrow{T} V$ is **symmetric** if its matrix A with respect to an orthonormal basis is symmetric. Alternatively, we could just say that $\langle Tv, w\rangle = \langle v, Tw\rangle$ (expand out the forms of $(AX)^*Y$ and X^*AY, noticing that $A = A^*$).

Then the Spectral Theorem holds here, too:

Theorem 264

There exists an orthonormal basis of eigenvectors for a symmetric operator. In other words, if A is a real symmetric matrix, then there exists some base change P, which is orthogonal, such that $P^*AP = D$ is diagonal.

Let's try to show this.

Remark 265. The following are just special cases of what we discussed above! For example, the lemma below is just a special case of the lemma from last class.

Lemma 266

Let T be a symmetric operator. If W is T-invariant, which means $TW \subset W$, then $W^\perp$ is also T-invariant.

Proof. If $u \in W^\perp$, then $\langle u, w\rangle = 0$ for all $w \in W$. Then

$$\langle Tu, w\rangle = \langle u, Tw\rangle = 0$$

since $Tw \in W$ and u is orthogonal to anything in W. Thus, Tu is orthogonal to all $w \in W$, so $Tu \in W^\perp$ as well. □

Lemma 267

Eigenvalues of a real symmetric matrix are real.

Again, this is just a repeat of a previous proof.

Proof of Theorem 264. Choose an eigenvector v_1 such that $Tv_1 = \lambda v_1$. Let W be the span of v_1; the form is nondegenerate here because it is positive definite, and it is also nondegenerate on $W^\perp$. Normalize v_1 to have length 1. Thus, $V = W \oplus W^\perp$, and by induction on the dimension we have some orthonormal basis $(v_2, \cdots, v_n)$ for $W^\perp$. Then V has an orthonormal basis $(v_1, \cdots, v_n)$ as desired. □

Let's use this to classify quadrics. We'll start with conics: this is the locus of points that satisfy

$$a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 + b_1x_1 + b_2x_2 + c = 0.$$

We'll classify such quadrics up to an isometry f of $\mathbb{R}^2$, which we can write in the form $t_v\phi$, where $t_v(x) = x + v$ is a translation and ϕ is some orthogonal operator (reflection or rotation).

In matrix form, if $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, our conic looks like

$$q(x_1, x_2) = X^TAX + BX + C.$$

But we want a symmetric matrix, so let's split the $a_{12}x_1x_2$ term into a part for x_1x_2 and a part for x_2x_1. In general for a quadric, this would look like

$$\sum_{i,j} x_i a_{ij} x_j + \sum_i b_i x_i + c$$

where $a_{ij} = a_{ji}$. Thus we've made A into a symmetric matrix, so we can diagonalize and **remove all cross terms** x_ix_j. In other words, applying the change of coordinates $X = PX'$ for some orthogonal matrix P, we find that $A' = P^TAP$ is diagonal, which results in (dropping the primes because it's not necessray)

$$q(x_1, x_2) = a_{11}x_1^2 + a_{22}x_2^2 + b_1x_1 + b_2x_2 + c$$

(For a general quadric in n variables, this is $\sum a_{ii}x_i^2 + \sum b_ix_i + c$.) But now we're allowed to apply a translation. We can **complete the square** for each variable and translate by $v_i = -\frac{b_i}{2a_i}$, and this will toss all of those linear terms out! So now our new form is (almost always)

$$q(x_1, x_2) = a_{11}x_1^2 + a_{22}x_2^2 + c$$

and in general, this means a quadric is (almost always) equivalent to $\sum a_{ii}x_i^2 = c$.

Remark 268. *Well, there is an edge case. If $a_{11} = a_{22} = 0$, that would just be a linear equation, so ignore that case. But it's possible that (for example) we have one of the quadratic terms disappearing, so $a_{22} = 0$. To avoid degeneracy in this case, we must have $a_{11}, b_2 \neq 0$. Then we can translate in x_2 to remove the constant and translate in x_1 to remove the x_1 term, and we have an equation of the form $a_{11}x_1^2 + b_2x_2 = 0$, which is a* **quadratic**.*)*

Assume we're not in that degenerate edge case, and WLOG let $a_{11} > 0$. Then we can write our equation as

$$a_{11}x_1^2 \pm a_{22}x_2^2 = c.$$

If we have a $+$, then we get an ellipse for $c > 0$, just the point 0 for $c = 0$, and nothing for $c < 0$. (The latter two are "degenerate conics.") On the other hand, we have a $-$, then that's a hyperbola for $c \neq 0$ and two parallel lines for $c = 0$. Again, the latter case is degenerate.

Well, we could do this in any number of variables if we wanted – we just look at the number of signs that are positive versus negative and classify into different surfaces.

Example 269

What about quadrics in three dimensions?

We have $q(x_1, x_2, x_3) = X^T AX + BX + C$; make A diagonal again using the Spectral Theorem. Now, if $a_{ii} \neq 0$ for all i, we translate to get rid of bs, which gives

$$a_{ii}x_1^2 + a_{22}x_2^2 \pm a_{33}x_3^2 = c$$

(we always have 0, 1, 2, or 3 plus signs, and we can multiply by -1 if we have 0 or 1). Three pluses give an ellipsoid for $c > 0$, two pluses give a hyperboloid if $c \neq 0$.

Now what about $a_{11}x_1^2 + a_{22}x_2^2 - a_{33}x_3^2 = 0$? This equation is homogeneous, so scaling still satisfies the equation. This turns out to be a cone through the origin! While this is a degenerate quadric, it is still pretty interesting.

Remark 270. *Suppose we want to sketch a picture of the locus of points $g(x_1, x_2) = c$ for some $c \approx 0$, given the graph of $g(x_1, x_2) = 0$. The idea is to draw regions in the plane where the function is positive and negative, and we just trace near the curve $g = 0$ either on the positive or negative side!*

So the locus of $a_{11}x_1^2 + a_{22}x_2^2 - a_{33}x_3^2 = c$ for $c > 0$ is a vase-shaped object, and the locus for $c < 0$ is two cup-shaped objects. And to complete our classification, we have our paraboloids, of which there are two kinds: the standard bowl (which looks like $x_3 = x_1^2 + x_2^2$) and the two-sheeted hyperboloid (which looks like $x_3 = x_1^2 - x_2^2$).

31 November 28, 2018

Today we're going to talk about a special group:

Definition 271

The **special unitary group** SU_2 is the set of matrices $A \in GL_2(\mathbb{C})$ that satisfy $A^*A = I$ with determinant 1.

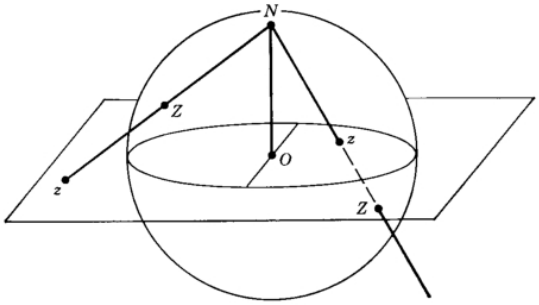
In particular, if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $A^* = \begin{bmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{bmatrix}$ must be equal to $A^{-1} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. This means $d = \overline{a}$ and $c = -\overline{b}$, and furthermore, $\overline{a}a + \overline{b}b = 1$.

If we write $a = x_0 + x_1 i$, $b = x_2 + x_3 i$, then a necessary and sufficient condition is that

$$x_0^2 + x_1^2 + x_2^2 + x_3^2 = 1.$$

This is the **unit 3-sphere** in $\mathbb{R}^4$, since the surface has dimension 3.

One way to analyze this is by **stereographic projection**. Let the x_0 axis point north, and note that the identity matrix I corresponds to the point $(1,0,0,0)$. So let I be the north pole; then for any matrix A, project A onto the plane $x_0 = 0$. This is defined everywhere, except the north pole I gets sent to infinity. Here's a visual depiction (from Google) of the stereographic projection of the 2-sphere:



Let's compute the formula for this projection explicitly: let A correspond to point (x_0, x_1, x_2, x_3). Then the line I will be the line $I + t(A - I)$ for some scalar t, which is equal to $(1 - t + tx_0, tx_1, tx_2, tx_3)$. Setting the first coordinate to zero, we want $t = \frac{1}{1-x_0}$. Thus, we've found the projection of SU_2 down to $\mathbb{R}^3$:

$$\pi(A) = \left(0, \frac{x_1}{1-x_0}, \frac{x_2}{1-x_0}, \frac{x_3}{1-x_0}\right).$$

When we project a sphere in 4-space down to 3-space, we'll get an ellipsoid. Consider the intersection of $x_0 = 0$ with the unit 3-sphere $x_0^2 + x_1^2 + x_2^2 + x_3^2 = 1$, and call this the **equator** $\mathbb{E}$. The interior of this region can be represented as $x_1^2 + x_2^2 + x_3^2 \leq 1$, which is B^3, the unit 3-ball. Notice that there are two points that are directly above or below each point inside the equator (except the boundary).

The lower half of the sphere goes inside the equator and the outer half of the sphere goes outside the equator under this stereographic projection. Also, every point in V (the hyperplane containing $\mathbb{E}$) corresponds to a point on the sphere; we just need to add a point at infinity to get I. (This is kind of like adding one point to a line to get a circle.) So if we let S^3 be a 3-sphere, we can represent this as

$$S^3 \sim V \cup \{I\}$$

which is $\mathbb{R}^3$ plus a point, or

$$S^3 = B^3 \underset{\mathbb{E}}{\cup} B^3.$$

SU_2 is very symmetric; let's undo some of that symmetry! Let $\mathbb{E}$ be the set of points in SU_2 that satisfy $x_0 = 0$. The trace of the matrix $\begin{bmatrix} a & b \\ -\overline{b} & \overline{a} \end{bmatrix}$ is $2x_0$, since $a = x_0 + x_1 i$. In other words, the equator $\mathbb{E}$ consists of those points where **the trace of A is 0**.

Definition 272

A **latitude** of SU_2 is the locus of points $x_0 = c$ for some $-1 < c < 1$. This is equivalent to saying that the trace of A is $2c$.

In particular, the points in SU_2 on a given latitude are those that satisfy $x_1^2 + x_2^2 + x_3^2 = 1 - c^2$.

Proposition 273

The latitude $x_0 = c$ is a conjugacy class.

Proof. We'll use the Spectral Theorem! A is unitary, so there exists P unitary so that P^*AP is diagonal. This diagonal matrix will be of the form $D = \begin{bmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{bmatrix}$, and this still has determinant 1, so λ must have absolute value 1. If P had determinant 1, we'd be done (because we would show that every matrix of a fixed trace is conjugate to D). Well, the determinant of P^* and the determinant of P are complex conjugates of each other, and $P^*P = I$, so the determinant of P is equal to δ for some $|\delta| = 1$.

Now just change P to $\begin{bmatrix} \delta^{-1/2} & 0 \\ 0 & \delta^{-1/2} \end{bmatrix} P$, and the determinant of P (and therefore also P^*) will be 1. P^*AP will still be diagonal, so every A in SU_2 is conjugate to a diagonal matrix. Well, conjugate elements have equal eigenvalues. so we can't have two diagonal matrices of this type with different traces cannot be conjguate.

Thus our matrix is conjugate to a matrix of the form $\begin{bmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{bmatrix}$, where $\lambda + \overline{\lambda}$ is $2c$ (so $\lambda = c \pm \sqrt{1-c^2}i$). These two matrices are also conjugate to each other (if we conjugate by $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$), so $x_0 = c$ is indeed a conjugacy class, as desired. □

Now, let's define a Hermitian form on SU_2:

Definition 274

Say that a matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SU_2$ corresponds to (x_0, x_1, x_2, x_3). If B similarly corresponds to (y_0, y_1, y_2, y_3), define the bilinear form $\langle A, B \rangle = x \cdot y$.

In other words, we carry over the (real-valued) dot product on $\mathbb{R}^4$ over to SU_2, and there's a nice way to describe this:

Proposition 275

The Hermitian form on SU_2 satisfies $\langle A, B \rangle = \frac{1}{2}\text{tr}(A^*B)$.

Proof. This is a direct computation: Let $A = \begin{bmatrix} a & b \\ -\overline{b} & \overline{a} \end{bmatrix}$, and let $B = \begin{bmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{bmatrix}$. Then

$$A^*B = \begin{bmatrix} \overline{a}\alpha + b\overline{\beta} & ? \\ ? & \overline{b}\beta + a\overline{\alpha} \end{bmatrix}.$$

So the trace is

$$(\overline{a}\alpha + a\overline{\alpha}) + (\overline{b}\beta + b\overline{\beta}).$$

Now $(\overline{a}\alpha + a\overline{\alpha}) = (x_0 - x_1 i)(y_0 + y_1 i) + (x_0 + x_1 i)(y_0 - y_1 i)$; the cross-terms cancel and we're left with $2x_0y_0 + 2x_1y_1$. The other term gives us $2x_2y_2 + 2x_3y_3$, and plugging everything in indeed yields our result. □

One nice property of this form is that $\langle A, A \rangle = \frac{1}{2}\text{tr}(A^*A) = 1$. Note that if A is on the equator,

$$\langle I, A \rangle = \frac{1}{2}\text{tr}(IA) = \frac{1}{2}\text{tr}(A) = 0.$$

So I and A are orthogonal unit vectors for any $A \in \mathbb{E}$, which makes sense: the north pole is orthogonal to the equator.

Definition 276

A **longitude** of SU_2 is a 1-dimensional great-sphere L through the north and south pole I and $-I$.

Here's one way to describe a longitude more explicitly: take a two-dimensional subspace W of $\mathbb{R}^4$ which contains $(\pm 1, 0, 0, 0)$, and intersect it with S^3. This will intersect the equator at two points: call them A and $-A$ (they are antipodes). Then (I, A) forms an orthonormal basis for W, since the two vectors are orthogonal.

This longitude L is then a unit circle in our two-dimensional subspace W. Since I and A are orthonormal, L consists of the points $\{\cos\theta I + \sin\theta A\}$. This will have length 1 because

$$\langle cI + sA, cI + sA\rangle = c^2\langle I, I\rangle + s^2\langle A, A\rangle = c^2 + s^2 = 1.$$

By the way, it's interesting that the 1-sphere has a group law and so does the 3-sphere, but the 2-sphere does not. It turns out those two are the only spheres with group laws, and that's a theorem in topology. This is related to the Hairy Ball theorem, and we should search up a paper by Milnor for more information!

32 November 30, 2018

Today we're going to talk about SO_3, but it'll take a while to get there. Recall the definition of SU_2: it is the set of all matrices of the form $A = \begin{bmatrix} a & b \\ -\overline{b} & \overline{a} \end{bmatrix}$, where $\overline{a}a + \overline{b}b = 1$. This can be bijected to the 3-sphere in $\mathbb{R}^4$: letting $a = x_0 + x_1 i$, $b = x_2 + x_3 i$, we want $x_0^2 + x_1^2 + x_2^2 + x_3^2 = 1$. The characteristic polynomial of A is $t^2 - 2x_0 t + 1$, so the eigenvalues of our matrix A are $x_0 \pm \sqrt{1 - x_0^2}\, i$, which is $\cos\theta \pm i\sin\theta$ for some θ.

Also recall that we defined a form on SU_2 as well: if A corresponds to X in $\mathbb{R}^4$ and B corresponds to Y, we just want $\langle A, B\rangle = X \cdot Y$. This turns out to be equal to $\frac{1}{2}\mathrm{tr}A^*B$.

We also had a geometric interpretation of all of this: define the **equator** $\mathbb{E}$ to be the $\{A \in SU_2 \mid x_0 = 0\}$. A being on the equator is equivalent to having a trace of 0, which is equivalent to saying the eigenvalues are $\pm i$, which is equivalent to saying A^2 has a double eigenvalue at -1, so $A^2 = -I$ if A has to be of this form. This is additionally equivalent to having $I \perp A$ (since $\langle I, A\rangle = 0$). Finally, this is equivalent to A being **skew-Hermitian**: $A^* = -A$, since a must be pure imaginary. Thus, the equator is a real vector space of dimension 3 (b can be anything and a is pure imaginary); an orthonormal basis of the equator is

$$\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$$

Denote these $\underline{i}, \underline{j}, \underline{k}$ respectively. These satisfy the **quaternion relations**

$$\underline{ij} = -\underline{ji} = \underline{k}, \underline{jk} = -\underline{kj} = \underline{i}, \underline{ki} = -\underline{ik} = \underline{j}.$$

As a sidenote, these are actually i times the **Pauli matrices**. If we multiply a skew-symmetric matrix by i, we get a Hermitian matrix, so we end up with the three Hermitian matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Recall that the equator $\mathbb{E}$ is a conjugacy class in SU_2, so we can say that SU_2 operates on $\mathbb{E}$. Given $P \in SU_2, A \in \mathbb{E}$, the operation is that

$$P * A = PAP^*.$$

Geometrically, the equator $\mathbb{E}$ is a 2-sphere in $V(x_1, x_2, x_3)$ (since $x_0 = 0$).

Theorem 277

The operation of $P \in SU_2$ on the equator $\mathbb{E}$ is a rotation.

One way we could show this is by writing $A = \begin{bmatrix} a & b \\ -\overline{b} & \overline{a} \end{bmatrix}$, do a similar thing for P, and show that PAP^* gives another matrix similar to A. But this is disgusting computation; instead, let's use Euler's theorem.

Proof. We know that rotations are elements of SO_3 if they are linear operators, orthogonal, and have determinant 1. Thus, we just need to verify that conjugation by P satisfies all of these! For linearity, we just need to show $P(A_1 + A_2)P^* = PA_1P^* + PA_2P^*$ (true by the distributive law) and that $P(rA)P^* = rPAP^*$ (which is also true).

To show orthogonality, we show that the operator preserves the Hermitian form: we want $\langle A, B\rangle = \langle PAP^*, PBP^*\rangle$. We know that

$$\langle A, B\rangle = \frac{1}{2}\text{tr}(A^*B).$$

Note that

$$\langle PAP^*, PBP^*\rangle = \frac{1}{2}\text{tr}(PA^*P^*PBP^*) = \frac{1}{2}\text{tr}(PA^*BP^*),$$

and conjugation preserves eigenvalues, so the two above expressions are equal since the trace is the sum of the eigenvalues of a matrix!

Finally, for determinant, we argue by continuity. SU_2 is path-connected, so draw a path from I to P. Compute the value of MAM^* as M ranges along this path – note that orthogonal matrices have determinant 1 or -1. But MAM^* is orthogonal, so we can't get to -1, and therefore PAP^* must have determinant 1 as well. □

So what is the rotation? We want to determine both the axis and the angle given conjugation by some matrix P.

To find the axis of rotation, we want to find the fixed point (since it'll just be the line through the origin and that fixed point). So given P, we want to find a matrix A such that $PAP^* = A$. Let's say P is on some given **longitude** (recall this is a great-circle through I and $-I$). If the great-circle intersects the equator at a point B (it'll intersect at two points, pick one), then we can write

$$P = \cos\theta I + \sin\theta B.$$

So a direct computation shows that

$$PAP^* = (cI + sB)A(cI^* + sB^*) = (cI + sB)A(cI - sB)$$

(B is on the equator, so $B^*B = I$ and $B^2 = -I$. This means $B^* = -B$; that is, it is a skew-hermitian matrix). Now note that B commutes with $cI - sB$, so

$$PBP^* = (cI + sB)(cI - sB)B = (c^2B - s^2B^2) = (c^2 + s^2)B = B$$

and so B is fixed – we've found our fixed point!

This just leaves our other question: what is the angle of rotation? If we conjugate by I or by $-I$, we get the identity, so it seems like the angle is moving twice as fast as we do around our 3-sphere. So we can make the following guess:

Proposition 278

(Let B be a point on the equator.) The angle α of rotation by $P = cI + sB$, where $c = \cos\theta, s = \sin\theta$, is $\pm 2\theta$.

Proof. Take a point on the equator $A \in \mathbb{E}$ such that $A \perp B$. Letting $C = AB$, we want to check that (A, B, C) is an orthonormal basis of $\mathbb{E}$, with relations $AB = -BA = C, BC = -CB = A, CA = -AC = B, A^2 = B^2 = C^2 = -I$.

Notably, since $\operatorname{tr}(A^*B) = 0$ (by construction) and $A^* = -A$ (we showed that matrices on the equator are skew-hermitian), $\operatorname{tr}(AB) = 0$ as well, so $C \in \mathbb{E}$. Verifying the other relations is to do calculations like $-(AB) = (-B)(-A) = BA$. So now

$$PAP^* = (cI + sB)(A)(cI - sB) = c^2A - s^2BAB + cs(BA - AB)$$

and since $AB = C, BA = -C$, and $BAB = BC = A$, this is equal to

$$(c^2 - s^2)A - 2csC = \cos(2\theta)A - \sin(2\theta)C$$

so $\alpha = -2\theta$. □

So now consider the map

$$SU_2 \overset{\text{operates on } \mathbb{E}}{\to} SO_3.$$

The kernel of this map is $\{\pm I\}$, and the map is surjective, so we finally have a way of representing our group SO_3:

$$\boxed{SO_3 \cong SU_2/\{\pm I\}}.$$

In one dimension, this is saying that if we take half a circle and glue the two endpoints together, we get a circle. In two dimensions, if we take half a sphere and glue the opposite diametrical points, we get a Mobius band. This is a non-orientable surface, and it's called the real projective plane RP^2. Finally, given a 3-sphere, the picture is just completely confusing. But SO_3 is also called the **real projective 3-space** RP^3.

33 December 3, 2018

Recall that the exponential function has the power series

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots.$$

Definition 279

The **matrix exponential** is defined as

$$e^A = 1 + \frac{A}{1!} + \frac{A^2}{2!} + \cdots.$$

Professor Artin imagines that someone was completely clueless and came up with this as a definition. This has some properties:

- This series uniformly converges on any bounded set of matrices. This is analysis, and it's not too hard to prove.
- Conjugating e^A with P does something nice:

$$Pe^AP^{-1} = e^{PAP^{-1}}.$$

This can be easily checked by just plugging both sides into the definition and using distributivity.

- The eigenvalues of e^A are e^λ, where λ is an eigenvalue of A. This is because if we have $\lambda v = Av$, then

$$e^A v = \left(I + \frac{A}{1!} + \frac{A^2}{2!} + \cdots\right) v = v + \frac{A}{1!}v + \frac{A^2}{2!}v + \cdots$$

and this is equal to

$$v + \frac{\lambda}{1!}v + \frac{\lambda^2}{2!}v \cdots = e^\lambda v.$$

- If $AB = BA$, then $e^{A+B} = e^A e^B$.

Let's do an example: take $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Then

$$e^A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{1!}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 0 + \cdots = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Similarly, if we take $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, since $B^2 = B$, we get

$$e^B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{1!}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{2!}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \cdots = \begin{bmatrix} e & 0 \\ 0 & 1 \end{bmatrix}.$$

But notice that $AB \neq BA$, and $e^{A+B} \neq e^A e^B$ in general.

(One comment: it's easy to compute the exponential for a diagonal matrix; just exponentiate each entry.)

- Now let's write down the matrix exponential e^{At} as a function of t:

$$e^{At} = I + \frac{1}{1!}At + \frac{1}{2!}(At)^2 + \cdots.$$

Notice that t is a scalar, so we must have

$$e^{As+At} = e^{As}e^{At}$$

for any scalars s, t, because As and At commute. Thus, we have a homomorphism $\mathbb{R}^+ \to GL_n$ via $t \to e^{At}$.

- e^{At} is a differentiable function of t; in particular, define

$$\frac{d}{dt}e^{At} = \lim_{\Delta t \to 0} \frac{e^{A(t+\Delta t)} - e^{At}}{\Delta t}.$$

This can be written as

$$\lim_{\Delta t \to 0} \frac{e^{A\Delta t}e^{At} - e^0 e^{At}}{\Delta t} = \lim_{\Delta t \to 0} \frac{e^{A(0+\Delta t)} - e^{A\cdot 0}}{\Delta t}e^{At} = e^{At}\frac{d}{dt}(e^{At})_{t=0}.$$

But working with this isn't actually necessary for computation. Instead,

$$\frac{d}{dt}(I + \frac{1}{1!}At + \frac{1}{2}A^2t^2 + \cdots) = 0 + \frac{1}{1!}A + \frac{2}{2!}A^2 t + \frac{3}{3!}A^3t^2 + \cdots = A(1 + \frac{1}{1!}At + \cdots) = Ae^{At}.$$

What's important here is that e^{At} solves the differential equation $\frac{d\psi}{dt} = A\psi$.

Definition 280

A **one-parameter group** in GL_n is a differentiable homomorphism

$$\mathbb{R}^+ \xrightarrow{\phi} GL_n$$

satisfying $\phi(s+t) = \phi(s)\phi(t)$. Its **derivative** is defined to be

$$\frac{d\phi}{dt} = \lim_{\Delta t \to 0} = \frac{\phi(t+\Delta t) - \phi(t)}{\Delta t} = \left.\frac{d\phi}{dt}\right|_{t=0} \phi$$

(just like above).

So any one-parameter group satisfies $\frac{d\phi}{dt} = A\phi$ for some matrix A. It turns out all one-parameter subgroups look like matrix exponentials:

Proposition 281

If $\frac{d\phi}{dt} = A\phi$, where ϕ is a one-parameter group, then $\phi(t) = e^{At}$.

Proof. Consider $e^{-At}\phi(t)$. By the product rule,

$$\begin{aligned}\frac{d}{dt}\left(e^{-At}\phi(t)\right) &= -Ae^{-At}\phi(t) + e^{-At}\frac{d\phi}{dt} \\ &= -Ae^{-At}\phi + e^{-At}(A\phi) \\ &= 0\end{aligned}$$

since A commutes with e^{-At} (which is just a sum of powers of A). Thus $e^{-At}\phi(t)$ must be a constant matrix, so

$$e^{-At}\phi(t) = C \implies \phi(t) = Ce^{At}.$$

Putting $t = 0$, $e^0\phi(0) = C$, but e^0 is the identity and $\phi(0)$ is as well (because we have a homomorphism). Thus $C = I$ and $\phi(t) = e^{At}$. □

Example 282

What are the one-parameter groups in the orthogonal group $O_n \subset GL_n$?

In other words, we want a differential homomorphism $\mathbb{R}^+ \to GL_N$ with image in O_n. Thus, we want $\phi = e^{At}$ to be orthogonal for all t, which means

$$(e^{At})^*(e^{At}) = I$$

for all t. Let's differentiate: $(e^{At})^* = e^{A^*t}$ (look at each term in the definition), so by the product rule, we have

$$(A^*e^{A^*t})(e^{At}) + (e^{A^*t})(Ae^{At}) = 0$$

which implies that we want $\boxed{A^* + A = 0}$. Plugging this back in, all such matrices work:

$$e^{-At}e^{At} = e^0 = I.$$

Corollary 283

The one-parameter groups in O_n are of the form e^{At} for a **skew-symmetric** matrix A.

In the $n = 2$ case, skew-symmetric matrices are multiples of $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Then

$$e^{At} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{1!}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} t + \frac{1}{2!}\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} t^2 + \frac{1}{3!}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} t^3 + \cdots = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix},$$

so the one-parameter groups are the obvious rotation matrices.

34 December 5, 2018

Recall that a **one-parameter group** in GL_n is a differentiable homomorphism $\mathbb{R}^+ \xrightarrow{\phi} GL_n$; we know that $\phi(t) = e^{At}$ must be a matrix exponential, where A is the derivative of the homomorphism at $t = 0$.

We found the one-parameter groups in $O_n \subset GL_n$ to be e^{At}, where $A^* = -A$ is skew-symmetric. We'll continue this kind of characterization now:

Example 284

What are the one-parameter groups in $SL_n \subset GL_n$?

We want A such that e^{At} has determinant 1 for all t. Luckily, we have a nice formula:

$$\det e^A = e^{\operatorname{tr} A}.$$

This is true since the eigenvalues of e^A are e^λ, where λ is an eigenvalue of A, and the determinant of a matrix is the product of its eigenvalues.

So if we want e^{At} to have determinant 1, and t is a scalar, we want

$$e^{t(\operatorname{tr} A)} = 1 \, \forall t.$$

This happens if and only if **the trace of A is** 0, and all such matrices correspond to one-parameter groups e^{At}.

For example for $n = 2$, we can have $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, which yields $e^{At} = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$. We can also have $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, which gives $e^{At} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$. $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ gives $\begin{bmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{bmatrix}$, and so on. (Keep in mind that the one-parameter groups always trace out a continuous path in our matrix space.)

Example 285

What are the 1-parameter groups in some subgroup of triangular matrices? Let's say, for example, that we want 1s on the diagonal and 0s below the diagonal.

Then $\frac{d}{dt}\left(e^{At}\right)$ has 0s on the diagonal, and evaluating

$$\frac{d}{dt}(e^{At}) = Ae^{At}$$

at $t = 0$, we want A to have zeros on and below the diagonal.

Example 286

What are the one-parameter groups in SU_2?

We're supposed to know that SU_2 is a 3-sphere in $\mathbb{R}^4$, and longitudes are circles of the form $\cos\theta I + \sin\theta A$ for some A with trace 0.

Well, now we're working in real numbers anymore. Let's say we want to look at $SU_2 \subset GL_2(\mathbb{C})$. What does it mean for such a map to be differentiable? We can think of functions of a complex variable as real and imaginary parts. In fact, the maps here are given by complex solutions to $\frac{d\phi}{dt} = A\phi$, and the same proof still shows that $\phi = e^{At}$ – nothing funny happens here.

We still want A to satisfy

$$(e^{At})^* = (e^{At})^{-1}, \det e^{At} = 1 \,\forall t.$$

The second condition means A has trace 0. Note that

$$(e^A)^* = I^* + \frac{1}{1!}A^* + \frac{1}{2!}(A^2)^* + \cdots$$

and since $(A^2)^* = (A^*)^2$, we see that this means

$$(e^{At})^* = e^{A^*t}.$$

(It is important that t **is real**, so it's equal to its complex conjugate and we don't have to worry about conjugating it here.) So we want

$$e^{A^*t} = (e^{At})^{-1} \implies e^{A^*t}e^{At} = 1.$$

for all t. Differentiating, by the product rule,

$$\left(A^* e^{A^*t} e^{At}\right) + \left(e^{A^*t} A e^{At}\right) = 0.$$

Putting $t = 0$, $A^* + A = 0$, so A should be **skew-Hermitian** and have **trace** 0. So the one-parameter groups are actually tracing out the longitudes in SU_2, since a basis for these matrices is

$$\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}.$$

These correspond to the one-parameter groups $\begin{bmatrix} e^{it} & 0 \\ 0 & e^{-it} \end{bmatrix}$, $\begin{bmatrix} c & s \\ -s & c \end{bmatrix}$, and $\begin{bmatrix} c & is \\ is & c \end{bmatrix}$.

Let's go back to looking in SL_2. I and $-I$ are in every 1-parameter group, but the rest of the longitudes partition SU_2. So there's always a path to any matrix in SU_2: can we always get to a given matrix P in SL_2 along 1-parameter groups? (Some confusion came up during class, but the answer was discussed next class.)

35 December 10, 2018

Definition 287

The **Lie algebra** of a matrix group G is the space of **tangent vectors** to G at I.

Let's try to define this based off the calculus path tangent vector. If we have a path $x(t)$ and $p = x(t_0)$, then the derivative $v = \frac{dx}{dt}\big|_{t=t_0}$. We'll make an analogous definition for matrices:

Definition 288

Let $x(t)$ be a matrix-valued function. Then the **derivative** $\frac{dx}{dt}$ is also a matrix

$$\lim_{\Delta t \to 0} \frac{x(t+\Delta t) - x(t)}{\Delta t}$$

Note that we just take the derivatives of the individual matrix entries with respect to t.

Example 289

What is the Lie algebra of SU_2?

A vector (aka matrix) A is tangent to SU_2 at I if there exists a path $x(t)$ with the following properties:

- It exists for some interval $-r < t < r$.
- $x(t) \in SU_2$ for all t, and $x(0) = I$.
- x is differentiable, and $\frac{dx}{dt}$ at $t = 0$ evaluates to A.

Well, for $x(t) \in SU_2$ to be true, we must have $x(t)^* x(t) = I$ and $\det x(t) = 1$ for all t. Let's take the **derivatives** of those two statements. Note that $\left(\frac{dx}{dt}\right)^* = \frac{dx^*}{dt}$, so by the product rule, differentiating $x^* x = I$ yields

$$\frac{dx^*}{dt} x + x^* \frac{dx}{dt} = 0.$$

Putting in $t = 0$, we have $x(0) = x^*(0) = I$, and the derivatives become A^* and A respectively. Thus, $A^* + A = 0$, and A must be **skew-Hermitian**.

Next, if the determinant of $x(t)$ is 1, then the derivative of $\det x(t)$ is 0. Thus the derivative of $x_{11}x_{22} - x_{12}x_{21}$ is

$$x_{11}' x_{22} + x_{11} x_{22}' - x_{12}' x_{21} - x_{12} x_{21}' = 0.$$

Plugging in $t = 0$, since $x(0) = I$, $x_{11} = x_{22} = 1$ and $x_{12} = x_{21} = 0$, and we have

$$x_{11}'(0) + x_{22}'(0) = 0.$$

Thus the **trace of A is** 0 as well.

Remark 290. *We could have guessed this because we know some paths already: $x(t) = e^{At}$ is a path in GL_2 where the derivative of x at $t = 0$ is A, and it's in SU_2 if $A^* = -A$ and the trace of A is* 0.

Corollary 291

A is in the Lie algebra of SU_2 if and only if e^{At} is a one-parameter group.

This actually turns out to be true when we replace SU_2 with any matrix group!

By the way, we'll quickly address the question from last class:

Fact 292

Matrices in $SL_2(\mathbb{R})$ that can be obtained along a 1-parameter group are those where the eigenvalues of P are positive and real or complex, but not negative real except for the matrix $-I$.

Let L be a Lie algebra (let's say of SU_2 for now).

Definition 293
The Lie algebra is not a group, but it has a **bracket operation** which is a law of composition sending $A, B \in L$ to $[A, B] = AB - BA$.

Let's check that this is still in the Lie algebra for SU_2. We know that

$$(AB - BA)^* = B^*A^* - A^*B^* = (-B)(-A) - (-A)(-B) = BA - AB = -(AB - BA),$$

so $AB - BA$ is skew-Hermitian. Then the trace of BA is (preserved under conjugation) equal to the trace of $B^{-1}BAB$, which is AB. So the difference of AB and BA has trace 0, which is what we want! So the bracket operation indeed keeps us in the Lie algebra L.

Proposition 294
Lie algebras satisfy the **Jacobi identity**

$$[[A, B], C] + [[B, C], A] + [[C, A], B] = 0.$$

Let's check this: it's equal to

$$\sum_{\text{cyc}} [(AB - BA)C - C(AB - BA)] = \sum_{\text{cyc}} [ABC - BAC - CAB + CBA]$$

which is clearly 0. It turns out the bracket acts a lot like the **commutator**: let's consider $xyx^{-1}y^{-1}$, and write $x = 1 + a$ and $y = 1 + b$ such that a^2, b^2 are small enough to be neglected, but not ab. Then the commutator is

$$(1 + a)(1 + b)(1 - a)(1 - b) = (1 + a + b + ab)(1 - a - b + ab) \cdots = 1 + ab - ba = 1 + [a, b].$$

Definition 295 (Actual definition)
A real vector space L is a **Lie algebra** if it has a bracket operation that is bilinear and skew-symmetric ($[A, B] = -[B, A]$) satisfying the Jacobi identity.

Example 296
Take $\mathbb{R}^3$ with the cross-product $[a, b] = a \times b$. It can be verified that this indeed satisfies the Jacobi identity!

36 December 12, 2018

Recall this concept from earlier in the class:

Definition 297
A group G is **simple** if it has no proper normal subgroup.

We know that the icosahedral group is simple, and that alternating groups A_n for $n \geq 5$ are all simple. Recall that a **normal subgroup** is a group that is closed under multiplication, inverses, and conjugation by elements of G. In particular, if $x \in N, g \in G$, then $gxg^{-1} \in N$. It's also good to recall the **commutator** $gxg^{-1}x^{-1}$ for $x \in N, g \in G$. This is also in N, and intuitively this gives a lot of elements if G is "very noncommutative."

Proposition 298

SO_3, the group of orthogonal matrices with determinant 1, is simple.

Proof. It is easier to look at SU_2. There exists a spin homomorphism $SU_2 \to SO_3$, which tells us how a given element of SU_2 acts on the equator. The kernel is $\pm I$ (since those are the only elements of SU_2 that fix the equator when we conjugate by them).

Take a normal subgroup of SO_3. Then this maps to a subgroup of SU_2, so the analogous statement for SU_2 is that the only proper normal subgroup of SU_2 is $\{\pm I\}$.

To prove this, let's say N is normal and contains some $Q \neq \pm I$. We'll show that $N = SU_2$. The entire conjugacy class of Q, which is the latitude of matrices that have the same trace as Q, is in N. This is a subset of N, and the set of commutators of the form $PQP^{-1}Q^{-1}$, where Q is given and P is arbitrary, is also some subset of N. But multiplication by Q^{-1} is continuous, so this also contains some points arbitrarily close to the identity and therefore all the conjugacy classes of points near the identity. One way to understand this is to draw a little path starting from Q in the conjugacy class $\{PQP^{-1}\}$. After moving to the other sphere, we get a little path starting from I that is in N, which gives lots of things that are small distances away from the identity.

So now just color in the whole sphere: given a longitude of the form $\{\cos\theta I + \sin\theta A\}$ for some A on the equator, we know that small θ values are in N. But $\cos\theta I + \sin\theta A = e^{A\theta}$, since $A^2 = -I$ for all matrices on the equator! So now we get the whole longitude by taking powers of $e^{A\theta}$ for some small θ. □

The next result, similar to the one we've just proved, is that the only proper normal subgroup of SL_2 is $\{\pm I\}$. In fact, we can replace $\mathbb{R}$ with any field:

Theorem 299

Let F be a field with at least 4 elements. Then the only proper normal subgroup of $SL_2(F)$ is $\{\pm I\}$.

Modding out by $\{\pm I\}$ is called $PSL_2(F)$ for some reason. One thing to mention: finite fields have order p^n for prime p, and if $p = 2$, then $I = -I$. But that's not too important right now. For example, say F is a finite field of order p. Then $SL_2(F)$ has order $\frac{(q^2-1)(q^2-q)}{q-1} = (q-1)(q)(q+1)$. For example, if $|F| = 5, 7, 11$, then $|SL_2(F)|$ has order $120, 336, 1320$, and $|PSL_2(F)|$ has order $60, 168, 660$.

Proof of theorem. Let N be a normal subgroup of $SL_2 = SL_2(F)$, and let's say it contains some matrix $A \neq \pm I$. Our goal is to show $N = SL_2$.

First of all, $\left\{ \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix} \right\}$ generate SL_2, so it suffices to show that N contains them. Let $s \in F$, where $s \neq \pm 1$ (so $s \neq s^{-1}$). The matrices with eigenvalues s, s^{-1} form a conjugacy class in SL_2.

Claim 300. *Suppose a matrix Q has eigenvalues s, s^{-1}, and $s \neq \pm 1$. Then the two eigenvectors v_1 and v_2 (for those eigenvalues) aren't multiples of each other (since $s \neq s^{-1}$), so they form a basis.*

Let B be the matrix with eigenvectors v_1, v_2. We can scale these eigenvectors by scalars so that B has determinant 1 (and is therefore in SL_2). Now $Be_i = v_i$ for our standard basis vectors e_i; let $P = \begin{bmatrix} s & 0 \\ 0 & s^{-1} \end{bmatrix}$. Then BPB^{-1} sends v_1 to $BPe_1 = sBe_1 = sv_1$, and similarly $BPB^{-1}v_2 = s^{-1}v_2$. So $BPB^{-1} = Q$, and therefore Q is conjugate to $\begin{bmatrix} s & 0 \\ 0 & s^{-1} \end{bmatrix}$.

Claim 301. *If N contains some matrix with eigenvalues s, s^{-1}, then $N = SL_2$.*

This is because N contains both $\begin{bmatrix} s^{-1} & 0 \\ 0 & s \end{bmatrix}$ and $\begin{bmatrix} s & sx \\ 0 & s^{-1} \end{bmatrix}$, so it contains their product, which is $\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}$. We can get the lower triangular version too (with an analogous argument), and now we can generate all elements of SL_2, meaning that $N = SL_2$.

So now we're ready for the actual proof. Let's say N contains some element $A \neq \pm I$. Choose a vector v_1 that is not an eigenvector; then $v_2 = Av_1$ is not a multiple of v_1, so we have a basis (v_1, v_2).

Choose P such that v_1, v_2 are eigenvectors with eigenvalues r, r^{-1}. Then the commutator $C = PAP^{-1}A^{-1}$ is an element of N, so

$$Cv_2 = PAP^{-1}A^{-1}v_2 = PAP^{-1}v_1 = r^{-1}PAv_1 = r^{-1}Pv_2 = r^{-2}v_2$$

so C is an element of N with one eigenvalue r^{-2}. (The other eigenvalue is therefore r^2.) The only thing that remains is to show that there exists r whose square is not ± 1. There are at most 4 solutions in a field for $r^2 = \pm 1$, so if the field has more than 5 elements, this works. So we now just need to think about $p = 5$, but luckily that also works! $\square$

See 18.702 for the continuation of this class.