Building New Fields

Last semester we considered how to factor polynomials over particular fields, such as $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$. We saw that some polynomials which are irreducible over $\mathbb{Q}$ can be factored if we move to $\mathbb{R}$, for example. But this is overkill - $\mathbb{R}$ contains numbers like π which we never need to help factor polynomials over $\mathbb{Q}$. All we really need to do is add a few new elements to our field to ensure we can factor a given polynomial. This builds a new field. Today we will begin to explore this idea.

- Let's start with an example. Consider the polynomial $a(x) = x^3 + 3x^2 - x + 2$ in $\mathbb{Q}[x]$. By the rational root theorem, if $a(x)$ has a root, it must be ± 1 or ± 2, but none of these are roots, so there are no roots in $\mathbb{Q}$. Since $a(x)$ has degree 3, this means that $a(x)$ is irreducible over $\mathbb{Q}$.
- But of course there are roots... somewhere. Let's make up a new number, call it $\beth$. What properties should $\beth$ have?
- Well, it should be a root of our particular $a(x)$. That is $\beth^3 + 3\beth^2 - \beth + 2 = 0$.
- Now we want to *extend* the field $\mathbb{Q}$ to a larger field which contains $\beth$. Let's call this new field $\mathbb{Q}(\beth)$.
- What does $\mathbb{Q}(\beth)$ look like? Well it contains all the rational number, plus also $\beth$. What about $1 + \beth$? Sure. What else?
- In fact, we can get any $a\beth^2 + b\beth + c$ where $a, b, c \in \mathbb{Q}$. Why don't we have $\beth^3$ in the field?
- Well we do of course, but $\beth^3 = -3\beth^2 + \beth - 2$, which we already counted. Similarly, any larger power of $\beth$ is already in there.
- What about inverses? Well additive inverse are easy: $-\beth = -1\beth$. What about $\beth^{-1}$?
- We want $\beth^{-1}\beth = 1$. Maybe $\beth^{-1} = a\beth^2 + b\beth + c$. We can actually solve for a, b, and c:

$$\beth^{-1} = -\frac{1}{2}\beth^2 - \frac{3}{2}\beth + \frac{1}{2}$$

  (try multiplying it out). We got this by solving for a, b, and c in $\beth(a\beth^2 + b\beth + c) = 1$, or equivalently that $(-3a+b)\beth^2 + (a+c)\beth - 2a = 1$. We need to be able to deduce that $(-3a + b) = 0$ and $a + c = 0$ - we know this because $a(x)$ is the *minimum* polynomial for $\beth$.
- Sweet, so if we have $\beth^{-1}$, do we have inverses of every element of the form $a\beth^2 + b\beth + c$? This is not obvious, but in fact we do.
- Slow down: why are we doing this? We wanted to be able to factor $a(x) = x^3 + 3x^2 - x + 2$. Can we factor this in $\mathbb{Q}(\beth)$? Try dividing $a(x)$ by $x - \beth$. We will get a remainder of 0. (The quotient will be $x^2 + (3 + \beth)x - 1 + 3\beth + \beth^2$.)

- What can we say about $\mathbb{Q}(\aleph)$? We claim this is a field, but inverses are a little tricky. Is it a field we have seen before? How can we get our hands on it.
- Maybe let's start from the other side. We know that $\aleph$ is a root of $a(x)$. Let's work with $a(x)$ in $\mathbb{Q}[x]$. In fact, consider $\mathbb{Q}[x]/\langle a(x)\rangle$.
- What does this set look like? It is a bunch of cosets. One coset is $\langle a(x)\rangle + x^5 + 7x^2 + 3$. What else is in this coset? What is the smallest degree polynomial in this coset?
- Why does it never make sense to "name" a coset with an x^3 in it? We can always replace that x^3 with a smaller degree, using $a(x)$. Similar to how we never needed a $\aleph^3$ in any element of $\mathbb{Q}(\aleph)$.
- In fact, the more we work with it, the more $\mathbb{Q}[x]/\langle a(x)\rangle$ looks just like $\mathbb{Q}(\aleph)$. Could these be isomorphic?
- How does that correspondence go? Which element in $\mathbb{Q}(\aleph)$ should correspond to $\langle a(x)\rangle 3x^2 + 7x - 1$? Which element in $\mathbb{Q}[x]/\langle a(x)\rangle$ should correspond to $\aleph^2 + 4\aleph + 2$?
- This seems quite straight forward. We match up polynomials $b(x) \in \mathbb{Q}[x]$ with the element $b(\aleph)$ in $\mathbb{Q}(\aleph)$. But that's not quite right. We want to match up cosets, not polynomials. Also, there are "more" polynomials than elements in $\mathbb{Q}(\aleph)$ (such as all the degree 3, 4, 5... ones).
- This is the magic of quotient groups. We define a homomorphism $\varphi : \mathbb{Q}[x] \to \mathbb{Q}(\aleph)$ by $\varphi(b(x)) = b(\aleph)$. But then we "reduce" $b(\aleph)$ using $a(x) = 0$. The function φ is not injective. So we mod out by the kernel.
- The kernel of φ is the set of polynomials which when evaluated at $\aleph$ give zero. But $\aleph$ is a root of $a(x)$, so this will be all multiples of $a(x)$. That is, the kernel is the ideal $\langle a(x)\rangle$.
- Now that we have an isomorphism between $\mathbb{Q}[x]/\langle a(x)\rangle$ and $\mathbb{Q}(\aleph)$, we can transfer facts about one to the other. We were not quite sure that $\mathbb{Q}(\aleph)$ was a field. But we can fairly easily prove that $\mathbb{Q}[x]/\langle a(x)\rangle$ is a field. Here's how:
- To show this is a field, we need to prove that every non-zero element has an inverse. Take $\langle a(x)\rangle + b(x) \neq 0$. This means $b(x) \notin \langle a(x)\rangle$. But since $a(x)$ is irreducible, this means $\gcd(a(x), b(x)) = 1$.
- Then by Bezout's lemma, we get polynomials $s(x)$ and $t(x)$ such that $1 = s(x)a(x) + t(x)b(x)$. This then says that $1 \in \langle a(x)\rangle + t(x)b(x) = (\langle a(x)\rangle + t(x))(\langle a(x)\rangle + b(x))$. But 1 is a polynomial so is in $\langle a(x)\rangle + 1$. This tells us that $\langle a(x)\rangle + t(x)$ is the inverse of $\langle a(x)\rangle + b(x)$.
- Note also that by using the Euclidean Algorithm backwards, we can find $t(x)$. This is probably easier than doing the system of equations to find inverse. Of course we then need to apply our isomorphism to $t(x)$, but that is easy.

Review of Ideals of Polynomials

Our goal is to understand *extension fields*, which last time we saw could be interpreted as quotient rings. To do this, we need to make sure we understand quotient rings, especially those of rings of polynomials.

- To set the stage, remember we are considering E extending a field F.
- For some element $c \in E$, consider the *evaluation function* $\sigma_c : F[x] \to E$ defined by $\sigma_c(p(x)) = p(c)$. That is, σ_c takes a polynomial over F and evaluates it at the number c, to get a number in E.
- σ_c is a surjective homomorphism! The kernel is an ideal of $F[x]$. So what ideal? Call it J_c.
- What do ideals in $F[x]$ look like? The are sets of polynomials, closed under addition, subtraction and multiplication (they are subrings) which absorb products.
- For example, suppose an ideal J in $\mathbb{Q}[x]$ contained the polynomials $x^2 - x - 6$ and also $x^2 - 4$. What else must it contain? In particular, will it contain any polynomials of degree less than 2? Will it contain any constants?
- If we get stuck: back in $\mathbb{Z}$, consider an ideal that contains 36 and also 40. What else must it contain? Is there a single element that generates all of the ideal?
- It turns out that in $\mathbb{Z}$ and in $F[x]$ (with F a field), every ideal is *principle*. That is, each ideal is generated by a single element.
- Let's prove this. In $F[x]$, with ideal J, let $p(x)$ be any polynomial in J of smallest degree. We could take $p(x)$ to be monic (divide out by the leading term).
- We claim that every element of J is a multiple of $p(x)$. So let $a(x) \in J$ be such an element. How can we decide whether $a(x)$ is a multiple of $p(x)$? Well, try dividing.
- Remember the division algorithm? When we try to divide $a(x)$ by $p(x)$ we get $a(x) = q(x)p(x) + r(x)$ for some polynomials $q(x)$ and $r(x)$ with $r(x) = 0$ or else $\deg(r(x)) < \deg(p(x))$. Which is it here?
- Well, what can we say about the location of $r(x)$? Since $r(x) = a(x) - q(x)p(x)$ we see that $r(x) \in J$. But there is no polynomial in J of degree smaller than $p(x)$, so it must be that $r(x) = 0$.
- This is really amazing. Find any polynomial in J of minimum degree, and that generates all of J. Since we can divide by the leading coefficient, this says that there is always one unique monic "minimal" polynomial that generates J.
- Now what if we start with a number $c \in E$? If c is the root of some polynomial in $F[x]$, we say that c is *algebraic over* F, otherwise we say that c is *transcendental*. Consider just the algebraic case. There are lots of polynomials that have c as a root then. In fact, the set of all such polynomials is exactly J_c.

- Let $p(x)$ be the monic generating polynomial for J_c. Then $p(x)$ must be irreducible. Why?
- We call $p(x)$ the *minimum polynomial of c over F*.
- Now what is special about ideals generated by irreducible polynomials? Remember, irreducible in $F[x]$ is the analogous notion to *prime* in $\mathbb{Z}$.
- Some different sorts of ideals: An ideal J in a ring A is *prime* provided for any two elements $a, b \in A$, if $ab \in J$ then $a \in J$ or $b \in J$.
- A proper ideal J in a ring A is *maximal* provided every ideal K containing J (so $J \subseteq K \subseteq A$) we have $J = K$ or $K = A$. That is, there are no proper ideals of A strictly larger than J.
- In $F[x]$, what can we say about ideal of the form $\langle p(x) \rangle$? What if $p(x)$ is irreducible? Explain why for $p(x)$ irreducible, $\langle p(x) \rangle$ is both prime and in fact maximal.
- It turns out that the reason we care about prime and maximal ideals is that their quotient rings have nice properties. J is prime iff A/J is an integral domain. J is maximal iff A/J is a field. You will explore this more in your homework.
- Notice this tells us that $F[x]/\langle p(x) \rangle$ is a field as long as $p(x)$ is the minimum polynomial for c over F.
- The Fundamental Homomorphism Theorem tells us that the range of σ_c is a subfield of E. What does it look like? Well it is just $\{a(c) \ : \ a(x) \in F[x]\}$
- It is also the *smallest* field containing F and c. This is because any field containing F and C would also contain elements of the form $a_0 + a_1 c + a_2 c^2 + a_3 c^3 + \cdots + a_n c^n$ with the $a_i \in F$, and this is exactly what you get when you plug in c to polynomials in $F[x]$.
- For notation, we write the range of σ_c as $F(c)$.
- Note: if d is also a root of $p(x)$ (the minimum polynomial of c over F) then $F(d) \cong F(c)$. For example $F(\sqrt{3}) \cong F(-\sqrt{3})$.

Start by finishing up what we were doing last time.

- We had that $F[x]/\langle p(x)\rangle$ was a field as long as $p(x)$ is irreducible.
- In fact, if $p(x)$ is the minimum polynomial for a number $c \in E$, then $p(x)$ will definitely be irreducible and what is more, by the Fundamental Homomorphism Theorem, $F[x]/\langle p(x)\rangle \cong E$. So E is a field.
- What does it look like? E is the range of σ_c. It is just $\{a(c) \ : \ a(x) \in F[x]\}$.
- It is also the *smallest* field containing F and c. This is because any field containing F and c would also contain elements of the form $a_0 + a_1c + a_2c^2 + a_3c^3 + \cdots + a_nc^n$ with the $a_i \in F$, and this is exactly what you get when you plug in c to polynomials in $F[x]$.
- For notation, we write the range of σ_c as $F(c)$.
- Note: if d is also a root of $p(x)$ (the minimum polynomial of c over F) then $F(d) \cong F(c)$. For example $F(\sqrt{3}) \cong F(-\sqrt{3})$.

Now the way we know that E is a field is by using a more general result that if you mod out by a maximal ideal, the quotient ring is a field. This is interesting in its own right (and you are asked to prove it in homework), but in this particular case, we can actually see that the quotient ring is a field directly. Do that by going through Activity 2. This would only work in integral domains in which you can do the Euclidean algorithm. Such integral domains are called *Euclidean domains*. These are the nice integral domains: each is a unique factorization domain, and each is a principle ideal domain (although the converses are both fals.)

New Fields as Quotient Rings

A quick review of what we have done so far in building extension fields.

- Given a field F and a larger (extension) field E, we can either:
- ...start with an element $c \in E$ and build $F(c)$, the smallest field containing F and c, or,
- ...we can start with an irreducible polynomial $p(x) \in F[x]$ and consider the quotient ring $F[x]/\langle p(x)\rangle$.
- These are the same provided c is a root of $p(x)$ and $p(x)$ is the least degree polynomial which has c as a root.
- Up until now we started with $c \in E$ and found $p(x)$ by considering the *evaluation homomorphism* $\sigma_c : F[x] \to E$ given by $\sigma_c(f(x)) = f(c)$. The kernel of σ_c was $\langle p(x)\rangle$ where $p(x)$ was the *minimum polynomial of c over F*.
- Last time we started going the other way: if you start with just F and consider an irreducible polynomial $p(x) \in F[x]$, we would like to build an extension field in which $p(x)$ has a root.

Theorem. *Let F be a field and $p(x)$ a nonconstant polynomial in $F[x]$. There exists an extension field E of F and an element c in E such that c is a root of $p(x)$.*

Proof. We can assume $p(x)$ is irreducible, otherwise factor it and prove the theorem for one of the irreducible factors.

Since $p(x)$ is irreducible, $\langle p(x)\rangle$ is maximal. And if $\langle p(x)\rangle$ is maximal, then $F[x]/\langle p(x)\rangle$ is a field.

Write $J = \langle p(x)\rangle$. Is $F[x]/J$ an extension of F? Not really. $F[x]/J$ contains cosets of J. But we can identify $a \in F$ with the coset $J + a$. Think of picking representatives ($a \in J + a$ is the representative for the whole coset). More precisely, define $h : F \to F[x]/J$ by $h(a) = J + a$. It is clearly a homomorphism, and must be injective because F is a field. The range of h is the set of all cosets $J + a$ where a is a constant polynomial. So F is isomorphic to this subfield of $F[x]/J$.

Now another element of $F[x]/J$ is $J + x$. We will show this is a root of $p(x) = a_0 + a_1x + \cdots + a_nx^n$. But to make this make sense, we need to work in the range of h.

$$(J + a_0) + (J + a_1)(J + x) + (J + a_2)(J + x)^2 + \cdots + (J + a_n)(J + x)^n = J + p(x) = J$$

□

- Note that while it looks strange, what we are doing here is nothing new. Compare to what we really do when we build $\mathbb{Z}_n \cong \mathbb{Z}/n\mathbb{Z}$.
- $\mathbb{Z}/n\mathbb{Z}$ is really a ring of cosets, but we pick representatives from each coset and call the coset by the representative's name.

- Now, consider another theorem that seems very similar to the above.

Theorem. *Let E be a field extension of F and $\alpha \in E$ be algebraic over F. Then $F(\alpha) \cong F[x]/\langle p(x)\rangle$, where $p(x)$ is the minimal polynomial of α over F.*

- This theorem captures the "other direction" in that we are now starting with an element in a field and saying something about the smallest field extending F that contains it.
- What is more, this allows us to say something about the structure of $F(\alpha)$, and to compare it to $F(\beta)$ for other algebraic elements. **Ex:** What does $\mathbb{Q}(\sqrt[3]{2})$ look like? **Ex:** What does $\mathbb{Q}(\sqrt{2+\sqrt{3}})$ look like?
- One interesting consequence: if α and β are two different roots of the same irreducible polynomial, then the theorem above says that $F(\alpha) \cong F(\beta)$. This is not surprising for polynomial like $x^2 - 3$, but for more complicated ones, it could be that the extensions are not equal, even if they are isomorphic. **Ex:** How does $\mathbb{Q}(\sqrt{2})$ relate to $\mathbb{Q}(\sqrt{3})$? What about to $\mathbb{Q}(\sqrt{2}+1)$?

Ruler and Compass Constructions

Introduction

One of the themes this semester is to see how abstract algebra can help us solve questions outside of abstract algebra. You might call these applications, but note we often apply to other areas of math instead of the "real world."

Looking at these applications has two purposes. First, it is nice to see how to solve these problems in other areas of math. However, the more relevant reason for us is it helps motivate us to learn new things about algebra. This will be the case with our first example: geometry.

- This is not "modern" geometry; quite the opposite. We take our lead from the Greeks of antiquity. For them, doing geometry meant *actually doing geometry*. For example, that a line through a given point, parallel to a given line, existed meant you should be able to built it. The tools they thought reasonable were a straight edge (unmarked) and a compass (which could hold its shape).

- Grab a pair of these tools and some papyrus. Play. What can you *construct*? What geometric figures *exist*?

- If you get frustrated with these ancient tools, boot up GeoGebra. Of course this has more tools than straight-edge and compass. Thus, make sure you only use the "new point" tool (to place points at the intersections of lines, circles, or both), the "line through two points" tool, and the "compass" tool (under the circle menu). You can also use the arrow to drag things around if you need to.

- Some challenges: can you construct a square? If you can construct a square, can you construct a square of twice the size? Three times the size?

- Can you construct a 60 degree angle? A 30 degree angle? What if you have an angle constructed: can you construct an angle twice the size? Half the size? A third of the size?

- Suppose you constructed a line segment which you think of as the side length of a cube. Can you construct a line segment to be the side length of a cube twice the volume?

- You can obviously construct a circle. Can you construct a circle of twice the area? Can you construct a square with the same area as the circle?

Constructible Numbers

Last time we considered what geometric shapes we could or could not construct. We were left with three big questions: is it possible to *trisect an angle*, to *double a cube*, or to *square a circle*. To answer these questions, we must "algebratize" algebraic constructions.

Start with two constructible points 0 and 1 *one unit* apart. We define *constructible* recursively from this base case:

- A *constructible line* is a line passing through two constructible points.
- A *constructible circle* is a circle whose radius is a constructible number and whose center is a constructible point.
- A *constructible point* is the intersection of two constructible lines, two constructible circles, or a constructible line and a constructible circle.

We say a number a is *constructible* provided $a = 0$ or there are two constructible points distance $|a|$ apart. So far we have constructible numbers 0, 1, and -1. What else is constructible?

For this activity, use GeoGebra. Use only the "new point" tool (to place points at the intersections of lines, circles, or both), the "line through two points" tool, and the "compass" tool (under the circle menu). You can also use the arrow to drag things around if you need to.

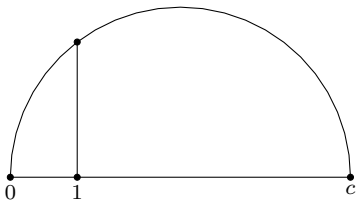
1. Show that the numbers 2 and 4 are constructible. Then show that the number $3 = 4-1$ is constructible.

2. If a and b are constructible numbers, are the numbers $a+b$ and $a-b$ also constructible?

3. Suppose a and b are constructible. Construct a triangle containing a base of unit length adjacent to a side of length a. Construct a similar triangle with where the side corresponding the unit length side now has length b. What is the length of the side corresponding to the a-length side?

4. Explain how you can modify the above construction to prove that if a and b are constructible, then a/b is constructible.

5. Given constructible number c, explain how you can construct the figure below. The vertical line should be perpendicular to the horizontal line, which is the diameter of the circle.



   What is the length of the vertical line?

6. Let $\mathfrak{C}$ be the set of all constructible numbers. What sort of set is this? Is it a group? A ring? A field? Is it one of these we know about already?

Constructible Extensions

Last week you saw that the field of constructible numbers is an field, extending $\mathbb{Q}$ and closed under taking square roots of positive elements. Today we will consider the converse: are there any other constructible numbers other than those you can get from $\mathbb{Q}$ using field operations and square roots.

- Let's think about constructible points (a, b) in the plane $\mathbb{R}^2$. Certainly if a and b are constructible numbers (as defined previously) then (a, b) a constructible point (by using perpendicular lines).
- Also, if (a, b) is constructible, so is $(0, b)$ and $(a, 0)$.
- We already know that all rational numbers are constructible, so we now have all points in $\mathbb{Q} \times \mathbb{Q}$ constructible.
- Consider the field $K_1 = \mathbb{Q}(a, b)$ where the point (a, b) was constructed in one step from $\mathbb{Q} \times \mathbb{Q}$. Then $K_2 = K_1(c, d)$ where the point (c, d) was constructed in one step from points in $K_1 \times K_1$. And so on.
- If a point is to be constructible, then it will need to be constructed in some finite number of steps starting from $\mathbb{Q} \times \mathbb{Q}$.
- Recall that a point is constructible provided it is at the intersection of lines and circles, each of which are constructible.
- Consider the case for the intersection of two lines. Let L_1 pass through (a_1, b_1), and (c_1, d_1) and L_2 pass through (a_2, b_2) and (c_2, d_2). Can we find the coordinates of the intersection of these two lines?
- Yes, doing so requires solving a system of two linear equation. In particular, the new x coordinate is a linear combination of $a_1, a_2, b_1, \ldots, d_2$.
- So x is the root of a degree 1 polynomial over K_i, so $[K_{i+1} : K_i] = 1$. Similarly for y.
- What about the intersection of a line and a circle? Recall that we can write the equation of a circle as $(x - a)^2 + (y - b)^2 = k^2$. Substitute in the equation for the line to get a quadratic polynomial in x (or y).
- Finally, the intersection of two circles. Use the equation for a circle in the form

  $$x^2 + y^2 + dx + ey + f = 0$$

  and take the difference of the two equations to get a linear expression involving x and y which can be substituted into one of the equations for the circle as above.
- What does all this tell us? Well, if we extend $\mathbb{Q}$ to get a field containing the coordinates of a constructible point, then we must have extended *only* by taking square roots of elements. Of course, not necessarily elements in $\mathbb{Q}$, but definitely elements in some extension field we got along the way.

- More specifically, we have that if F is some field containing constructible numbers, then the points determined by the intersections of lines and circles in F lie in the field $F(\sqrt{\alpha})$ for some $\alpha \in F$.
- Since this is recursive, we can say that a real number α is constructible if and only if there exists a sequence of fields

$$\mathbb{Q} = F_0 \subset F_1 \subset \cdots \subset F_k$$

such that $F_i = F_{i-1}(\sqrt{\alpha_i})$ with $\alpha_i \in F_i$ and $\alpha \in F_k$.

- This tells us exactly what sort of field extension any constructible number belongs to. It also says that the field of *all* constructible numbers is an infinite algebraic extension of $\mathbb{Q}$.
- But what about doubling the cube? Note that if we started with a cube of side length 1, we would get a cube of volume 2, so side length $\sqrt[3]{2}$. Is this a constructible number? If it is, then it is in a field extension of $\mathbb{Q}$ like those described above. So is it? Not obviously so, since it is not the square root of a number. But perhaps we can get the $\sqrt[3]{2}$ in a field extension you make by just adding square roots.
- To convince ourselves this cannot happen, we need to consider the size or *degree* of a field extension and what that tells us about the form of its elements. Next time.

Review of Vector Spaces

Let's review some basic ideas from linear algebra.

- When you first thought about vectors, you worked with a specific example: vectors in $\mathbb{R}^n$. These were columns of numbers which we could add together, as well as multiply by constants.
- In general, a vector space is a collection of objects called *vectors* together with some field of *scalars* and two operations: addition on the vectors and scalar multiplication (allowing vectors to be multiplied by scalars).
- Just like with groups and rings, there are some axioms that a vector space must adhere to which say how addition and scalar multiplication work. For one, the vectors under addition form an abelian group. Also, scalar multiplication works like you expect (in terms of distributive properties and associativity).
- We are particularly interested to two concepts: linear independence and span.
- A set of vectors is linearly *dependent* if one of them is a linear combination of the others (in other words, you can get one by adding scalar multiples of the others). Another way to say this is that you can get the zero vector as a non-trivial linear combination of the others. That is
$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$$
with not all of $a_1, a_2, \ldots, a_n$ equal to 0.
- Thus a set of vectors is linearly independent if the only way to get 0 as a linear combination is for all the scalars to be 0.
- If we can get a vector $\mathbf{v}_0$ as a linear combination of, say $\mathbf{v}_1$ and $\mathbf{v}_2$, then we say that $\mathbf{v}_0$ is in the *span* of $\mathbf{v}_1$ and $\mathbf{v}_2$.
- If we have a set of vectors so that every vector in the vector space is in their span, we say the set of vectors *spans* the vector space.
- Now think of starting with a single vector and building the largest possible set of linearly independent vectors. Eventually, you will get to a point where you can't add any more. This is because every vector not in your set is in the span of the set. This maximal set of linearly independent vectors must span the vector space.
- Alternatively, start with a large set of vectors which span the vector space. If any of these are in the span of the remaining vectors, we can get rid of it (we don't need it to span the vector space). So trim down the set until we get a minimal spanning set. This set must be linearly independent (otherwise we could get rid of the vector which was a linear combination of the others).
- The number of vectors in any maximal independent set will be equal to the number of vectors in all minimal spanning sets. Such sets are called *bases* for the vector space, and the number of vectors in the set is called the *dimension* of the vector space.

Degrees of Field Extensions

Why do we care about vector spaces? It turns out you can view a field extension as a vector space. This is useful because we can borrow the concept of linear independence, span and dimension and apply them to better understand field extensions.

- Let F be a field and K and extension field. We can view the elements of K as vectors, and the elements of F as scalars. This makes K into a vector space.
- Vector spaces have dimension (the number of vectors in any basis). We are most interested in when that dimension is finite, say n. In this case we say that K is an *extension of degree n* or that the *degree of K over F* is n and write

$$[K : F] = n.$$

- Now consider our favorite field extension $F(c)$ over F. Recall this is the smallest field containing both F and c. What is the degree of $F(c)$ over F?
- Suppose c is *algebraic* over F (that is, it is the root to some polynomial $p(x)$ in $F[x]$). Well let $p(x)$ be the minimum polynomial of c over F, say of degree n. Then we can write every element of $F(c)$ as

$$a_0 + a_1c + a_2c^2 + \cdots + a_{n-1}c^{n-1}$$

  where the a_i are in F. In other words, the set $\{1, c, c^2, \ldots, c^{n-1}\}$ span $F(c)$.
- Actually, how do we know this? Well let $a(c)$ be any element of $F(c)$. We know for sure that $a(x)$ is some polynomial (possibly of degree larger than n) because $F(c)$ contains c and is closed under addition and multiplication.
- Now use the division algorithm on $a(x)$ and $p(x)$. We get

$$a(x) = q(x)p(x) + r(x)$$

  where $\deg(r(x)) \leqslant n - 1$. So plug in c: we get $a(c) = r(c)$. Thus we only need to go up to a c^{n-1} term.
- So $\{1, c, c^2, \ldots, c^{n-1}\}$ spans $F(c)$. Is the set linearly independent? Well consider a linear combination equal to 0:

$$a_0 + a_1c + a_2c^2 + \cdots + a_{n-1}c^{n-1} = 0$$

  If any of the coefficients a_i were non-zero, we would have a polynomial of degree $n - 1$ for which c was a root. But $p(x)$ of degree n was the minimum polynomial, so this is impossible. Thus the set is indeed linearly independent.

- Putting these together, we see that $\{1, c, c^2, \ldots, c^{n-1}\}$ is a basis for $F(c)$, and thus $F(c)$ is a degree n extension of F.

  Ex: What is the degree of $\mathbb{C}$ over $\mathbb{R}$?

  Ex: What is the degree of $\mathbb{Q}(\sqrt{5})$ over $\mathbb{Q}$? Well the minimum polynomial for $\sqrt{5}$ is $x^2 - 5$ (why is this minimal?). Thus $[\mathbb{Q}(\sqrt{5}) : \mathbb{Q}] = 2$. Notice this confirms our suspicion that every element of $\mathbb{Q}(\sqrt{5})$ can be written as $a + b\sqrt{5}$ for rational a and b.

  Ex: What is the degree of $\mathbb{Q}(\sqrt{5}, \sqrt{7})$ over $\mathbb{Q}(\sqrt{5})$? What about over $\mathbb{Q}$? Well if we start with $\mathbb{Q}(\sqrt{5})$ as our base field, we must find the minimum polynomial for $\sqrt{7}$. First, could $\sqrt{7}$ be in $\mathbb{Q}(\sqrt{5})$? If it were, then $\sqrt{7} = a + b\sqrt{5}$. Square both sides and rearrange to get that $\sqrt{5}$ is rational (it is not). Notice that $\sqrt{7}$ is a root of $x^2 - 7$. This is irreducible over $\mathbb{Q}$, but because $\sqrt{7} \notin \mathbb{Q}(\sqrt{5})$, it is also irreducible in $\mathbb{Q}(\sqrt{5})$. Thus $\mathbb{Q}(\sqrt{5}, \sqrt{7})$ has degree 2 over $\mathbb{Q}(\sqrt{5})$. Similarly, it has degree 2 over $\mathbb{Q}(\sqrt{7})$. What about over $\mathbb{Q}$? It turns out it has degree 4 over $\mathbb{Q}$. Why?

Iterated Extensions

Last time we saw how to find the basis of a field extension. What if we take an extension of that extension? For example, consider $\mathbb{Q}(\sqrt{5}, \sqrt{7})$. How can we work with this?

- First, we need to understand how bases work. Suppose F is a field, and that K is an extension of F, and that E is an extension of K (so $F \subseteq K \subseteq E$).
- Suppose $a_1, a_2, \ldots, a_m$ is a basis for K over F. Suppose $b_1, b_2, \ldots, b_n$ is a basis for E over K. We want to find a basis for E over F.
- Well we know that any $c \in E$ can be written as $k_1b_1 + k_2b_2 + \cdots + k_nb_n$ where each k_i is a scaler from K.
- But since K is a vector space over the field F, we know that $k_i = l_{i1}a_1 + l_{i2}a_2 + \cdots + l_{im}a_m$ where l_{ij} come from F. Now substitute and distribute.
- We get $c = \sum l_{ij}a_ib_j$. So perhaps $\{a_ib_j\}$ (the set of all products of basis elements for K and E) form a basis for E over F. Certainly this is a spanning set. Is it linearly independent?
- Yes. If $\sum l_{ij}a_ib_j = 0$ then by collecting like terms, $\sum k_jb_j = 0$. Since $\{b_j\}$ is linearly independent, that means that each $k_j = 0$. But $k_j = \sum l_{ij}a_i = 0$ means that each $l_{ij} = 0$ since $\{a_i\}$ is linearly independent.
- Now a little counting. Since the basis for E over F contains all the products of basis elements for K over F and basis elements of E over K, we get

$$[E : F] = [E : K][K : F]$$

- This works as long as both extensions are finite. Certainly if $K = F(c)$ for some *algebraic* element c, then $[K : F]$ is finite. It is the degree of the minimum polynomial for c. Then if $E = K(d)$ for some element d which is algebraic over K, then $[E : K]$ is finite. Thus $E = K(d) = F(c, d)$ is a finite extension of F.
- If $E = F(c)$, then we say that E is a *simple* extension of F. If $E = F(c_1, c_2, \ldots, c_n)$ then E is an iterated extension.
- What the above result says is that if $c_1, c_2, \ldots, c_n$ are all algebraic over F, then E is an extension of finite degree of F. What about the converse of this statement? Suppose E is a finite extension of F. We claim that every element in E is algebraic over F.
- Here's why: If E is a finite extension of F, say with degree n, then there is a basis for E over F, say $\{a_1, a_2, \ldots, a_n\}$ (it has n elements in it). But then $E = F(a_1, \ldots, a_n)$.
- This says that E is an iterated extension of F, but in fact every element in E must be algebraic over F: for any element $c \in E$ we have that $\{1, c, c^2, \ldots, c^n\}$ is linearly dependent (it contains more than n elements). Thus there are elements $a_i \in F$ (not all zero) such that $a_0 + a_1c + \cdots + a_nc_n = 0$. But then c is a root to the polynomial $a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \in F[x]$.

Constructible Numbers again

- We have that if F is some field containing constructible numbers, then the points determined by the intersections of lines and circles in F lie in the field $F(\sqrt{\alpha})$ for some $\alpha \in F$.

- Since this is recursive, we can say that a real number α is constructible if and only if there exists a sequence of fields

$$\mathbb{Q} = F_0 \subset F_1 \subset \cdots \subset F_k$$

such that $F_i = F_{i-1}(\sqrt{\alpha_i})$ with $\alpha_i \in F_i$ and $\alpha \in F_k$.

- This tells us exactly what sort of field extension any constructible number belongs to. It also says that the field of *all* constructible numbers is an infinite algebraic extension of $\mathbb{Q}$.

- But what about doubling the cube? Note that if we started with a cube of side length 1, we would get a cube of volume 2, so side length $\sqrt[3]{2}$. Is this a constructible number? If it is, then it is in a field extension of $\mathbb{Q}$ like those described above. So is it? Not obviously so, since it is not the square root of a number. But perhaps we can get the $\sqrt[3]{2}$ in a field extension you make by just adding square roots.

- What does all this tell us? Well, if we extend $\mathbb{Q}$ to get a field containing the coordinates of a constructible point, then we can be sure that the degree of the field extension is a power of 2.

- So what? Well what if we could double the cube? This would mean we could start with a cube of side length 1, and get a cube with volume 2. But the side length of such a cube would be $\sqrt[3]{2}$, and $\sqrt[3]{2}$ belongs to a degree 3 field extension, so any field containing $\sqrt[3]{2}$ would need to have degree a multiple of 3.

- Trisecting the angle? Consider an attempt to trisect 60°. We would be able to construct 20° then, so surely we could also get $\cos(20^\circ)$ as a constructible number (how?). But

$$\cos(3\theta) = 4\cos^3(\theta) - 3\cos(\theta)$$

and since $\cos(60^\circ) = 1/2$, this says that $\cos(20^\circ)$ is a root of the polynomial $8x^3 - 6x - 1$ or that $2\cos(20^\circ)$ is the root of $x^3 - 3x - 1$. We can show this is irreducible using the rational roots theorem. So again, any extension containing a trisected 60° would have a degree that was a multiple of 3.

- Finally, if we could square the circle, then we would be able to construct a square whose side length was $\sqrt{\pi}$. But this is transcendental, so definitely would not give an extension whose degree was a power of 2.

Galois Theory

We have been investigating fields for the last few weeks. One application of this was to see that certain numbers (and shapes) are not *constructible*. It is surprising that we can make such a concrete connection between geometry and algebra (although *analytic* geometry is just this: describing geometric objects with equations and numbers).

This was not our primary reason to study fields though. So why are we doing all this in an algebra class? Why does UNC think you should learn this before teaching high school algebra? Well basic algebra is largely about solving equations, and when those equations involve polynomials of degree greater than 2, this becomes very challenging.

Our goal for the next few classes is to better understand the relationship between polynomials and field extensions. This will lead us to our final application: deciding when you can solve a polynomial using radicals.

Derivatives and Roots

- Recall what we know about extension fields and their relationship to minimum polynomials.
- If a polynomial $p(x)$ has degree n, what can you say about the number of roots it has (in a large enough extension field)?
- We know it must have at most n roots, but these might not be distinct. But what if $p(x)$ is irreducible?

  Theorem. *Let $p(x) \in F[x]$ be irreducible, and assume that F has characteristic 0. Then $p(x)$ does not have any repeated roots.*

  Proof. The cool thing about this proof is that it uses derivatives. But we define these purely formally. But still the usual way.

  Suppose $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ does have a repeated root (in some extension field), call it c. Then $p(x) = (x-c)^2q(x)$ for some polynomial $q(x)$. Now consider $p'(x)$:

  $$p'(x) = 2(x-c)q(x) + (x-c)^2q'(x) = (x-c)(2q(x0 + (x-c)q'(x))$$

  So $x - c$ is a factor of $p'(x)$ so c is a root.

  Since $p(x)$ is irreducible and has c as a root, $p(x)$ must be the minimum polynomial for c. This says that every polynomial which has c as a root is a multiple of $p(x)$. In other words, $p'(x)$ is a multiple of $p(x)$. But unless $p'(x) = 0$, this is impossible because $p'(x)$ has smaller degree than $p(x)$.

  So $p'(x) = na_nx^{n-1} + \cdots + a_1 = 0$ would imply that $na_n = 0$. This cannot happen in a field of characteristic 0 (although if we were in $\mathbb{Z}_5[x]$ this would be problematic).

  □

- From here on out, let's consider fields of characteristic 0 only (that is, extensions of $\mathbb{Q}$). Now irreducible polynomials have only distinct roots.
- Now do the field extension thing for these roots. What can happen?
- We can adjoin a root, and then in that bigger field, the polynomial factors. But does it factor into linear terms? Maybe, maybe not. If not, then we could find a root to one of the factors and adjoin that as well. Eventually, we will adjoin enough roots to get an extension field in which the polynomial factors completely (into linear factors).
- This extension field is call the *root field* or *splitting field* of the polynomial (because in that field, the polynomial "splits" into linear factors).

  Ex: What is the splitting field of $p(x) = x^3 + 2$? What is its degree over $\mathbb{Q}$?

  Ex: What is the degree of the splitting field of $p(x) = x^4 - 10x^2 + 25x - 5$? Well, $p(x)$ is irreducible, so there is at least one root α which we can adjoin to $\mathbb{Q}$ to get the degree 4 field extension $\mathbb{Q}(\alpha)$. Now what? From the polynomial's graph, we know that there are 2 real roots, so 2 complex roots. If α was one of the real roots, then we are not yet in the splitting field. So we need to add another root β. This might be of degree 2 or 3, depending on how $p(x)$ factored in $\mathbb{Q}(\alpha)$. It get's messy.
- Notice that if c and d are two distinct roots of $p(x)$, then $F(c) \cong F(d)$. Why?
- In fact, what can we say about isomorphisms between field extensions?

Let's pick up where we left off yesterday, with the theorem about repeated roots.

Theorem. *Let $p(x) \in F[x]$ be irreducible, and assume that F has characteristic 0. Then $p(x)$ does not have any repeated roots.*

Proof. The cool thing about this proof is that it uses derivatives. But we define these purely formally. But still the usual way.

Suppose $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ does have a repeated root (in some extension field), call it c. Then $p(x) = (x-c)^2q(x)$ for some polynomial $q(x)$. Now consider $p'(x)$:

$$p'(x) = 2(x-c)q(x) + (x-c)^2q'(x) = (x-c)(2q(x0 + (x-c)q'(x))$$

So $x - c$ is a factor of $p'(x)$ so c is a root.

Since $p(x)$ is irreducible and has c as a root, $p(x)$ must be the minimum polynomial for c. This says that every polynomial which has c as a root is a multiple of $p(x)$. In other words, $p'(x)$ is a multiple of $p(x)$. But unless $p'(x) = 0$, this is impossible because $p'(x)$ has smaller degree than $p(x)$.

So $p'(x) = na_nx^{n-1} + \cdots + a_1 = 0$ would imply that $na_n = 0$. This cannot happen in a field of characteristic 0 (although if we were in $\mathbb{Z}_5[x]$ this would be problematic).

□

- From here on out, let's consider fields of characteristic 0 only (that is, extensions of $\mathbb{Q}$). Now irreducible polynomials have only distinct roots.
- Now do the field extension thing for these roots. What can happen?
- We can adjoin a root, and then in that bigger field, the polynomial factors. But does it factor into linear terms? Maybe, maybe not. If not, then we could find a root to one of the factors and adjoin that as well. Eventually, we will adjoin enough roots to get an extension field in which the polynomial factors completely (into linear factors).
- This extension field is call the *root field* or *splitting field* of the polynomial (because in that field, the polynomial "splits" into linear factors).

  Ex: What is the splitting field of $p(x) = x^3 + 2$? What is its degree over $\mathbb{Q}$?

  Ex: What is the degree of the splitting field of $p(x) = x^4 - 10x^2 + 25x - 5$? Well, $p(x)$ is irreducible, so there is at least one root α which we can adjoin to $\mathbb{Q}$ to get the degree 4 field extension $\mathbb{Q}(\alpha)$. Now what? From the polynomial's graph, we know that there are 2 real roots, so 2 complex roots. If α was one of the real roots, then we are not yet in the splitting field. So we need to add another root β. This might be of degree 2 or 3, depending on how $p(x)$ factored in $\mathbb{Q}(\alpha)$. It gets messy.
- Notice that if c and d are two distinct roots of $p(x)$, then $F(c) \cong F(d)$. Why?
- In fact, what can we say about isomorphisms between field extensions?

Isomorphisms of extension fields

To understand splitting fields better, and to understand how the roots of polynomials relate to each other, we want to consider isomorphisms between extension fields.

- Suppose you have two extensions $\mathbb{Q}(\alpha)$ and $\mathbb{Q}(\beta)$ of $\mathbb{Q}$. How might these be related to each other?
- Specifically, what if α and β are both roots of the same polynomial? Then each are isomorphic to $\mathbb{Q}[x]/\langle p(x)\rangle$, so isomorphic to each other.
- What does that isomorphism look like? Where can it send elements from $\mathbb{Q}$? Where does it send α?
- Since the 0 and 1 need to be sent to themselves, all of $\mathbb{Q}$ must be sent to itself!
- Elements in $\mathbb{Q}(\alpha)$ look like $a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{n-1}\alpha^n$. Apply the isomorphism to this! If we send α to β, this works.
- More in general, if you have an isomorphism from $\mathbb{Q}(\alpha)$ to $\mathbb{Q}(\beta)$, consider what happens to polynomials in $\mathbb{Q}[x]$ under this isomorphism. In particular, consider the minimum polynomial for α and β.

Field Automorphisms

- Let's consider the set of all automorphisms of a field, call it $\mathrm{Aut}(F)$. It is easy to see that $\mathrm{Aut}(F)$ is a group (under composition).
- Now consider a field extension E of F and the elements of $\mathrm{Aut}(E)$. Instead of looking at all the automorphisms, look only at the automorphisms that *fix* F. That is, $\sigma(a) = a$ for all $a \in F$.
- Call the set of these automorphisms $\mathrm{Gal}(E : F)$. This is called the *Galois group* of E over F. It is called a group because it is! Show this.

  Ex: Consider $E = \mathbb{Q}(\sqrt{3}, \sqrt{5})$. This is an extension of $\mathbb{Q}$, but also of $\mathbb{Q}(\sqrt{3})$ and $\mathbb{Q}(\sqrt{5})$. Describe $\mathrm{Gal}(E : \mathbb{Q})$. What group is this? (It is $\mathbb{Z}_2 \times \mathbb{Z}_2$.)
- So what does this have to do with polynomials?
- Well if $p(x)$ is a polynomial in $F[x]$, and suppose E is an extension of F containing (some of the) roots of $p(x)$. Then any automorphism in $\mathrm{Gal}(E : F)$ defines a permutation of the roots of $p(x)$ that lie in E, since roots are sent to roots.
- The converse of the above is also true. That is, if a and b are roots of the same polynomial, then there is an isomorphism $\varphi : F(a) \to F(b)$ that fixes F.
- Now we are ready to think about how large the Galois group can be. Say $p(x)$ is a polynomial in $F[x]$ and E is the splitting field for $p(x)$ over F. As long as f has no repeated roots, then $|\mathrm{Gal}(E : F)| = [E : F]$.

  Proof. Proceed by induction on the degree of $p(x)$. If the degree is 1, then $E = F$ and we are done. Assume the result holds for all polynomials of degree $k < n$ and that $p(x)$ has degree n.

  Now let $q(x)$ be an irreducible factor of $p(x)$ (maybe $q(x) = p(x)$), with degree m. The roots of $q(x)$ lie in E, so pick one of them, call it a. We have

  $$E : F(a)] = n/m \quad \text{and} \quad [F(a) : F] = m$$

  Now $q(x)$ has exactly m roots, all of them in E. For each root b we have an isomorphism from $F(a)$ to $F(b)$ which fixes F. So there are m such isomorphisms.

  Consider $\mathrm{Gal}(E : F(a))$. This is the set of all automorphisms of E that fix $F(a)$. By our inductive hypothesis, we have $|\mathrm{Gal}(E : F(a))| = [E : F(a)] = n/m$. Each of these composed with an isomorphism between $F(a)$ and $F(b)$ gives an element of $\mathrm{Gal}(E : F)$. □
- We did not get to this one:

  Ex: Find the Galois group of $p(x) = x^4 + x^3 + x^2 + x + 1$ over $\mathbb{Q}$. Note this is irreducible over $\mathbb{Q}$ by Eisenstein (and substitution of $x + 1$ for x). To find the roots, first multiply by $(x - 1)$ to get $x^5 - 1$. Let a be a fifth root of unity. Then $\mathbb{Q}(a)$ is the splitting

field and has degree 4 over $\mathbb{Q}$. What are the automorphisms? Well, send a to a^i for $i = 1, 2, 3, 4$. The automorphism σ_2 which sends a to a^2 generates $\mathrm{Gal}(\mathbb{Q}(a) : \mathbb{Q})$, so the group is $\mathbb{Z}_4$.

Subgroups of the Galois Group

Recall that the Galois group of a splitting field K over a base field F is the group of all automorphisms of K that leave F fixed. These automorphisms are determined by where they send the roots of a (really any) polynomial $p(x) \in \mathscr{F}[x]$ that has K as its splitting field: roots of an irreducible polynomial must be sent to other roots of the same irreducible polynomial. So we can determine quite a bit about the splitting field, polynomials that split in the field, and the Galois group using the other two of these to say something about the third.

Today, let's look at what can happen in *subgroups* of the Galois group.

- Start with a familiar example. $K = \mathbb{Q}(\sqrt{3}, \sqrt{5}) = \mathbb{Q}(\sqrt{3} + \sqrt{5})$ has Galois group $\mathbb{Z}_2 \times \mathbb{Z}_2$. We can list out the automorphisms exactly, but let's call them ε, α, β and γ.
- We can think of $\varepsilon = (0, 0)$, $\alpha = (1, 0)$, etc. where a zero means we send $\sqrt{3}$ or $\sqrt{5}$ to itself, a 1 means we sent it to its negative.
- What do the subgroups look like? There are 5 subgroups: the two trivial subgroups and the subgroups $\{\varepsilon, \alpha\}$, $\{\varepsilon, \beta\}$, and $\{\varepsilon, \gamma\}$. Draw the lattice of subgroups.
- Look at $H = \{\varepsilon, \alpha\}$, where α sends $\sqrt{3}$ to $-\sqrt{3}$ and $\sqrt{5} \mapsto \sqrt{5}$. Note that all (both) the automorphism of K here fix every element in $\mathbb{Q}(\sqrt{5})$.
- We call $\mathbb{Q}(\sqrt{5})$ the fixed field of the subgroup H. For short, call it the *fixfield* of H. Will every subgroup of $G = \text{Gal}(K/\mathbb{Q})$ have a fixfield?
- What about the other direction? Suppose you have a subfield E of K contained in $\mathbb{Q}$ (an *intermediate field*). Of course we can consider the Galois group $G(K/E)$, the set of all automorphisms of K that fix E. Try this with $\mathbb{Q}(\sqrt{3})$. You get a subgroup of G, which will will call the *fixer* of E. Does every intermediate field have a fixer?
- There is a very strong relationship between the fixfields of subgroups of G and fixers of intermeidate fields. Every subgroup has a fixfield, and the fixer of the fixfield is the subgroup. Every intermdiate field has a fixer, and the fixfield of the fixer is the intermediate field!
- Note that so far we have thought of the intermediate fields in our example as just $\mathbb{Q}(\sqrt{3})$ and $\mathbb{Q}(\sqrt{5})$. But since there are three intermediate subgroups, there must be three intermediate fields. Can we find the last fixfield? It's $\mathbb{Q}(\sqrt{15})$.
- There is more! What can we say about the degree of extensions in the field lattice and the size of the subgroups? Well, the size of a subgroup H of G will be equal to $[K : I]$, where I is the fixfield of H. Note that this really shows that the lattice of subgroups matches up with the lattice of intermediate fields, just flipped upside down.

Group Theory (again)

Motivation

Main question: What do the roots of $x^5 - 5x - 2$ look like? Is there something analogous to the quadratic formula that will tell us the roots?

Approach: Describe the splitting field E, and consider $\mathrm{Gal}(E : \mathbb{Q})$ and its subgroups. Use properties of groups to say something interesting about the subgroups. Translate this back to say something interesting about the splitting field (and its subfields). Translate that to say something about the polynomial.

But before we can do that, we need to remind ourselves about groups and learn some more advanced topics in group theory. Let us begin.

- Remember the definition of a group. There are three axioms.
- Some examples to keep in mind: D_4, S_3, $\mathbb{Z}_n$ (infinite groups like $\mathbb{Z}$, $\mathbb{Q}$, etc will be of less interest).
- How should we *represent* our groups? Remember that there are lots of groups that look different but end up being the same (isomorphic). So we should think a bit about how to present groups in some standard way.
- Do the activity: Cayley Permutations.

Cayley's Theorem

- One example of a group from last semester was S_n, the group of permutations on n elements. The symmetric groups, and subgroups of them, are nice to work with, since it is relatively easy to say how to multiply elements.
- You might wonder if there are any other groups. In some sense the answer is yes: D_4, $\mathbb{Z}_7$, $\mathbb{Z}$, $\mathbb{R}^*$, etc. But some of these turned out to be isomorphic to a symmetric group or at least a subgroup of a symmetric group. Is this always the case?
- Surprisingly YES! Cayley's theorem says that every group is isomorphic to a group of permutations. That is, to a subgroup of some symmetric group. In fact, if $|G| = n$ (that is, G has exactly n elements) then G is isomorphic to a subgroup of S_n.
- We will go through a careful proof of Cayley's theorem next time. Today, let's see it in action.

  Ex. Find a group of permutations isomorphic to $G = \{a, b, c, d\}$ whose table is given below

	a	b	c	d
a	a	b	c	d
b	b	a	d	c
c	c	d	a	b
d	d	c	b	a

Solution. There are many answers. We will give what is called the (left) regular representation of G. We will show that G is a subgroup of S_4. The idea is that we think of the elements as 1, 2, 3, and 4, and assign to each element $g \in G$ the permutation which moves each element to the result of multiplying it by g on the left.

So we will have 4 permutations: π_a, π_b, π_c, and π_d. π_a says where each element goes when you multiply it by a. Well, the elements stay put, since a is the identity. So $\pi_a = (1)$. Now π_b is the permutation: "multiply by b on the left." So $\pi_b = (12)(34)$. This is because multiplying by b causes the first and second elements to swap places, and the 3rd and 4th as well: $ba = b$, $bb = a$, $bc = d$ and $bd = c$. Next $\pi_c = (13)(24)$ and finally $\pi_d = (14)(23)$.

Proof of Cayley's Theorem

- Recall, Cayley's theorem says that every group is isomorphic to a group of permutations.
- Let's prove this. Now remember, to show G is isomorphic to H, we must find an isomorphism from G to H. We want H to be a subgroup of S_A. So for each $a \in G$, we want $f(a)$ to be a permutation of the elements in A. What should those elements be?
- How about... THE ELEMENTS OF G!?!?!? Yes, take A to be the set G.
- Consider the permutation $\pi_a : G \to G$ defined by $\pi_a(x) = ax$. That is, the permutation takes an element in G and sends it to the result of multiplying that element by a (on the left).
- Back in the chapter 6 homework, you proved that this is a bijection. (Should we do it again?) Thus π_a is a permutation of G for each element $a \in G$.
- We have a different permutation π_a for each different $a \in G$. So consider the set $G^* = \{\pi_a : a \in G\}$. This is a set of permutations in S_G, but it might not be all of them. We should show that G^* really is a subgroup of S_G. Do it.
- Now, is G really isomorphic to G^*? We claim that $f : G \to G^*$ given by $f(a) = \pi_a$ is an isomorphism. We need to check that it is bijection, and satisfies the isomorphism property:
  - f is injective: suppose $f(a) = f(b)$. Then $\pi_a = \pi_b$. So in particular $\pi_a(e) = \pi_b(e)$ which means $ae = be$ so $a = b$.
  - f is surjective: consider $y \in G^*$. So $y = \pi_a$ for some $a \in G$. So $y = f(a)$ for some $a \in G$.
  - $f(a)f(b) = \pi_a \circ \pi_b$. But $\pi_a \circ \pi_b(x) = \pi_a(\pi_b(x)) = \pi_a(bx) = abx = \pi_{ab}(x)$. So $\pi_a\pi_b = \pi_{ab} = f(ab)$. So $f(a)f(b) = f(ab)$, so f satisfies the isomorphism property.

Finite Permutation Groups

Let's dive in and really get to know the groups S_n for different values of n. Start with the activity to remind ourselves about cycle notation. The goal is to see which permutations can be written in which ways.

Transpositions

- Here is a bar bet. Suppose you have 5 playing cards (all with different numbers) laid out in a row. You and a friend take turns, swapping the position of any two cards at a time. You must work together to get the cards in the correct order and must do so after both of you have had an equal number of turns.
- The bet: you won't be able to do it, no matter what you try. That is, if I start with the correct arrangement. So how can I pick arrangements which will prevent you from completing the task?
- Notice that what we are really doing is starting with a permutation of the set $\{1, 2, 3, 4, 5\}$ and writing it as a product of transpositions. Actually, we are finding the inverse permutation, as a product of transpositions, but it should be clear that by reversing it (whatever that means) we get the original permutation as a product of transpositions. The requirement that both players move the same number of times asks for there to be an *even* number of transpositions.
- If there are some initial configurations which can not be "solved" that means that the (inverse of the) permutation *cannot* be written as the product of an even number of transpositions. There are permutations like this.
- In fact, a permutation can be written as an even number of transpositions if and only if it cannot be written as a product of an odd number of permutations. Given this fact, it makes sense to call a permutation itself either even or odd - doing so is well defined.

Even and odd permutations

- Let's prove: No permutation can be written as both an even number and an odd number of transpositions.
- Consider first how you might write the identity ε as a product of transpositions. We can prove that the number of transpositions must be even. We do this by rewriting the product using 2 fewer transpositions. This is enough, because it is obvious that ε can't be written as one transposition.
- The idea: fix some number x and find the last time x occurs (the right-most x). We move this x to the left until the transposition it is in is next to a copy of itself, in which case you can remove both copies.
- Show this process with an example: $\varepsilon = (45)(23)(13)(25)(14)(24)(35)(15)$ Use $x = 3$.
- More precisely, what are the possibilities for the transposition directly to the left of (xa). If it is (xa), remove both. If it is (bc), we can swap them, since disjoint cycles commute. If it is (xb), then write $(xb)(xa) = (xa)(ab)$. If it is (ab), then write $(ab)(xa) = (xb)(ab)$.
- Once we know that ε is even, suppose π is both even and odd. Then so would be π^{-1}. But then $\pi \circ \pi^{-1} = \varepsilon$ would be odd.
- Neat. It now makes sense to call a permutation *even* if it can be written as an even number of permutations, and *odd* otherwise. That is, the *parity* of a permutation is a property of the permutation, not just the way we happen to write it.
- We will let A_n denote the set of all even permutations in S_n. What would it take to prove that A_n is a subgroup of S_n?
- How many elements are in A_n? That is, how many of the $n!$ elements of S_n are even?
- To prove that $|A_n| = \frac{1}{2}|S_n|$, we can define a bijection from the set of even permutations to the set of odd permutations. Do this by picking any transposition σ and sending τ to $\sigma\tau$ (that is, compose the even transposition τ with σ, but this really just means writing σ at the start of the transposition τ). We can prove this is injective and surjective. Thus the set of even permtuations will be the same size as the set of odd permutations.

Order

The *order* of an element g in a group is the least n such that $g^n = e$ (if there is such an n, otherwise we say the order is infinite). Start by investigating the order of elements in S_5. What can we say about the order of a permutation in S_n based on the length of the disjoint cycles we use to represent it?

Order of Group Elements

Now we want to do this more in general – in any group G, we can consider the order of the elements of G. First we should consider exponents.

- When G is written multiplicatively, we use the shorthand a^n to mean a multiplied with itself n times.
- Write a^{-n} to mean a^{-1} multiplied by itself n times. $a^0 = e$ by convention.
- We have the usual rules of exponents: $a^m a^n = a^{m+n}$, $(a^m)^n = a^{mn}$ and $a^{-n} = (a^{-1})^n = (a^n)^{-1}$
- Do you believe these? How would you prove them? (A good thing to think about if you are going to teach algebra).
- What if G was written additively? We don't use exponents then: instead of a^n we write $na = a + a + \cdots + a$. Do multiples follow the same rules as exponents?
- Unless we need to for some good reason, we stick to multiplicative notation, so will use exponents.

  Def. If there exists a nonzero integer m such that $a^m = e$, then the *order* of the element a is the least positive integer n such that $a^n = e$.

  If there does not exists such an integer, we say that a has *order infinity*.

  We write $\text{ord}(a)$ for the order of a.

- Some examples. What are the orders of the elements in $\mathbb{Z}_6$ (this should be written additively). What are the orders of the elements in D_4? What is the order of the elements in $\mathbb{Z}$?
- Now if the order of an element a is n, then n is the smallest positive integer such that $a^n = e$. But there might be other numbers t such that $a^t = e$ as well. What can we say about these?
- It should be easy to see that if t is a multiple of n, then $a^t = e$. But is this the *only* way to get the identity? That is, if $a^t = e$, must t be a multiple of n? Why?
- Yes. To prove it, we need to use the *division algorithm* (for $\mathbb{Z}$): If m and n are integers and n is positive, then there exists unique integers q and r such that $m = nq + r$ where $0 \leqslant r < n$. (q is the quotient, r is the remainder when we divide m by n.)

- Why does this help? Well suppose $a^t = e$. Now we can divide t by n, and we get $t = nq + r$. So we have

$$a^t = a^{nq+r} = a^{nq}a^r = (a^n)^q a^r = e^q a^r = a^r$$

But that means $a^r = e$ - which is impossible because $r < n$. Unless $r = 0$. But then $t = nq$ so t is a multiple of n.

Orders of elements and subgroups

Last time we explored the orders of elements in groups. We noted that if the order of $a \in G$ was n, then the only way for $a^t = e$ is if t is a multiple of n. This used the division algorithm.

We also looked at a specific example: $G = S_5$. Here, we found that for any $\alpha \in S_5$ it was the case that $\alpha^{120} = (1)$ (where 120 is the number of elements of S_5). We proved that this is the case because we were able to prove that the only orders of elements in S_5 were $1, 2, \ldots, 6$, and each of these divide 120. But will a similar thing happen in all groups?

- First, we will consider what the different powers of $a \in G$ look like. Say the order of a is n. So we know that $a^n = e$, and for any $0 \leqslant k < n$, we have $a^k \neq e$.
- We can say more. No two powers of a less than n are the same: there must be exactly n different powers of a: $a^0, a^1, a^2, \ldots, a^{n-1}$. How do we know this?
- First note that there cannot be any *other* powers of a: consider a^m. Write $m = nq + r$. Then
$$a^m = a^{nq+r} = a^{nq}a^r = a^r$$
But $0 \leqslant r < n$ so a^m is always one of the n powers of a.
- Do those n powers of a all need to be different? Well what if they weren't? Then $a^r = a^s$ for $0 \leqslant s < r < n$ (say r is the larger of the two). In other words, $0 < r - s < n$. Now multiply both sides of $a^r = a^s$ by a^{-s}:
$$a^r a^{-s} = a^s a^{-s}$$
$$a^r a^{-s} = e$$
$$a^{r-s} = e$$
But $r - s < n$ which is the *least* power of a which give the identity. So $r - s = 0$, or in other words, $r = s$ so $a^r = a^s$.
- By the same argument, if a has order infinity, different powers of a always give different elements of G. That is, if $r \neq s$, then $a^r \neq a^s$.
- This has the following crucial consequence. If $\text{ord}(a) = n$, then $|\langle a \rangle| = n$. That is, the order of the subgroup generated by a is equal to the order of the element a. (Note the two uses of the word order).
- Remember, we are trying to decide whether $a^{|G|} = e$ for all $a \in G$. This will happen as long as $\text{ord}(a)$ divides $|G|$. But now we know that there is a subgroup of G that has size $\text{ord}(a)$.
- Perhaps we can say something about the size of the subgroups of G? Yes we can!

  Theorem (Lagrange's theorem)**.** *Let G be a finite group and H any subgroup of G. Then the order of G is a multiple of the order of H.*

- It is hard to express just how amazing this theorem is. We proved it last semester. But let's quickly recall how the proof went.
  - We simply count the number of cosets you get from a subgroup H.
  - Given any subgroup $H < G$, we can always form cosets $\{aH \ : \ a \in G\}$.
  - Each of these cosets will have the same number of elements, the same number of elements as H.
  - Further, the set of all cosets forms a *partition* of G: eveyr element is in one and only one coset.
  - So how many cosets are there? That is, what is the index of H in G? We have $[G : H] = |G|/|H|$.
  - But this is a whole number, so it must be that $|H|$ divides $|G|$.
- You might ask if this is really useful. Well what if, for example, the order of G is prime? Does G have any subgroups?
- We will consider other implications of this theorem, including to RSA cryptography, next week.
- And back to our question: Since for any element $a \in G$, we can take $\langle a \rangle$, and the order of this cyclic subgroup is just the ord(a). But that means that ord(a) must be a factor of the order of G!
- We end with an application of this result.

Ex. If p is prime and a is a positive integer less than p, then a^p divided by p has remainder a. That is, $a^p \equiv a \pmod{p}$. Hint: look at the group $\mathbb{Z}_p^*$. (This famous result is known as *Fermat's Little Theorem.*)

Proof. Recall that the group $\mathbb{Z}_p^* = \{1, 2, 3, \ldots, p-1\}$ under multiplication modulo p. In other words, after we multiply elements in $\mathbb{Z}_p^*$, we look for the remainder when the number is divided by p. So to find the remainder when a^p is divided by p, we can simply calculate a^p in the group $\mathbb{Z}_p^*$.

As long as p is prime, $\mathbb{Z}_p^*$ really is a group. The identity is 1. Now let a be any number less than p. What can we say about the order of a in $\mathbb{Z}_p^*$? Well by Lagrange's theorem, ord(a) must divide $p-1$, since $p-1$ is the order of $\mathbb{Z}_p^*$. Call ord$(a) = k$. So $p - 1 = mk$ for some m. Thus $a^{p-1} = a^{mk} = 1$. Then $a^p = 1 \cdot a = a$.

□

Today we investigated Euler's theorem, that $a^{\varphi(n)} \equiv 1 \pmod{n}$ as long as a and n are relatively prime, where $\varphi(n)$ denotes the number of numbers less than n that are relatively prime to n.

Some things we noticed: $\varphi(p) = p - 1$ for prime numbers p. It also appears that $\varphi(pq) = \varphi(p)\varphi(q)$ where p and q are prime. We will prove this next time.

Public Key Cryptography

Here is the idea: we want to publicly publish an encryption key that can be used to encode data, that cannot be decrypted unless you know the private decryption key.

- How do we encode the message? First, suppose you have some standard method to make the message into a number (or sequence of numbers). Call such a number x.
- Now we want to transform x into a encrypted number y. We do this by computing $x^E \pmod n$ for some numbers E and n. Then y can be sent.
- For this to work, we need a way to transform y back into x. We will do this by computing $y^D \pmod n$. This would only be secure if D was completely private (so we better not be able to find D from E and n). It would only work if this really did give x back.
- So we want $(x^E)^D \cong x \pmod n$. How can we make sure this happens?
- What if $DE = \varphi(n) + 1$? Or even $DE = k\varphi(n) + 1$? Well, if it happens that x is relatively prime to n, we would have

$$(x^E)^D = x^{DE} = x^{k\varphi(n)}x \equiv 1^k x \pmod n$$

- If x is not relatively prime to n, then we will also be okay, because of how we will pick n.
  - We will have $n = pq$ for primes p and q. Say x is a multiple of p (but not of q).
  - So $x = rp$ for some $r < q$. Then

$$x^{km} = x^{k\varphi(pq)} = (x^{\varphi}(q))^{k\varphi(p)} \equiv 1 \pmod q$$

  - This means that $x^{k\varphi(pq)} = 1 + tq$ for some t. We get

$$(x^E)^D = x^{k\varphi(pq)}x = (1 + tq)x = x + tq(rp) = x + trn \equiv x \pmod n$$

- Great. So we want $DE = k\varphi(n) + 1$ which is the same as saying $DE \equiv 1 \pmod{\varphi(n)}$.
- We also want $n = pq$, so $\varphi(n) = \varphi(p)\varphi(q) = (p-1)(q-1)$. Use $m = \varphi(n)$.
- So here is what we want to do:
  - Pick two really big prime number p and q; find $n = pq$ and $m = (p-1)(q-1)$ (these are easy if we start with p and q).
  - Find a number E that is relatively prime to m. Just guess and check until you find something.
  - Find D such that $DE \equiv 1 \pmod m$. You can do this using the Euclidean algorithm. You will be successful because $\gcd(E, m) = 1$

  - Don't tell anyone what p, q, m or D are. Tell everyone what n and E are.
  - Note: we want E to be fairly large. Taking $E = 2$ would always work, but then you could undo the encryption by searching for a square root perhaps. Other considerations are whether it is efficient to compute x^E, which is easier if in binary E doesn't have too many 1's.

- Work through an example with small primes. Have a student write down their favorite number x (in secret) and then compute $x^E \pmod{n}$ (in secret). Then as a class, decrypt the result.

- One concern is how to compute the power mod n in any sort of reasonable way. We can use the method of repeated squares:

  - First, write E as a sum of powers of 2 (like the Egyptians would).
  - So now we just need to compute $x^{2^{k_1}+2^{k_2}+\cdots+2^{k_n}}$. This becomes a product of terms for the form x^{2^k} for different k.
  - Notice though that $x^{2^k} = (x^2)^{2^{k-1}}$. In other words, we can keep squaring x, k times.
  - This doesn't seem to help, except that we can reduce mod n at each step. So as long as we can square a number less than n, we are good.

Today we did the RSA Cryptography "lab" activity.

Cauchy's Theorem for Abelian Groups

Our goal is to understand the structure of groups better. What does this mean? One example of what we are after is Lagrange's Theorem: if you have a group containing exactly 14 elements, could any of those elements have order 5? Lagrange says no! In fact, we know that the only possible orders are 1, 2, 7, and 14, the divisors of 14. Note that knowing about order is one step in understanding what the subgroups of a group look like (since this tells us that the cyclic subgroups look like).

A natural question: does our arbitrary group of 14 elements necessarily have elements of order 2 and 7? Of 14? That is, to what extent is the *converse* of Lagrange's theorem true? Perhaps investigate this for the group A_4, which has size 12. What orders of elements are there? Is there any subgroup of size 6?

Today we will prove Cauchy's theorem for abelian groups: if G is a finite abelian group and a prime p divides $|G|$, then there is an element of order p in G.

- Before we get to that, think about the style of proof that will be needed here. We want to show something is true for all finite abelian groups. In other words, all abelian groups of all finite orders. Perhaps we could use induction?
- In fact, we will want to use *strong* induction, since it is not at all clear how to make a group containing 1 fewer element. For fun, we will give a version of strong induction proofs known as proof by "minimal criminal".
- Let's do an example of this with numbers: prove that every number greater than 1 is either prime or is the product of primes.
  - Suppose this was not the case. What would a counterexample look like? It can't be 2 or 3 or 4 or..., but if there were a counterexample, there would need to be a *least* counterexample. This would be the *minimal criminal*.
  - The minimal criminal n cannot be prime (otherwise it is not a counterexample). So because it is greater than 1, it must be composite. So $n = ab$ for some numbers a and b both greater than 1. But also a and b are less than n, so they are not counterexamples.
  - Thus a and b are both either prime or the product of primes? Write them as such.
  - This makes n the product of primes (or the product of products of primes, which is the same thing). A contradiction.
- How might this help us with groups? Well suppose we wanted to show something was true for all finite groups. The standard induction framework would ask us to prove that it is true for all groups of order 1 (base case), and also show that *if* it is true of all groups of order less than n *then* it is true of all groups of order n.
- In terms of minimal criminals, we would assume there was a counterexample, and let G be a group of smallest order for which the thing we wanted to show was false. Now do something to get a group H of order less than n. The result holds of H. Use that fact to go back to G and conclude the same for G, contradicting the assumption there was a counterexample.

- We will prove that if G is a finite abelian group and a prime p divides $|G|$, then there is an element of order p in G.
- This is clearly true if $|G| = 1$, since there is no such prime. Now suppose G is an arbitrary group of order n and let p be a prime dividing n.
- First, what if there is an element a with $\text{ord}(a) = tp$ for some integer t? Will there be an element of order p then? This does not use induction.
- Now let a be an element which has order NOT a multiple of p. What can we say about $\langle a \rangle$ and the quotient group $G/\langle a \rangle$? Specifically, is the size of the quotient group less than n? Is it a multiple of p?
- Since $G/\langle a \rangle$ is a group (we know $\langle a \rangle$ is a normal subgroup since G is abelian) with order less than n, we know that the result applies to it (this is the inductive hypothesis). Since p divides the order of $G/\langle a \rangle$, we know this quotient group contains an element of order p.
- Call that element Hb (since elements in the quotient group are cosets). So $\text{ord}(Hb) = p$. What does this say about $\text{ord}(b)$ in G? Well let $\text{ord}(b) = k$. Then $b^k = e$ so $(Hb)^k = He = H$. So k is a multiple of p. So $\text{ord}(b) = tp$ for some p, and thus there is an element in G of order p (namely b^t).

Structure of Groups

Direct Products of and in Groups

Do the activity on direct products.

Fundamental Theorem of Finite Abelian Groups

The goal today is to understand how to classify all finite abelian groups.

- Some abelian groups are cyclic. And if they are, we know exactly what they look like. They are isomorphic to $\mathbb{Z}_n$ for some n (and this happens precisely when there is an element of order n in the group).
- But not all abelian groups are cyclic. In fact, even $\mathbb{Z}_2 \times \mathbb{Z}_2$ is an example of this. So what does an abelian group look like in general?
- Note that if a group is not cyclic, then it is not generated by a single element. But it will be generated by some number of elements. There are groups that are *finitely generated* and others that are not (require an infinite set of generators). Of course finite groups are finitely generated.
- If g_1, g_2, $\ldots g_n$ is a set of generators, that means that every element of the group can be written as a product of powers of these elements. If the group is abelian, then we can say each element $g \in G$ is $g = g_1^{k_1} g_2^{k_2} \ldots g_n^{k_n}$. Some of the k_i might be negative or zero.
- Now the powers of g_1 are exactly the elements of the cyclic subgroup $\langle g_1 \rangle$, and similarly for all g_i. So this says that $G = \langle g_1 \rangle \langle g_2 \rangle \cdots \langle g_n \rangle$. So perhaps you would think that G could be written as an *internal* direct product of cyclic groups.
- To make sure that were true, we would also need $\langle a_i \rangle \cap \langle a_j \rangle = \{e\}$ (the other requirement, that the subgroups are normal, we get for free since we are in an abelian group).
- This will not always be true. In $\mathbb{Z}_{12}$, if we picked $a_1 = 3$ and $a_2 = 6$, then we are done for.
- So we must pick the generators carefully. We pick them based on what their orders are.
- Given any abelian group G, let $G(p)$ be the set of all elements whose orders are a power of p. For prime p, we call such groups *p-groups*. Our first step is to break G down as the direct product of p-groups.
- Specifically, say $|G| = p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}$. We can show that G is the internal direct product of $G(p_1)$, $G(p_2)$,$\ldots G(p_n)$.
- It is easy to verify that $G(p_i)$ is really a subgroup. Thus we must only argue that $G(p_i) \cap G(p_j) = \{e\}$ for all $i \neq j$ and that $G = G(p_1)G(p_2) \cdots G(p_n)$.
- The first is easy (by a homework problem you did in fact: no power of one is equal to a power of another).

- The other must be true because we know that the product of these subgroups is at very least a subgroup of G, but it must have the same size, and therefore be equal to G.

- So now we have the G is a direct product of p-groups. These p-groups might or might not be cyclic (consider the two examples we had on the activity from last time).

- We need to understand the structure of p-groups. It would be nice to pick generators so that the p-group could be expressed as the direct product of cyclic groups (all of which will also be p-groups).

- Again, do so by considering the order of elements. Say we have a p-group G_p. Let g be any one of the elements in G_p of maximal order. So $\text{ord}(g) = p^m$ for some m. We will show that $G_p = \langle g \rangle \times H$ for some subgroup H of G_p.

- This will be enough, because then we can proceed by induction to break down H (also a p-group, but of smaller order) similarly. Eventually we will be left with a second group that is cyclic (or trivial).

- Here is the basic idea. The proof is by induction. We take g to be of largest order in G_p. Let h be of smallest order of all elements not in $\langle g \rangle$. Then shift to the quotient group $G/\langle h \rangle$ and look at the element $\langle h \rangle g$. Argue that this has maximal order in $G/\langle h \rangle$ and by induction conclude that $G/\langle h \rangle \cong \langle \langle h \rangle g \rangle \times H/\langle h \rangle$ for some subgroup H of G containing $\langle h \rangle$. Then argue that this H makes $G_p \cong \langle g \rangle \times H$.

- The takeaway: The Fundamental Theorem of Finite Abelian groups tells us that any finite abelian group G is isomorphic to the direct product of cyclic groups, each of order a power of a prime. That is,

$$G \cong \mathbb{Z}_{p_1^{k_1}} \times \mathbb{Z}_{p_2^{k_2}} \times \cdots \times \mathbb{Z}_{p_n^{k_n}}$$

  (the p_i need not be distinct).

- As an example, let's classify all finite abelian groups of order $540 = 2^2 3^3 5$. One of these is $\mathbb{Z}_{2^2} \times \mathbb{Z}_{3^3} \times \mathbb{Z}_5$. Another is $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \times \mathbb{Z}_{3^3} \times \mathbb{Z}_5$. Or we can break down the 3^3 in two different ways. All together there will be 6 groups with this order.

Subnormal Series

The goal today is to gain familiarity with the definitions of subnormal and normal series, and to understand what a composition or principle series is. This will lead to the definition of a solvable group.

One preliminary is the notion of a *simple group*. These are groups that have no non-trivial normal subgroups. One of the great accomplishments of algebra recently is the complete classification of all finite simple groups in 2008. It turns out that the simple groups are only

- $\mathbb{Z}_p$ for prime p,
- A_n for $n \geqslant 5$,
- Belong to one of 16 families of groups of Lie type.
- One of 26 exceptions (called the sporadic groups, 20 of which are subquotients of the *Monster group* of order $2^{46}3^{20}5^{9}7^{6}11^{2}13^{3}17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71 \approx 8 \times 10^{53}$, the other 6 known as *p*ariahs).

For us: $\mathbb{Z}_p$ or A_n.
Now do the activity to remind ourselves how subgroups and quotient groups interact.

More Subnormal Series and Solvable Groups

- Motivating question: Let G be a group with normal subgroup H. Suppose G/H is not simple (that is, it has a normal subgroup). What does that tell us about the relationship between G and H?
- Perhaps consider the example $G = \mathbb{Z}_{60}$ and $H = \langle 6 \rangle$. What is $\mathbb{Z}_{60}/\langle 6 \rangle$? It looks like $\mathbb{Z}_6$. Now $\mathbb{Z}_6$ has a normal subgroup $\{0, 3\}$. Which subgroup of G/H is this? Can we recognize this as a quotient group itself?
- Finish up the activity from last time.
- The Jordan-Hölder Theorem: Any two composition series of G are isomorphic. By isomorphic, we mean there is a 1-1 correspondence between the quotient groups in the composition series.
- This is really remarkable. It says that however you decompose a group using a subnormal series, you will essentially get the same series (in as much as the quotient groups will match up). In particular, the *length* of composition series is well defined.
- Another useful definition is this: A group G is *solvable* if it has a subnormal series $\{H_i\}$ such that all the quotient groups H_{i+1}/H_i are abelian. Next week we will see why calling such groups "solvable" makes sense.

Today we reviewed composition series, motivated by a group quiz on the topic.

Back to Fields

Here is the main question we asked after the last exam: What do the roots of $x^5 - 5x - 2$ look like? Is there something analogous to the quadratic formula that will tell us the roots?

Approach: Describe the splitting field E, and consider $\text{Gal}(E : \mathbb{Q})$ and its subgroups. Use properties of groups to say something interesting about the subgroups. Translate this back to say something interesting about the splitting field (and its subfields). Translate that to say something about the polynomial.

Today, let's remind ourselves about what we know from Galois theory.

- Given a polynomial $p(x)$, we can find its *splitting field* E, the smallest extension field in which the polynomial factors into linear terms (splits).
- We can look at automorphisms of the splitting field that *fix* the base field (usually $\mathbb{Q}$, which has to be fixed anyway).
- These automorphisms must send roots of irreducible polynomials to roots of the same irreducible polynomial. If $p(x)$ is not irreducible, then we need to look at its irreducible factors to see what the automorphisms look like.
- The set of all $\mathbb{Q}$-fixing automorphisms of E forms a group (under composition) called the *Galois group of E over $\mathbb{Q}$*, written $\text{Gal}(E : \mathbb{Q})$.
- There is a strong connection between the field extension and the Galois group, which is the content of the *Fundamental Theorem of Galois Theory*.
- One of these is that the size of the Galois group is equal to the degree of the field extension. But there is more.
- If we look at a subgroup H of $\text{Gal}(E : \mathbb{Q})$, we get a subset of the automorphisms of E that fix $\mathbb{Q}$. We consider the set of elements in E that are fixed by all the automorphisms in H. This is a field F, and thus an *intermediate field* between $\mathbb{Q}$ and E. We call it the *fixed field* of H.
- On the other hand, if we take an intermediate field I (so $\mathbb{Q} \subset I \subset E$) we can consider $\text{Gal}(E : I)$, the group of all automorphisms of E that fix I. We call this group the *fixer* of I.
- These match up exactly! We could prove: If H is the fixer of I, then I is the fixed field of H, and if I is the fixed field of H, then H is the fixer of I.
- And it gets even better. Suppose that I is not just an intermediate field, but is itself a splitting field from some polynomial. In this case it makes sense to consider $\text{Gal}(I : \mathbb{Q})$ as well.
- These are automorphisms of I that fix $\mathbb{Q}$. Now technically none of these automorphisms belong to $\text{Gal}(E : \mathbb{Q})$, but they are related. By a homomorphism!

- Given an automorphism in $\mathrm{Gal}(E : \mathbb{Q})$, consider its restriction to I. That is, simply restrict the domain to only include elements from I. It turns out that the function that takes a automorphism to its restriction is a homomorphism. And where there's a homomorphism, there is a kernel. These will be the automorphisms of $\mathrm{Gal}(E : \mathbb{Q})$ which are the identity on I. That is, exactly the elements of $\mathrm{Gal}(E : I)$.

- So by the Fundamental Homomorphism Theorem,

$$\mathrm{Gal}(I : \mathbb{Q}) \cong [\mathrm{Gal}(E : \mathbb{Q})]/[\mathrm{Gal}(E : I)]$$

  In particular, we see that the fixer of a splitting field is a normal subgroup.

  Ex: Show the Galois correspondence for the whole diagram of intermediate fields of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.

  Ex: The field $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{7})$ has Galois group isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$. Consider the two subnormal series

$$\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \supset \langle (1,0,0), (0,1,0) \rangle \supset \{(0,0,0)\}$$

$$\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \supset \langle (0,0,1) \rangle \supset \{(0,0,0)\}$$

  What intermediate fields do these match up with?

Polynomials Solvable by Radicals

Today we will try to understand why there is nothing like the quadratic formula for polynomials in general. Specifically, there is no quintic (or higher) formula.

- What do we mean by something like the quadratic formula? We mean solvability by radicals. That is, is there some way to express the roots of a polynomial in terms of field operations and nth roots?
- This question can be framed in the language of extension fields. Remember, we know that every polynomial has roots in some extension field. Asking what the roots look like is asking what the extension field looks like.
- In particular, we need to see if the roots lie in a field extension obtained by taking nth roots of elements in lower fields.
- We say that an extension E of a field F is an *extension by radicals* if there is a chain of subfields
$$F = F_0 \subseteq F_1 \subseteq F_2 \subseteq \cdots \subseteq F_r = E$$
such that for $i = 1, 2, \ldots, 3$ we have $F_i = F_{k-1}(\alpha_i)$ and $\alpha_i^{n_i} \in F_{i-1}$.
- Then a polynomial $f(x)$ is *solvable by radicals* over F if the splitting field K of $f(x)$ over F is contained in an extension of F by radicals.
- An easy example of a polynomial that is solvable by radicals is $x^n - a$ (for any $a \in \mathbb{Q}$). We can say exactly what the roots look like. Let ω be a primitive nth root of unity (what does this mean?), so the roots are $\sqrt[n]{a}\omega^k$ for $k \in \{0, 1, \ldots, n-1\}$.
- We wish to make a connection to groups. We do so using Galois theory. Let the splitting field for $x^n - a$ over $\mathbb{Q}$ be E. What does $\mathrm{Gal}(E : \mathbb{Q})$ look like?
- First extend $\mathbb{Q}$ by ω. We can check that $\mathrm{Gal}(\mathbb{Q}(\omega) : \mathbb{Q})$ is abelian, and thus solvable. Then extend again to get E (we know $E \supseteq \mathbb{Q}(\omega)$). We can argue that $\mathrm{Gal}(E : \mathbb{Q}(\omega))$ is also abelian, and thus solvable. Then put it together (look at the two series).
- What this says is that at least in this special case, since the polynomial is solvable by radicals, the splitting field's Galois group is solvable. This holds in general. Essentially, just repeat the process for each extension (putting in the primitive roots of unity as required).
- The converse happens to also be true. So a polynomial is solvable by radicals if and only if the Galois group of its splitting field is a solvable group.
- Now consider the polynomial $p(x) = x^5 - 6x^3 - 27x - 3$. This is irreducible by Eisenstein. Thus there are 5 distinct roots in some splitting field E.
- What do the roots look like? By either graphing or doing some calculus, we can verify that there are exactly three real roots (there are only two critical points). Thus there are exactly two complex roots.

- The automorphisms of the splitting field are determined by what they do to the five roots. So each automorphism can be viewed as an element of S_5.
- One automorphism is complex conjugation. This leaves the three real roots fixed, and permutes the other two. So in S_5, this is a transposition.
- We also know that $\text{Gal}(E : \mathbb{Q})$ has order that is a multiple of 5. We get this by considering the tower of extensions of $\mathbb{Q}$ in E: the first extension must be of degree 5, so the degree $[E : \mathbb{Q}]$ is a multiple of 5, and this is the same as the order of $\text{Gal}(E : \mathbb{Q})$. By Cauchy's theorem, $\text{Gal}(E : \mathbb{Q})$ must have an element of order 5. In S_5 this is necessarily a 5-cycle.
- So $\text{Gal}(E : \mathbb{Q})$ is a subgroup of S_5 that contains a 2-cycle and a 5-cycle. We can show that this means that $\text{Gal}(E : \mathbb{Q}) \cong S_5$. In fact, S_5 is generated by any 2-cycle and any 5-cycle.
- Now consider the composition series for S_5:

  $$S_5 \supset A_5 \supset \{0\}$$

  Since A_5 is simple, we get that S_5 is not solvable. And this means that $p(x)$ is not solvable by radicals. Ta Daa.

Application to Counting

Let's end the year with an example of how understanding group structure can help solve problems in other areas of mathematics. Consider questions of combinatorics. In Discrete you learned a variety of methods for quickly counting the number of outcomes (this is also useful in probability of course).
Ex: How many ways are there for King Arthur and his 9 knights of the round table to sit around the round table? Why is the answer not just 10 factorial?

- One way to think about the above problem is to say that there are 10 seats, with 10 choices for who sits in the first, 9 choices for who sits in the second, and so on. But this is too many, since we just care who sits next to whom, not which chairs they sit in. So divide by 10 to account for the 10 different seating arrangements that should really count as one outcome.
- We have a similar way of correcting when we think of counting *combinations* instead of permutations.
- What about this problem: You want to put out place mats for Arthur and his knights. The mats come in two colors: blue and gold. How many ways can you arrange the place mats around the table?
- This is harder. We could start by saying there are 2^{10} arrangements (each spot has 2 choices). But how do we divide out by the duplicates?
- The problem is that the number of duplicates depends on the number of each color. For example, we have only counted 1 outcome where all the mats are gold. But we have counted 10 with 1 blue and 9 gold, and they are all "the same". We have counted 45 with 2 blue and 8 gold. How many of those should we have counted?
- This is getting complicated fast. It seems like while this way of dividing up the outcomes could work, it is certainly no easy task.
- So here is an alternative suggestion. Instead of creating cases based on how many of each color we have, break up the set of distributions by the type of duplication.
- For example, we wanted to group together the 1 blue/9 gold placements because each could be achieved by rotating the table from another. So maybe lets group together all placements that are equivalent under this symmetry.
- The advantage: we understand symmetries fairly well: they give us a group structure. Here, the group of all rotations of a 10-gon, which is a subgroup of D_{10}.
- To make this precise, we will talk about the group *acting* on the set of outcomes (colorings). So let's see what we can say about "group actions."

Group Actions

- Consider a group G and a set X. An *action* of G on X is a function that sends pairs in $G \times X$ to things in X. We write $(g, x) \mapsto gx$, satisfying

  1. $ex = x$ for all $x \in X$
  2. $(g_1 g_2)x = g_1(g_2 x)$ for all $x \in X$ and all $g_1, g_2 \in G$.

  We call X a *G-set*.

- Two elements $x, y \in X$ are *G-equivalent* (written $x \sim y$) provided there is some $g \in G$ such that $gx = y$.

- The *orbit* of an element $x \in X$ (written $\mathcal{O}_x$) is the set of all elements $y \in X$ that are G-equivalent to x.

- The *fixed point set* of an element $g \in G$ (written X_g) is the set of all $x \in X$ such that $gx = x$.

- The *stabilizer subgroup* of an element $x \in X$ (written G_x) is the set of all $g \in G$ such that $gx = x$.

Do an example with D_4 acting on the vertices (or edges) of a square.

We would like to make some observations about how G, X, $\mathcal{O}_x$, G_x and X_g are related, in terms of sizes. We start with the *Orbit-Stabilizer Theorem*:

Theorem. *Let G be a finite group acting on a set X. Then for any $x \in X$,*

$$|G| = |\mathcal{O}_x| \cdot |G_x|$$

Proof. Fix $x \in X$ and consider the elements of $\mathcal{O}_x$. These are the elements $y \in X$ such that $gx = y$ for some $g \in G$. But of course there might be more than one element in G which moves x to y. Say $g_1 x = y$ and $g_2 x = y$. Then $g_1 x = g_2 x$ so $g_2^{-1} g_1 x = x$. In other words, $g_2^{-1} g_1 \in G_x$.

Consider the cosets G/G_x (remember G_x is a subgroup of G, so this makes sense). Each coset has the form gG_x for some $g \in G$. But look at what we said above: $g_2^{-1} g_1 \in G_x$ precisely when g_1 and g_2 both move x to the same element in $\mathcal{O}_x$, and now we are saying that this happens exactly when g_1 and g_2 are in the same coset. So each different element of $\mathcal{O}_x$ corresponds exactly to a coset of G/G_x. In other words:

$$|\mathcal{O}_x| = |G/G_x| = [G : G_x] = |G|/|G_x|$$

where the last equality is by Lagrange's theorem. □

Let the *fixed point set* of an element $g \in G$ (written X_g) be the set of all $x \in X$ such that $gx = x$. Note here we are indexing by elements from G, instead of from X (as in $\mathcal{O}_x$ or G_x).

Now we can prove Burnside's lemma:

Theorem. *Let k be the number of distinct orbits defined by the action of G on X. Then*

$$k = \frac{1}{|G|} \sum_{g \in G} |X_g|.$$

In other words, the number of orbits is the average number of points fixed by the elements in G.

Proof. Let's count $\sum_{g \in G} |X_g|$. This counts all the pairs (g, x) such that g fixes x. What if we break this down by x instead of g. We get

$$\sum_{g \in G} |X_g| = \sum_{x \in X} |G_x|$$

Now break it down even further: divide up the sum by orbits. For any single orbit $\mathcal{O}_x$ we want to find $\sum_{y \in \mathcal{O}_x} |G_y|$. But if x and y are in the same orbit, then $|G_x| = |G_y|$[1] so

$$\sum_{y \in \mathcal{O}_x} |G_y| = |\mathcal{O}_x| \cdot |G_x| = |G|$$

If we do this for all k orbits, we get

$$\sum_{x \in X} |G_x| = k \cdot |G|$$

but the left hand side is really $\sum_{g \in G} |X_g|$, as we need. □

Here are some examples.

Ex: How many place mat placements are possible around Arthur's 10 seat round table using blue and gold place mats? We count the number of orbits of elements in X under a group action of G. What are X and G? Let X be the set of all "colorings" (place mat placements) and let G be the group of rotations of a 10-gon (a subgroup of D_{10}). We need to find $|X_g|$ for each element $g \in G$ and then average them. How many colorings are fixed by the identity? All of them. What about by a rotation by 36 degrees? Just 2. Same for rotations by 3, 7, 9. Other rotations? We get:

$$\frac{1}{|G|} \sum_{g \in G} |X_g| = \frac{1}{10}(2^{10} + 4 \cdot 2^1 + 4 \cdot 2^2 + 2^5) = 108$$

Ex: How many 6-bead bracelets can you make using 3 colors of beads? Note we need all of D_6 here.

$$\frac{1}{|G|} \sum_{g \in G} |X_g| = \frac{1}{12}(3^6 + 2 \cdot 3^1 + 2 \cdot 3^2 + 3^3 + 3 \cdot 3^4 + 3 \cdot 3^3 = 92$$

Ex: How many ways can you color the faces of a cube using 4 colors? We need to decide on the group of symmetries of a cube. This will be homework.

[1] How do we know this? Define $\varphi : G_x \to G_y$ by $\varphi(a) = gag^{-1}$ where g is the element which gives $gx = y$. You can show that $gag^{-1}y = y$. This is a isomorphism.