Lecture 1: Definition and Examples of Rings

Rings: Motivation and Definition

Throughout your education, you've encountered many different mathematical objects that can be "added" and "multiplied":

Integers, rationals, real numbers: You learned what it meant to "add" numbers and "multiply" these numbers in primary school.

Polynomials: You learned to "add" and "multiply" polynomials with real coefficients in high school

$$(3x^2 + 1x - 0.1) + (5x + 9) = (3 + 0)x^2 + (1 + 5)x + (-0.1 + 9)$$
$$(3x^2 + 1x - 0.1) * (5x + 9) = (3 * 5)x^3 + (3 * 9 + 1 * 5)x^2 + (1 * 9 + (-0.1 * 5))x + (-0.1 * 9)$$

Square matricies: You learned to "add" square matricies by adding the corresponding terms of the matricies, and to "multiply" square matricies by the matrix multiplication algorithm.

Functions: You "add" functions by constructing a new function whose output values are the sum (as numbers) of the original functions, and you "multiply" functions by constructing a new function whose output values are the product (as numbers) of the original functions.

If you were a very apt student, you might have complained to your teachers that the terms "add" and "multiply" are really overloaded! Why use the same two words to describe such different operations on such different kinds of mathematical objects?

One reason is that there are properties of these "adding" and "multiplying" operations that are true in all of these examples – properties such as $a + b = b + a$ and $a(bc) = (ab)c$ and $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$. In this way, the process of "adding" and "multiplying" square matricies (or polynomials, or functions) *behaves like* the process of "adding" and "multiplying" numbers. As you get further in mathematics, you will encounter *even more* mathematical objects that have "addition" and "multiplication" operations with these properties, so mathematicians have a special name for them, a *ring*.

Definition: A **ring** is a set R with two operations, called *addition* (written $a+b$) and *multiplication* (written $a * b$ or $a \cdot b$ or ab) such that

- $(R, +)$ is an abelian group. In case you forget, this means:
  - Addition is associative: $a + (b + c) = (a + b) + c$
  - There is an identity element 0, with $0 + a = a + 0 = a$ for all a.
  - Every element has an inverse $a + (-a) = 0$.
  - Addition is commutative: $a + b = b + a$.
- Multiplication is associative: $a * (b * c) = (a * b) * c$.
- Multiplication distributes over addition:

$$a * (b + c) = a * b + a * c \text{ and } (b + c) * a = b * a + c * a$$

Notation: We often write "0" for the additive identity of a ring. If n is a natural number then r^n means r multiplied by itself n times.

Examples of Rings

We've seen some examples of rings above.

Example 0: The integers $\mathbb{Z}$, the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$, the polynomials with real coefficients $\mathbb{R}[x]$, the $n \times n$ matricies of integers $M_n(\mathbb{Z})$, the continuous functions on the real line $C(\mathbb{R})$.

We can also invent new rings by taking abelian groups we know from group theory, and inventing a multiplication rule that satisfies the ring axioms.

Example 1: Recall that the elements of the abelian group $\mathbb{Z}_4 (= \mathbb{Z}/4\mathbb{Z})$ are the cosets

$$\begin{array}{ll} [0] = \{4k \mid k \in \mathbb{Z}\} & [1] = \{1 + 4k \mid k \in \mathbb{Z}\} \\ [2] = \{2 + 4k \mid k \in \mathbb{Z}\} & [3] = \{3 + 4k \mid k \in \mathbb{Z}\}. \end{array}$$

We can define a multiplication operation on $\mathbb{Z}_4$ according to the table below.

*	[0]	[1]	[2]	[3]
[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]
[2]	[0]	[2]	[0]	[2]
[3]	[0]	[3]	[2]	[1]

In fact, we can make any of the abelian groups $\mathbb{Z}/n\mathbb{Z}$ into a ring by defining multiplication as $[a][b] = [ab]$.

Example 2: For any real number n, the abelian group $n\mathbb{Z} = \{nk \mid k \in \mathbb{Z}\}$ is a ring under the standard multiplication of numbers. When n is an integer, $n\mathbb{Z}$ is a subset of $\mathbb{Z}$ which is closed under addition, multiplication, and taking additive inverses. In other words, $n\mathbb{Z}$ is a subset of the ring $\mathbb{Z}$ which is itself a ring.

Definition: A subset $S \subseteq R$ of a ring which is closed under the addition and multiplication operations, as well as taking additive inverses, is called a **subring** of R.

Example 3: For *any* ring R, we can consider the polynomials with coefficients in R, written $R[x]$. The rules for addition and multiplication are below.

$$(r_2x^2 + r_1x + r_0) + (r_2'x^2 + r_1'x + r_0') = (r_2 + r_2')x^2 + (r_1 + r_1')x + (r_0 + r_0')$$

$$\begin{aligned}(r_2x^2 + r_1x + r_0) * (r_2'x^2 + r_1'x + r_0') = (r_2r_2')x^4 &+ (r_2r_1' + r_1r_2')x^3 + \\ &(r_2r_0' + r_1r_1' + r_0r_2')x^2 + (r_1r_0' + r_0r_1')x + (r_0r_0')\end{aligned}$$

Example 4: If R and S are two rings, the **product** $R \times S$ is the ring whose elements are pairs (r, s), where $r \in R$ and $s \in S$, and the operations are defined as follows.

$$(r, s) + (r', s') = (r + r', s + s')$$

$$(r, s) * (r', s') = (r * r', s * s')$$

WARNINGS

Although we can think of rings as "mathematical objects that you can add and multiply, sort of like numbers," there are some properties of numbers that are *not* true in all rings.

Warning 0: Not all rings have an element 1 that satisfies the condition $1a = a1 = a$ for all ring elements a, but the most commonly encountered rings in mathematics do. Some (misguided) people include the requirement "1 exists" in the *definition* of a ring.

Definition: An element 1 of a ring R that satisfies $1a = a1 = a$ for all $a \in R$ is called a **multiplicative identity** of R. A ring with a multiplicative identity element 1 is called a **ring with unity** or **ring with identity**.

If a multiplicative identity exists, it must be unique. To see this, suppose 1_a and 1_b are both multiplicative identities. Then $1_a 1_b = 1_b$ and $1_a 1_b = 1_a$, so $1_b = 1_a$.

Warning 1: Sometimes, $ab \neq ba$. For example, try taking two elements a and b in the ring of 2 by 2 matricies of integers and calculating ab and ba. Unless you got very lucky when you chose a and b, you will find that $ab \neq ba$. There is a special name given to a ring for which $ab = ba$ is true for *all* elements a and b.

Definition: A ring R is **commutative** if for all $a, b \in R$, $ab = ba$.

Warning 2: You can't "divide" in rings. In the ring of real numbers, the expression $\frac{a}{b}$ is shorthand for ab^{-1}, and the symbol b^{-1} means the *multiplicative inverse* of b (i.e. the real number such that $b^{-1} \cdot b = b \cdot b^{-1} = 1$). In an arbitrary ring R, not all elements have multiplicative inverses. First, it's possible that a ring doesn't even contain a multiplicative identity (so the concept of a "multiplicative inverse" doesn't even make sense). Even if a ring contains 1, some elements won't have a multiplicative inverse. Elements that do are called *units*.

Definition: For $a \in R$, an element $b \in R$ is called a **multiplicative inverse** of a if $ab = ba = 1$. If a has a multiplicative inverse, it is called **invertible** or **a unit**. If every nonzero element of R is a unit then R is called a **division ring**. A commutative division ring is called a **field**.

Warning 3: For elements a, b in a ring R, sometimes $ab = 0$ even when a and b are both nonzero. For example, $[2] \cdot [3] = 0$ in $\mathbb{Z}/6\mathbb{Z}$.

Definition: If a and b are nonzero elements of R such that $ab = 0$ or $ba = 0$, then a and b are called **zero divisors**. A commutative ring R with identity is called an **integral domain** if it has no zero divisors.

For these reasons, you have to be very careful when you're proving things about rings, because you might get tempted to use properties that *feel true* but aren't. On the other hand, there are some properties of rings that *are* true. I've proved one of them below.

Proposition: Let R be a ring. For any $a \in R$, $0a = a0 = 0$

Proof: Because 0 is an additive identity, $0 + 0 = 0$. Combining this with the distributive property gives

$$0a = (0 + 0)a = 0a + 0a.$$

Adding (-0a) to both sides of the equation proves $0 = 0a$. The proof that $a0 = 0$ is similar.

Proposition: Let R be a ring with identity. If b_1 and b_2 are both multiplicative inverses of a, then $b_1 = b_2$.

Proof: Observe that $b_1ab_2 = b_1 1 = b_1$, and also $b_1ab_2 = 1b_2 = b_2$. Therefore, $b_1 = b_2$

Lecture 2: Homomorphisms and Ideals

Homomorphisms

Consider the ring R whose elements are sequences of integers $(a_0, a_1, a_2, \dots)$ which have only finitely many nonzero elements. Addition and multiplication are given by

$$(a_0, a_1, a_2, \dots) + (b_0, b_1, b_2, \dots) = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \dots)$$
$$(a_0, a_1, a_2, \dots) * (b_0, b_1, b_2, \dots) = (a_0 b_0, a_1 b_0 + b_0 a_1, a_2 b_0 + a_1 b_1 + a_0 b_2, \dots)$$

This ring seems to have a *lot* in common with the ring of polynomials with integer coefficients $\mathbb{Z}[x]$. In fact, from a polynomial $p(x) \in \mathbb{Z}[x]$ you can construct an element of R by letting $(a_0, a_1, a_2, \dots)$ be the sequence of coefficients of $p(x)$.

$$3 + 0x - 5x^2 \mapsto (3, 0, -5, 0, 0, 0, \dots)$$

and you can check that the rule for adding and multiplying elements in $\mathbb{Z}[x]$ is the same as the rule for adding and multiplying elements in R.

So, although R and $\mathbb{Z}[x]$ are *not* the same ring, the only difference between them is a superficial one: the names we give to the elements. Soon, we'll have a fancy word to describe this notion of sameness, *isomorphism*, but the most important concept from this example is the following: when we're trying to compare two different rings, it's often helpful to study maps from the elements of one ring to elements of the other that "preserve the structure." This idea is formalized by the definition of a *ring homomorphism*.

Definition: A **ring homomorphism** φ from a ring R to a ring S is a mapping from the elements of R to the elements of S that preserves the operations $+$ and $*$. That is, for all $a, b \in R$ it satisfies

$$\varphi(a + b) = \varphi(a) + \varphi(b) \text{ and}$$
$$\varphi(a * b) = \varphi(a) * \varphi(b).$$

If φ is injective (one-to-one) and surjective (onto), it is an **isomorphism**. The **kernel** of φ is $\ker(\varphi) := \{r \in R \mid \varphi(r) = 0\}$, and the **image** of φ is $\operatorname{im}(\varphi) := \varphi(R)$.

In the definition above, the expressions $a + b$ and $a * b$ use the addition and multiplication operations from R, while the expressions $\varphi(a) + \varphi(b)$ and $\varphi(a) * \varphi(b)$ use the operations from S.

Remark: Just like in group theory, a ring homomorphism φ is injective if and only if $\ker(\varphi) = \{0\}$. This follows from the fact that every ring homomorphism is also a group homomorphism of the underlying abelian group, so the same proof from group theory carries over to the context of ring homomorphisms.

Examples

Example 1: Let $\varphi : \mathbb{R} \to \mathbb{R}[x]$ send a number a to the constant polynomial a.

Example 2: Let $\varphi : C(\mathbb{R}) \to \mathbb{R}$ send a function f to the number $f(5)$.

Example 3: Let $\varphi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ send a to the coset $[a]$.

Example 4: Let R be a ring with identity, and let $\varphi : \mathbb{Z} \to R$ send k to $1 + 1 + \cdots + 1$ (where there are k copies of "1" added together).

Example 5: (from topology). Let $\psi : X \to Y$ be a continuous map of topological spaces, and let $C(X), C(Y)$ be the rings of continuous functions on X and Y respectively. Define $\varphi : C(Y) \to C(X)$ by $f \mapsto f \circ \psi$.

Non-examples: The following are *not* examples of ring homomorphisms. Explain why.

- Define $\varphi : \mathbb{Z} \to 2\mathbb{Z}$ by $\varphi(a) = 2a$.
- Define $\varphi : \mathbb{Z} \to \mathbb{Z}$ by $\varphi(a) = a^2$.
- Let $C^\infty(\mathbb{R})$ be the set of functions on $\mathbb{R}$ which are infinitely-many times differentiable. Then define $\varphi : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ by $\varphi(f) = \frac{df}{dx}$.

Proposition: If $\varphi : R \to S$ is a ring homomorphism, then $\ker(\varphi)$ is a subring of R, and $\mathrm{Im}(\varphi)$ is a subring of S

Proof: We must show that $\ker(\varphi)$ and $\mathrm{im}(\varphi)$ are closed under addition, additive inverses, and multiplication. Because ring homomorphisms are group homomorphisms of the underlying abelian group, we know from group theory that $\ker(\varphi)$ and $\mathrm{im}(\varphi)$ are subgroups. Therefore, they are closed under addition and additive inverses. Then for $a, b \in \ker(\varphi)$.

$$\varphi(ab) = \varphi(a)\varphi(b) = 0$$

proves that $\ker(\varphi)$ is closed under multiplication, and

$$\varphi(a)\varphi(b) = \varphi(ab)$$

proves that $\mathrm{im}(\varphi)$ is closed under multiplication.

In group theory, the kernels of group homomorphisms weren't just *any* subgroups, they were a special sort of subgroup called a *normal subgroup*. Likewise, the kernels of ring homomorphisms will be a special kind of subring called an *ideal*. To prepare for the definition of an ideal, let's quickly review normal subgroups.

Review of Normal Subgroups

Suppose we have a group $(G, \cdot)$ which is not necessarily abelian. Given $g \in G$ and a subgroup $H \subseteq G$, we can define the *left coset* $g \cdot H := \{g \cdot h \mid h \in H\}$ and the *right coset* $H \cdot g := \{h \cdot g \mid h \in H\}$. As g varies, these left cosets will partition G, and the right cosets will also partition G. Now suppose we study the left cosets by asking ourselves the following question:

> Take an element a from a left coset A and multiply it by an element b from another left coset B. Which left coset will $a \cdot b$ be inside? Does it depend on the elements from A and B that we chose? If not, then this recipe defines an operation on the set of left cosets, which would be great!

The sad truth, which we learned in group theory, is that in general the left coset containing $a \cdot b$ will depend on exactly which a and b we pick from A and B. For example, take $a' = a \cdot h_a$ and $b' = b \cdot h_b$, so $a' \in A$ and $b' \in B$. Then

$$a' \cdot b' = a \cdot h_a \cdot b \cdot h_b$$

and as hard as we might try, we can't simplify the right hand side to prove that it's in the same left coset as ab. However, if we knew that there is some $h'_a \in H$ such that $h_a \cdot b = b \cdot h'_a$, then

$$a' \cdot b' = a \cdot h_a \cdot b \cdot h_b = a \cdot b \cdot (h'_a \cdot h_b)$$

would indeed belong to the same left coset as $a \cdot b$. This kind of reasoning leads to following fact in group theory:

Fact: For a subgroup H of G, the following three statements are equivalent

1. For every $h \in H$ and $g \in G$, there is some $h' \in H$ such that $hg = gh'$.
2. The partition of G into left cosets of H equals the partition of G into right cosets of H.
3. The operation $[a] \cdot [b] = [a \cdot b]$ on left (and right) cosets is well-defined.

A subgroup H is *normal* if the above (equivalent) conditions are true; in this case, the left cosets are the same as the right cosets (so we just use the word "cosets" to refer to both), and we call the group of cosets the *quotient group* of G by H, written G/H.

Another way your professor might have motivated the study of normal subgroups is by proving that for any group homomorphism $\varphi : G \to G'$, the kernel $\ker(\varphi)$ isn't just any subgroup – it has the nice property that for any $g \in G$ and $h \in \ker(\varphi)$, there is another $h' \in \ker(\varphi)$ such that $h \cdot g = g \cdot h'$. But this is exactly the property of being a *normal* subgroup that we saw in our discussion of "what makes the group operation on cosets well-defined".

These two ways of motivating the property of normality (as the condition that ensures that the cosets form a group, or as a curious property of kernels of homomorphisms) may seem different, but are actually related by the first isomorphism theorem of groups which states that for any homomorphism φ,

$$\frac{G}{\ker(\varphi)} \cong \mathrm{Im}(\varphi)$$

This isomorphism is one of the most conceptually important facts in all of group theory; it should be part of your subconscious intuition about homomorphisms and normal subgroups by now.

Now let's do the same thought experiment with rings.

Ideals: the "Normal Subgroups of Ring Theory"

Suppose S is a subring of a ring R. R is an abelian group with respect to the addition operation, so because every subgroup of an abelian group is normal, S is a normal subgroup of R. Therefore the operation

$$[a] + [b] = [a + b]$$

is well-defined, and R/S is a group. But what about multiplication? Can we multiply cosets to make R/S into a *ring*? To answer this, we first consider the following question

> Take an element a from a coset A and an element b from another coset B. We already know that if we add a and b, the coset containing $a+b$ doesn't depend on which elements from A and B that we chose. But now what if we *multiply* the elements a and b? Will the coset of ab depend on which elements that we chose? If not, then the operation
>
> $$[a][b] = [ab]$$
>
> would be well-defined!

The sad truth is that in general the coset containing ab will depend on exactly which a and b we pick from A and B. For example, take $a' = a + s_a$ and $b' = b + s_b$, so $a' \in A$ and $b' \in B$ (remember, the group operation that we're using to define the cosets is addition... this is why we've written $a' = a + s_a$ not $a' = as_a$). Then

$$a'b' = ab + as_b + s_a b + s_a s_b$$

If we knew that $as_b \in S$ and $s_a b \in S$, then $a'b'$ would be in the same coset as ab. One condition on the subring S that causes this property to be true leads to the definition of an *ideal*

Definition: Let S be a subring of R. If, for any elements $s \in S$ and $r \in R$, we have $rs \in S$ and $sr \in S$, then S is an **ideal** of R.

Notice that the definition of a subring already guarantees that if $s, r \in S$, then rs and sr are in S. The condition of being an ideal is stronger: it says that when you multiply an element of S by *any* element (even one outside of S), you get something inside S. You can think that an ideal "absorbs" the ring into the ideal. Check that the following are ideals:

Example 0: In any ring R, the zero ideal (consisting of just the element 0) and the whole ring R are both ideals of R.

Example 1: $n\mathbb{Z} \subseteq \mathbb{Z}$ is an ideal.

Example 2: In $\mathbb{R}[x]$, the subring of polynomials that have zero constant term is an ideal.

Example 3: In $C(\mathbb{R})$, the set of continuous functions that vanish at $x = 9$ is an ideal.

Relationship between ideals and quotient rings

Proposition: If $I \subseteq R$ is an ideal, then R/I is a ring, with addition and multiplication of cosets defined by

$$[a] + [b] := [a+b] \qquad \text{and} \qquad [a][b] := [ab]$$

Proof: The discussion above proves that the addition and multiplication operations are well-defined. Also, because $I \subseteq R$ is a normal subgroup, we know from group theory that R/I is a group with respect to the addition operation defined above. It is abelian because

$$[a] + [b] = [a+b] = [b+a] = [b] + [a]$$

The fact that multiplication is associative is shown by

$$[a]([b][c]) = [a][bc] = [a(bc)] = [(ab)c] = [ab][c] = ([a][b])[c]$$

and the fact that distribution holds is given by

$$[a]([b]+[c]) = [a][b+c] = [a(b+c)] = [ab+ac] = [a][b] + [a][c]$$

(the fact that $([b]+[c])[a] = [b][a] + [c][a]$ is proved in a similar way).

Proposition: Let $\varphi : R \to S$ be a ring homomorphism. Then $\ker(\varphi)$ is an ideal in R, and

$$\frac{R}{\ker(\varphi)} \cong \operatorname{Im}(\varphi)$$

Proof: We already proved that $\ker(\varphi)$ is a subring; now, we must verify that it is an ideal. Let $r \in R$ and $k \in \ker(\varphi)$. The fact that $kr, rk \in \ker(\varphi)$ is shown by

$$\begin{aligned} \varphi(rk) &= \varphi(r)\varphi(k) = \varphi(r)(0) = 0 \\ \varphi(kr) &= \varphi(k)\varphi(r) = (0)\varphi(r) = 0. \end{aligned}$$

To prove the isomorphism, define the map

$$\begin{aligned} \frac{R}{\ker(\varphi)} &\to \operatorname{Im}(\varphi) \\ [r] &\mapsto \varphi(r) \end{aligned}$$

This is well-defined, since $\varphi(r + k) = \varphi(r) + \varphi(k) = \varphi(r)$ for any $k \in \ker(\varphi)$.

The kernel of this map is the set of cosets $[r]$ for which $\varphi(r) = 0$. This is just the zero coset, so the kernel is zero and therefore the map is injective.

The fact that the map is surjective is seen by taking an arbitrary element s of $\operatorname{Im}(\varphi)$, so $s = \varphi(r)$ for some $r \in R$, and observing that $[r]$ is sent to $\varphi(r) = s$ under the map above. Therefore, the above map is injective and surjective, hence an isomorphism.

Question: Determine whether the following subsets of $\mathbb{Z} \times \mathbb{Z}$ are subrings, ideals, or neither. If it is an ideal I, describe $(\mathbb{Z} \times \mathbb{Z})/I$.

- $\{(a, a) \mid a \in \mathbb{Z}\}$.
- $\{(2a, 2b) \mid a, b \in \mathbb{Z}\}$.
- $\{(2a, 0) \mid a \in \mathbb{Z}\}$.
- $\{(a, -a) \mid a \in \mathbb{Z}\}$.

Question: Determine whether the following subsets of $\mathbb{Z}[x]$ are subrings, ideals, or neither. If it is an ideal I, describe $\mathbb{Z}[x]/I$.

- All polynomials whose constant term is a multiple of 5.
- All polynomials whose coefficient of x^5 is 5.
- All polynomials whose constant term is 0, and whose coefficient of x is 0.
- All polynomials p such that $p(7) = 0$.
- All polynomials p such that $p(7) = 1$.
- All polynomials where only even powers of x appear (this is sometimes written $\mathbb{Z}[x^2]$).
- All polynomials whose coefficients sum to zero.
- All polynomials whose coefficients are divisible by 3.

Properties of ideals: a preview

Next class, we will learn about all sorts of properties of ideals. I want to state just one today.

Proposition: Let I be an ideal in a ring R. If I contains a unit, then $I = R$.

Proof: Let $u \in I$ be a unit. This means that u has a multiplicative inverse $u^{-1} \in R$. Then for any $a \in R$, we see that $a = au^{-1}u$, so $a \in I$.

Corollary: The only ideals in a field F are $I = \{0\}$ and $I = R$.

Lecture 3: Ideals

New Ideals from Old

If we have two ideals I, J in ring R, we can construct new ideals related to I and J.

Definition: Let I and J be ideals in R.

- $I + J := \{a + b \mid a \in I, b \in J\}$.
- IJ consists of all finite sums of elements of the form ab, where $a \in I$, $b \in J$.
- $I \cap J$ is the intersection of I and J.

Question: Let $I = 4\mathbb{Z}$ and $J = 6\mathbb{Z}$, what are the ideals $I + J$, IJ, and $I \cap J$? Explain why $I \cup J$ is not an ideal.

Question: What are the containment relationships between $I, J, I + J$, and $I \cap J$?

There is also a way to construct an ideal out of any subset of elements of S.

Definition: Let A be a subset of the ring R. The **ideal generated by** A is the smallest ideal of R containing A, and is written (A). An ideal generated by a single element is called a **principal ideal**; an ideal generated by finitely many elements is a **finitely generated ideal**.

Sometimes the definition of (A) is hard to work with; here are other ways to think about (A).

- (A) is the intersection of all ideals containing the subset A.
- If R is commutative, then (A) consists of all elements of R that can be written as

$$r_1a_1 + ... + r_na_n$$

where the r_i's and a_i's are elements of R and A, respectively.

Question 1: In the ring $\mathbb{Z}$, describe (5) and $(6, 4)$.

Question 2: In the ring $\mathbb{Z}[x]$, describe $(x), (5)$, and $(5, x)$.

Question 3: Find generating sets for the kernels of the following homomorphisms.

- $\mathbb{Z}[x] \to \mathbb{Z}$ given by $a_0 + a_1x + a_2x^2 + ... \mapsto a_0$.
- $\mathbb{Z}[x] \to (\mathbb{Z}/3\mathbb{Z})[x]$ given by $a_0 + a_1x + a_2x^2 + ... \mapsto [a_0] + [a_1]x + [a_2]x^2 + ...$
- $\mathbb{Z}[x] \to \mathbb{Z}/3\mathbb{Z}$ by $a_0 + a_1x + a_2x^2 + ... \mapsto [a_0]$.

Types of Ideals

On the first day of class, we learned words to describe rings that had certain nice properties.

1. In a *field* (commutative ring with identity where every nonzero element is a unit), we can divide by nonzero elements.

2. In an *integral domain* (commutative ring with identity having no zero divisors), we can't always divide but we have the next best thing, the cancellation property: If $a \neq 0$ and $ab = ac$, then $b = c$.

Last class, we constructed quotient rings R/I from ideals $I \subseteq R$. Now, we'll see that certain properties of the ideal translate into other properties of the quotient ring. Specifically, we can determine whether R/I will be a field or an integral domain just by studying I.

When will R/I be a field?

To answer this question, we first classify fields as the commutative rings with identity that have no nontrivial proper ideals, then prove that the ideals of a quotient ring are precisely the ideals containing I. This will motivate the definition of *maximal* ideal as an ideal which is not contained in any ideal except R and itself: it is precisely these *maximal* ideals of a commutative ring with identity whose quotient R/I is a field.

Proposition: A commutative ring with identity R is a field if and only if its only ideals are (0) and R.

Proof: Because every nonzero element of a field is a unit, and every ideal that contains a unit is the entire ring, it follows that the only ideals of a field are (0) and the entire field. Conversely, let R be a ring such that the only ideals are 0 and R. Then for every nonzero $x \in R$, the ideal (x) must be all of R. This means that $1 = rx$ for some $r \in R$, so x is invertible.

The next proposition describes the ideals of a quotient ring.

Proposition: Let $\varphi : R \to S$ be a ring homomorphism. If I is an ideal in S, then $\varphi^{-1}(I)$ is an ideal in R.

Proof: Let $a, b \in \varphi^{-1}(I)$. Using the definition of homomorphisms, $\varphi(a+b), \varphi(ab), \varphi(-a) \in I$, so $\varphi^{-1}(I)$ is a subring. Now let $r \in R$, $j \in \varphi^{-1}(I)$, and notice that $\varphi(rj) = \varphi(r)\varphi(j) \in I$ since I is an ideal and $\varphi(j) \in I$. A similar computation shows $\varphi(jr) \in I$, so $rj, jr \in \varphi^{-1}(I)$, proving that $\varphi^{-1}(I)$ is an ideal.

In the above proposition, because every ideal of S contains $0 \in S$, it follows that every ideal of R of the form $\varphi^{-1}(I)$ contains $\ker(\varphi)$.

Proposition: Let I be an ideal of a ring R, and let $\varphi : R \to R/I$ be the quotient map. The maps below are inverses (and hence define a bijection).

$$\left\{ \begin{array}{c} \text{Ideals in} \\ \text{R that contain I} \end{array} \right\} \leftrightarrow \left\{ \begin{array}{c} \text{Ideals} \\ \text{in } R/I \end{array} \right\}$$

$$J \mapsto \varphi(J)$$

$$\varphi^{-1}(K) \mapsfrom K$$

Proof: First, we verify that $\varphi(J)$ is indeed an ideal in R/I.

$\varphi(J)$ **is a subring:** This follows from the fact that φ is a homomorphism (check this!), and $\varphi(J)$ is the image of the homomorphism $\varphi\big|_J$, and we know that images of homomorphisms are subrings.

$\varphi(J)$ **is an ideal:** Any element of $\varphi(J)$ can be written as $[j]$ for some $j \in J$, and let $[s] \in R/I$. Then $[j][s] = [js]$ and $[s][j] = [sj]$ are also inside $\varphi(J)$ (since $js, sj \in J$), so $\varphi(J)$ is an ideal.

So the map is well-defined. We also know (from the above proposition) that for any ideal $K \in R/I$, $\varphi^{-1}(K)$ is an ideal. To complete the proof it suffices to show that $\varphi(\varphi^{-1}(K)) = K$ for all ideals K in R/I, and $\varphi^{-1}(\varphi(J)) = J$ for all ideals J in R containing I (so that $K \mapsto \varphi^{-1}(K)$ and $J \mapsto \varphi(J)$ are inverse maps). The first claim follows from the fact that φ is surjective. To prove the second, notice that if $\varphi(a) = \varphi(j)$ for some $j \in J$, then $a - j \in I \subseteq J$, so $a \in J$. Therefore, if $\varphi(a) \in \operatorname{im}(J)$, then $a \in J$. This proves that $\varphi^{-1}(\varphi(J)) = J$.

Combining this with the above, we see that a quotient ring R/I of a commutative ring with identity is a field precisely when there are no ideals "in between" I and R.

Definition: An ideal I of R is called **maximal** if there are no ideals J satisfying $I \subsetneq J \subsetneq R$

Corollary: Let I be an ideal in a commutative ring R with identity. Then R/I is a field if and only if I is maximal.

When will R/I be an integral domain?

Let R be a commutative ring with identity. In an integral domain, a product ab is zero only if a or b is zero. In a quotient ring R/I, the zero element is the coset I itself. Unsurprisingly, the condition that R/I is an integral domain translates into the condition that $ab \in I$ only if a or b is in I.

Definition: An ideal I of a commutative ring R is *prime* if $I \neq R$ and whenever $ab \in I$, then either $a \in I$ or $b \in I$.

Proposition: Let I be an ideal in a commutative ring R with identity. Then R/I is an integral domain if and only if I is prime.

Proof: In the ring R/I, then zero element is the coset $[0] = I$. In other words, $[a] = 0$ if and only if $a \in I$. After this observation, the proof is a matter of following the definitions:

$$\begin{aligned} R/I \text{ is an integral domain} &\iff [a][b] = 0 \text{ only if } [a] = 0 \text{ or } [b] = 0 \\ &\iff ab \in I \text{ only if } a \in I \text{ or } b \in I \\ &\iff I \text{ is a prime ideal} \end{aligned}$$

Lecture 4: Fractions and the Chinese Remainder Theorem

Today, we discuss two unrelated topics. The first topic is a way to *enlarge* a ring by introducing new elements that act as multiplicative inverses of non-units, just like making fractions of integers enlarges $\mathbb{Z}$ into $\mathbb{Q}$. The second topic will be the first substantial theorem of the course, the chinese remainder theorem.

Fractions

Most integers don't have integer inverses. This is annoying, because it means we can't even solve linear equations like $2x = 5$ in the ring $\mathbb{Z}$. Of course, one way to resolve this issue is to work in $\mathbb{Q}$ instead, where the solution is $x = 5/2$. Because $\mathbb{Q}$ is a field, any equation $ax = b$ with $a, b \in \mathbb{Q}$ and $a \neq 0$ will always have a solution, namely b/a, because we find by dividing both sides of the equation by a. But what if we want to solve $ax = b$ in a general ring R (so $a, b \in R$)? Can we always find a field that contains R so that we can divide by elements?

Let R be a commutative ring with identity. If we try to enlarge R into a field in the most naive way possible, using as inspiration the way we enlarge $\mathbb{Z}$ into $\mathbb{Q}$, we might try the following definition.

Naive definition 1: Let R' be the ring whose elements are of the form a/b, where $a, b \in R$. Addition is given by $a/b + c/d = (ad + bc)/bd$, and multiplication is given by $(a/b)(c/d) = ac/bd$.

Sad fact: This isn't even a ring because it fails to be a group under the addition operation: the additive identity element would need to be $0/1$, and there's no way to find an inverse for a/b ($(-a)/b$ is not an inverse because $a/b + (-a)/b = 0/(b^2)$).

Your reaction to the failure of this definition is probably "Oh, of course you need some condition saying that $0/(b^2)$ is the same as $0/1$. In $\mathbb{Q}$, two fractions a/b and c/d are the same number when $ad = bc$; let's modify our definition so that the objects are *equivalence classes* of fractions under this relation."

Naive definition 2: Let R' be the ring whose elements are equivalence classes of fractions a/b, where a/b is equivalent to c/d if $ad = bc$. Addition is given by $a/b + c/d = (ad + bc)/bd$, and multiplication is given by $(a/b)(c/d) = ac/bd$.

Sad fact: Under this equivalence relation, *everything* is equivalent to the fraction $0/0$, so R' only has a single element.

Again taking our inspiration from fractions of integers, we might decide to prohibit 0 from being a denominator.

Naive definition 3: Let R' be the ring whose elements are equivalence classes of fractions a/b, $b \neq 0$, where a/b is equivalent to c/d if $ad = bc$. Addition is given by $a/b + c/d = (ad + bc)/bd$, and multiplication is given by $(a/b)(c/d) = ac/bd$.

Sad fact: This won't (in general) be an enlargement of R, because it won't (in general) contain R as a subring. Suppose, for example, that R contains zero divisors, so $ab = 0$ for some nonzero a, b. Then $a/1 = 0/b$, and $0/b = 0/1$, so $a/1 = 0/1$. We were hoping that the elements $\{r/1 \mid r \in R\}$ would be a copy of R inside R' and we have failed, because $a/1 = 0/1$ for any zero divisor a.

Somehow, the presence of zero divisors in our ring prevents us from using this fraction construction. In retrospect, this sort of makes sense – in a field, you *never* have zero divisors. And when you “enlarge” a ring, the zero divisors will still be zero divisors in the enlarged ring. So we were naive to think we could ever enlarge a ring with zero divisors into a field. However, for integral domains (which don’t have zero divisors), we can.

Definition: Let R be an integral domain. The **field of fractions of** R is the field $\mathcal{F}$ consisting of equivalence classes of pairs (a, b) of elements of R, $b \neq 0$, where $(a, b) \sim (c, d)$ if $ad = bc$. Normally, an equivalence class $[(a, b)]$ is written a/b. Addition and multiplication on $\mathcal{F}$ are defined by

$$\begin{aligned} \frac{a}{b} + \frac{c}{d} &= \frac{ad + bc}{bd} \\ \frac{a}{b}\frac{c}{d} &= \frac{ac}{bd}. \end{aligned}$$

Remark: There is a lot to check in the above definition: the equivalence relation is indeed an equivalence relation, the addition and multiplication operations are well-defined and make $\mathcal{F}$ into a ring, and that the ring is indeed a field.

We motivated the construction of the field of fractions as a way to “enlarge” a ring into a field; the next proposition verifies that the field of fractions is indeed an “enlargement” of R.

Proposition: The map φ from an integral domain to its field of fractions given by

$$r \mapsto \frac{r}{1}$$

is an injective homomorphism.

Proof: To check that φ is a homomorphism:

$$\begin{aligned} \varphi(r + s) &= \frac{r + s}{1} = \frac{1r + 1s}{1 \cdot 1} = \frac{r}{1} + \frac{s}{1} = \varphi(r) + \varphi(s) \\ \varphi(rs) &= \frac{rs}{1} = \frac{r}{1} \cdot \frac{s}{1} = \varphi(r)\varphi(s) \end{aligned}$$

To check that φ is injective:

$$\begin{aligned} \ker(\varphi) &= \left\{ r \in R \mid \frac{r}{1} = \frac{0}{1} \right\} \\ &= \{r \in R \mid r \cdot 1 = 0 \cdot 1\} = \{0\} \end{aligned}$$

We can generalize this construction. The only reason we need R to be an integral domain is that strange things happen when you let zero divisors be denominators of fractions. So, even when R is *not* an integral domain, we can pick a subset $D \subseteq R$ of “things that are allowed to be denominators,” and consider the ring of elements of the form a/b, where $a \in R$ and $b \in D$. As long as D contains no zero divisors, D is closed under multiplication (so that the addition and multiplication rules make sense), and D doesn’t contain zero, the construction still works.

Definition: Let R be a commutative ring with identity, and let D be a nonempty subset of R such that $0 \notin D$, D contains no zero divisors, and D is closed under multiplication. The ring $D^{-1}R$ is the ring consisting of equivalence classes of pairs (a, b), where $a \in R$ and $b \in D$, and $(a, b) \sim (c, d)$ if $ad = bc$. Addition and multiplication are defined by

$$(a, b) + (c, d) = (ad + bc, bd)$$
$$(a, b)(c, d) = (ac, bd).$$

Normally, an element (a, b) is written a/b.

If R is an integral domain, its field of fractions is the same as $D^{-1}R$ where $D = R\backslash\{0\}$.

We were motivated to study fractions so that linear equations would have solutions. But what about equations involving higher-order polynomials? In fact, there *is* a way to start with a field F and a polynomial with coefficients F, and enlarge F in such a way that the polynomial is *forced* to have a root in the enlarged field. This is the beautiful theory of field extensions, which we will study later in the course.

Chinese Remainder Theorem

Consider the following riddle (allegedly first posed by Brahmagupta, a 7th century Indian mathematician):

> An old woman is carrying eggs to sell in the market when the eggs are crushed by a horse. The horseman wants to know how many eggs she had so that he can repay her, but she only knows the following: when she arranged them into groups of 2 or 3 or 4 or 5 or 6, there was always precisely one left over. But when she arranged them into groups of 7, there were none left over. What is the smallest number of eggs she might've had?

One way to express this problem mathematically is the following:

> Consider the homomorphism
>
> $$\mathbb{Z} \to \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_4 \times \mathbb{Z}_5 \times \mathbb{Z}_6 \times \mathbb{Z}_7$$
>
> given by the product of the quotient homomorphisms $\mathbb{Z} \to \mathbb{Z}_i$. What is the smallest number in the preimage of $([1], [1], [1], [1], [1], [0])$?

The topic of this section, the chinese remainder theorem, does *not* answer this riddle, but it tells us important information about the kernel and image of the homomorphism. In fact, it does it in the more general setting of *any* product of quotient homomorphisms from *any* commutative ring with unity R

$$R \to \frac{R}{I_1} \times \frac{R}{I_2} \times \cdots \times \frac{R}{I_n}$$

not just the case when $R = \mathbb{Z}$. Here's why we might care about the kernel and image of this map using the egg riddle as motivation.

1. There is no way to determine the *exact* number of eggs the woman was carrying because there are infinitely many elements in the preimage. For example, the number $2 * 3 * 4 * 5 * 6 * 7 = 5040$ is in the kernel of this homomorphism, so if you find any element in the preimage, you can add 5040 to get another element in the preimage. But is (5040)

the entire kernel? The chinese remainder theorem will tell us exactly what the kernel is (in our case, (420)), which answers the question "Up to what additive factor can you determine an element of $\mathbb{Z}$ just by knowing its image in $\mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_4, \mathbb{Z}_5, \mathbb{Z}_6$, and $\mathbb{Z}_7$?"

2. The old woman had redundant information: if you group eggs into groups of four and have one left over, then you already know that if you group the eggs into groups of two, you'll have one left over. The old woman wasted time in making the groups of two. This inefficiency in the woman's choice of grouping sizes is reflected in the fact that the homomorphism is not surjective. For example $([1],[0],[0],[0],[0],[0])$ is not in the image. If the woman wanted to be efficient and guarantee that none of her information is redundant, she should have chosen her "egg grouping sizes" to be pairwise coprime. In the more general ring-theory context, the condition of being coprime translates into the condition that the ideals I_k are pairwise *comaximal*.

Definition: Ideals I and J in a ring R are *comaximal* if $I + J = R$

In other words, the chinese remainder theorem will answer the following riddle.

> An old woman is carrying an element of a commutative ring to sell in the market when the ring element gets crushed by a horse. The woman forgets exactly which ring element it was, but knows which coset it belongs to modulo the comaximal ideals $I_1, I_2, \ldots, I_k$. Exactly how much information does this determine about the original ring element?

Theorem (Chinese Remainder Theorem): Let $I_1, \ldots, I_n$ be ideals in a commutative ring R with identity, and consider the homomorphism

$$\begin{aligned} R &\to \frac{R}{I_1} \times \frac{R}{I_2} \times \cdots \times \frac{R}{I_n} \\ r &\mapsto ([r],[r],\ldots,[r]) \end{aligned}$$

(the first bracket signifies a coset of I_1, the second is a coset of I_2, etc.)

1. The kernel equals $I_1 \cap I_2 \cap \cdots \cap I_n$
2. If the ideals are pairwise comaximal (i.e., I_j and I_ℓ are comaximal when $j \neq \ell$), then the map is surjective and

$$I_1 \cap I_2 \cap \cdots \cap I_n = I_1 I_2 \ldots I_n$$

so the first isomorphism theorem gives

$$\frac{R}{I_1 I_2 \ldots I_n} \cong \frac{R}{I_1} \times \frac{R}{I_2} \times \cdots \times \frac{R}{I_n}.$$

Proof: To prove the first claim, observe that $r \in R$ is in the kernel precisely when r is a representative of the zero coset with respect to each ideal. But this is just another way of saying that r is contained in each ideal, so the kernel is $I_1 \cap \cdots \cap I_n$.

We prove the second claim only in the $n = 2$ case (the full proof has an extra inductive step which is not very educational).

Case $n = 2$: Let I, J be two comaximal ideals. To show that the homomorphism is surjective, pick $i \in I$ and $j \in J$ satisfying $i + j = 1$ (this is possible by the comaximality condition). If we reduce both sides of this equation modulo i, we see that $[j] = [1]$ modulo i, and also by reducing modulo j, we see that $[i] = [1]$ modulo j. This shows that

$$\begin{aligned} R &\to R/I \times R/J \\ j &\mapsto ([1], [0]) \\ i &\mapsto ([0], [1]) \end{aligned}$$

So in particular, any element $([r_i], [r_j]) \in R/I \times R/J$ is the image of $r_i j + r_j i$. This proves surjectivity.

To show that $I \cap J = IJ$, notice that because everything in IJ is both in I and also in J, the inclusion $I \cap J \supseteq IJ$ is clear. To see the opposite inclusion, let $r \in I \cap J$. Then $r = r1 = r(i + j) = ir + rj$. Because $r \in J$, it follows that $ir \in IJ$. And because $r \in I$, it follows that $rj \in IJ$. Therefore, $ir + rj = r \in IJ$.

Lecture 5: $\mathbb{R}[x]/(x^2+1)$ and Polynomial Long Division

Study of $(x^2+1) \subseteq \mathbb{R}[x]$

Question: Describe (in words) the ideal $I = (x^2+1)$ in $\mathbb{R}[x]$ and give several examples of different elements of I.

Question: Find several different coset representatives of $[0] \in \mathbb{R}[x]/I$ (sometimes written $\overline{0}$) and $[1] = \overline{1} \in \mathbb{R}[x]/I$.

Helpful tip: In the ring $\mathbb{R}[x]/I$, the element $[x^2]$ is *the same element* as $[-1]$. This means we can "simplify" any element of $\mathbb{R}[x]$ by replacing instances of x^2's with -1's. For example,

$$\begin{aligned}[3x^4 + x^3 - x^2 + x + 9] &= [3][x^2][x^2] + [x][x^2] - [x] + [9] \\ &= [3][-1][-1] + [x][-1] - [x] + [9] \\ &= [3 - x - x + 9] \\ &= [-2x + 12]\end{aligned}$$

Note: this works for more complicated ideals too. For example, in the ring $\mathbb{R}[x]/(x^3 - 2x^2 + x + 9)$, the element $[x^3]$ equals the element $[2x^2 - x - 9]$, so $[x^3 + x + 1] = [2x^2 - 8]$.

This tip tells us that every element in $\mathbb{R}[x]/I$ can be written in the form $[ax+b]$. If you multiply two elements of $\mathbb{R}[x]/I$ that are written in this form, we get

$$[ax+b][cx+d] = [(ax+b)(cx+d)] = [acx^2 + (ad+bc)x + bd] = [(ad+bc)x + (bd-ac)]$$

This should remind us of multiplication of complex numbers, $(ai+b)(ci+d) = (ad+bc)i + (bd-ac)$. Let's prove that $\mathbb{R}/I$ is isomorphic to $\mathbb{C}$.

Claim: The map

$$\begin{gathered}\varphi : \mathbb{R}[x] \to \mathbb{C} \\ a_0 + a_1 x + \cdots + a_n x^n \mapsto a_0 + a_1 i + \cdots + a_n i^n\end{gathered}$$

is a surjective homomorphism with kernel (x^2+1).

Proof:

Homomorphism: For a polynomial $p \in \mathbb{R}[x]$, $\varphi(p)$ is the number you get when you evaluate p at the complex number i, $p(i)$. Then

$$\begin{gathered}\varphi(p+q) = (p+q)(i) = p(i) + q(i) = \varphi(p) + \varphi(q) \\ \varphi(pq) = (pq)(i) = p(i)q(i) = \varphi(p)\varphi(q)\end{gathered}$$

so φ is a homomorphism.

Surjective: The complex number $ai + b$ is the image of $ax + b \in \mathbb{R}[x]$.

Kernel is (x^2+1): Because $\varphi(x^2+1) = i^2+1 = 0$, it is clear that $x^2+1 \in \ker(\varphi)$, and therefore that $(x^2+1) \subseteq \ker(\varphi)$. To show the opposite inclusion, suppose $p \in \ker(\varphi)$. By dividing p by (x^2+1) using polynomial long division, we can write $p = (x^2+1)q+r$ for some polynomials q and r with $\deg(r) < 2$. Then $0 = \varphi(p) = p(i) = r(i)$, but the only polynomial of degree < 2 with $r(i) = 0$ is the zero polynomial, so $r = 0$. This proves that $p = (x^2+1)q$, so $p \in (x^2+1)$.

Corollary: $\mathbb{R}[x]/(x^2+1) \cong \mathbb{C}$

Comments:

- The fact that $\mathbb{R}[x]/(x^2+1)$ is isomorphic to a field tells us that $\mathbb{R}[x]/(x^2+1)$ is itself a field. Therefore (x^2+1) is maximal. This is related to the fact that you cannot factor x^2+1. (to understand this last sentence, try to prove that (x^2-1) is *not* maximal!).
- Notice that x^2+1 is a polynomial with coefficients in $\mathbb{R}$ and no roots in $\mathbb{R}$. By taking the quotient of $\mathbb{R}[x]$ by (x^2+1), we enlarged $\mathbb{R}$ to a new field $\mathbb{C}$ in which x^2+1 has roots. Today, this should seem magical. But when we study field extensions in a few weeks, we will understand on a deep level why this construction works and generalize it.

Polynomial long division

You learned the following fact in secondary school:

> **Observation:** When you divide a polynomial $f \in \mathbb{R}[x]$ by $g \in \mathbb{R}[x]$, the degree of the remainder polynomial is less than $\deg(g)$.

Example: Divide $3x^4 + 2x^3 + 4 \in \mathbb{R}[x]$ by $2x^2 + 1 \in \mathbb{R}[x]$ using polynomial long division.

The next theorem is a mathematically precise formulation of the above well-known fact.

Theorem: Let $f, g \in \mathbb{R}[x]$ with $g \neq 0$. There are unique polynomials $q, r \in \mathbb{R}[x]$ such that

$$f = qg + r$$

and either $r = 0$ or $\deg(r) < \deg(g)$.

Proof: To prove existence, note if $\deg(f) < \deg(g)$, then $q = 0, r = f$. Otherwise, write

$$f = a_n x^n + \text{lower order terms}$$
$$g = b_k x^k + \text{lower order terms}$$

Then f and $(a_n b_k^{-1} x^{n-k})g$ have the same leading term, so $r_1 = f - (a_n b_k^{-1} x^{n-k})g$ has smaller degree than f. We can write

$$f = (a_n b_k^{-1} x^{n-k})g + r_1$$

If $\deg(r_1) < \deg(g)$, then we're done. Otherwise, we repeat the process on r_1: write

$$r_1 = c_m x^m + \text{lower order terms}$$
$$g = b_k x^k + \text{lower order terms}$$

and

$$r_1 = (c_m b_k^{-1} x^{m-k})g + r_2, \quad \text{so}$$
$$f = (a_n b_k^{-1} x^{n-k})g + (c_m b_k^{-1} x^{m-k})g + r_2 = (a_n b_k^{-1} x^{n-k} + c_m b_k^{-1} x^{m-k})g + r_2$$

We can repeat the process until $\deg(r_i) < \deg(g)$, proving existence. To prove uniqueness, suppose

$$f = qg + r \qquad \text{and} \qquad f = \tilde{q}g + \tilde{r}$$

are two such expressions for f. Then

$$g(q - \tilde{q}) = \tilde{r} - r$$

Because the right hand side cannot be a polynomial of degree $\geq \deg(g)$, it must be the case that $q - \tilde{q} = 0$ and therefore that $\tilde{r} - r = 0$.

In the first half of today's class, we found a way to enlarge the field $\mathbb{R}$ by taking the quotient of its polynomial ring $\mathbb{R}[x]$ by a maximal ideal. We'll need to do this with fields that aren't necessarily $\mathbb{R}$ as well, but first we need to understand how polynomials behave when their coefficients come from some field other than $\mathbb{R}$. If you look at the proof above, we never used the fact that the field was $\mathbb{R}$! So the proof generalizes immediately to the following

Theorem: Let F be a field, and $f, g \in F[x]$ with $g \neq 0$. There are unique polynomials $q, r \in F[x]$ such that

$$f = qg + r$$

and either $r = 0$ or $\deg(r) < \deg(g)$.

Proof: The proof is the same as when $f, g \in \mathbb{R}[x]$.

Example: Divide $3x^4 + 2x^3 + 4 \in \mathbb{Z}_5[x]$ by $2x^2 + 1 \in \mathbb{Z}_5[x]$ using polynomial long division.

Practice problems (optional – do not hand in)

Gallian 16-1: Let $f(x) = 4x^3 + 2x^2 + x + 3$ and $g(x) = 3x^4 + 3x^3 + 3x^2 + x + 4$, where $f, g \in \mathbb{Z}_5[x]$. Compute $f + g$ and fg.

Gallian 16-2: In $\mathbb{Z}_3[x]$, show that $x^4 + x$ and $x^2 + x$ determine the same function (that is, $f(a) = g(a)$ for all $a \in \mathbb{Z}_3$)

Gallian 16-3: Let $f(x) = x^2 + 3x + 2 \in \mathbb{Z}_6[x]$. Show that there are four elements $a \in \mathbb{Z}_6$ such that $f(a) = 0$. Wednesday, we will learn that for any polynomial nonzero $f \in F[x]$ where F is a field (which $\mathbb{Z}_6$ is not), the number of elements $a \in F$ where $f(a) = 0$ cannot be larger than the degree of f.

Question: Find a way to write $2x^2 + 4$ as a product of two degree-one polynomials in $\mathbb{C}[x]$. Is it possible to write it as a product of degree-one polynomials in $\mathbb{R}[x]$?

Lecture 6: Irreducibility of Polynomials

Consequences of long division

Last class, we talked about how the polynomial long division algorithm works for polynomials with coefficients in *any* field, not just real numbers.

Recall: If F is a field and $f, g \in F[x]$ with $g \neq 0$, then there are unique $q, r \in F[x]$ such that $f = gq + r$ and either $r = 0$ or $\deg(r) < \deg(g)$. The polynomial r is called the **remainder term**.

Let's remind ourselves of a few definitions

Definition: Let $f, g \in D[x]$, where D is an integral domain. We say that g **divides** f, or that g **is a factor of** f, if $f = gq$ for some $q \in D[x]$. We say that $a \in D$ is a **root** of f if $f(a) = 0$.

Remember that every field is an integral domain, so these definitions apply to fields too. The division algorithm has some interesting corollaries. In all of the below, F is a field and $f \in F[x]$.

Corollary 1: For any $a \in F$, the remainder term when you divide f by $(x - a)$ is the constant polynomial $f(a)$.

Example: Divide $f(x) = x^2 + 3 \in \mathbb{Z}_5[x]$ by $x - 3 \in \mathbb{Z}_5[x]$. Verify that it equals the constant polynomial $f(3)$.

Proof: The divison algorithm gives $f = (x-a)q + r$, where r has degree < 1, so r is a constant polynomial. Evaluating both sides of the polynomial at $x = a$ reveals that $f(a) = 0 + r(a)$. Because r is a constant polynomial, it equals $f(a)$.

Corollary 2: For any $a \in F$, $f(a) = 0$ if and only if $(x - a)$ is a factor of $f(x)$.

Proof: By corollary 1, $f(a) = 0$ if and only if the remainder term when you divide f by $(x-a)$ is zero. This is true if and only if $f = (x - a)q$ for some $q \in F[x]$.

Corollary 3: The number of roots of f is at most $\deg(f)$.

Proof: This is clearly true for degree 1 polynomials, since if $ax + b = 0$, then $ax = -b$ and $x = -ba^{-1}$. Now use induction: assume we know that the number of roots of a degree $n-1$ polynomial has at most $n-1$ roots, and suppose f has degree n. If f has *no* roots, then the statement is obviously true. If f has at least one root c, then $f = (x - c)q$ for a degree $n - 1$ polynomial q. Because every root of f must be either a root of $(x - c)$ (which has one root) or q (which has at most $n - 1$ roots), it follows that f has at most n roots. This completes the induction.

Questions:

- What are the roots of $x^2 - x \in \mathbb{Z}_5[x]$?
- What are the roots of $x^2 - x \in \mathbb{Z}_6[x]$?
- What are the roots of $x^{100} \in \mathbb{R}[x]$?

The second corollary is particularly important: it says that roots of f are "the same thing" as linear factors of f. This also motivates the definition of the *multiplicity* of a root.

Definition: Let $f \in D[x]$, where D is an integral domain, and let $a \in D$ be a root of f. The **multiplicity** of a is the largest number $k \in \mathbb{Z}$ such that $(x-a)^k$ is a factor of f.

Question: What are all the roots of $f(x) = x^5 + 4x^4 + 4x^3 - x^2 - 4x + 1 \in \mathbb{Z}_5[x]$, and what are their multiplicities?

Can we always check whether f factors just by looking for roots? No: consider $x^4 + 2x^2 + 1$ in $\mathbb{R}[x]$. This polynomial is positive for every real number (so it has no roots), but it still factors as $(x^2+1)(x^2+1)$. So even though it has no linear factors (i.e. has no roots), it still has factors of higher degree.

In the same way that prime numbers in $\mathbb{Z}$ act as the "indivisible numbers that make up all other numbers", so too will "irreducible polynomials" act as the "indivisible polynomials that make up all other polynomials."

Irreducible Polynomials

Definition: Let F be a field. A nonconstant polynomial $f(x) \in F[x]$ is called **reducible over** F (or sometimes just "reducible") if $f(x) = g(x)h(x)$ for polynomials $g(x), h(x) \in F[x]$ with smaller degree than $f(x)$, and it is called **irreducible over** F (or just "irreducible") if $f(x)$ cannot be written as a product of two polynomials of smaller degree in $F[x]$.

Question: Let F be a field. What are the irreducible polynomials of degree one in $F[x]$? Can you think of an irreducible polynomial of degree 2 in $\mathbb{R}[x]$? What about $\mathbb{C}[x]$?

Theorem: Let F be a field, and $f(x) \in F[x]$ be a polynomial of degree 2 or 3. $f(x)$ is reducible over $F[x]$ if and only if $f(x)$ has a root in F.

Proof: If $f(x)$ has a root $a \in F$, then $x - a \in F[x]$ is a factor of $f(x)$, so $f(x)$ is reducible.

If $f(x)$ is reducible, then $f(x) = g(x)h(x)$ for some polynomials $g, h \in F[x]$. Because $\deg(f) = \deg(g) + \deg(h)$, then at least one of g and h must have degree 1. Every degree-one polynomial in a field has a root, so f has a root.

Question: What are all monic irreducible polynomials of degree 3 in $\mathbb{Z}_3[x]$?

Question: What are all the irreducible polynomials of degree 3 in $\mathbb{R}[x]$? (hint: it might be easier to answer the question "What are all the irreducible polynomials of degree n in $\mathbb{R}[x]$, where n is odd?")

The huge important result of today is next: a complete classification of which ideals in $F[x]$ are maximal (remember: we care about maximal ideals in $F[x]$ because their quotients will be fields). This classification will take two steps: first, we show that every ideal in $F[x]$ is principal (generated by a single element), then that the maximal ideals are exactly the ones generated by an *irreducible* polynomial.

Definition: An integral domain R is a **principal ideal domain** if every ideal $I \subseteq R$ is principal. (remember: an ideal is *principal* if it is generated by a single element)

Theorem: Let F be a field. $F[x]$ is a principal ideal domain.

Proof: We already know that $F[x]$ is an integral domain because F is an integral domain. Let $I \subseteq F[x]$ be an ideal. If $I = \{0\}$, then $I = \langle 0 \rangle$ which is a principal ideal. If $I \neq \{0\}$, then I contains nonzero elements, so we can let $f(x) \in I$ be an element of minimal degree. We will show that $\langle f(x) \rangle = I$.

Because $f(x) \in I$, it follows that $\langle f(x) \rangle \subseteq I$. To prove that $I \subseteq \langle f(x) \rangle$, let $g(x) \in I$, and by the division algorithm we have $g(x) = f(x)q(x)+r(x)$, where $r = 0$ or $\deg(r) < \deg(f)$. But since $r(x) = f(x)q(x) - g(x)$ is in I, and $f(x)$ is an element of minimal degree, it follows that $r(x) = 0$.

Theorem: Let F be a field, and $f(x) \in F[x]$. The ideal $\langle f(x) \rangle$ is maximal if and only if $f(x)$ is irreducible over F.

Proof: Suppose $\langle f(x) \rangle$ is maximal, and suppose $f(x) = g(x)h(x)$. Then $\langle f(x) \rangle \subseteq \langle g(x) \rangle \subseteq F[x]$, so by maximality of $\langle f(x) \rangle$, either $\langle g(x) \rangle = F[x]$ or $\langle g(x) \rangle = \langle f(x) \rangle$. In the first case, $g(x)$ must have degree zero, so $h(x)$ has the same degree as $f(x)$. In the second case, $g(x)$ has the same degree as $f(x)$, so $h(x)$ has degree zero. We have proven that for every factorization $f(x) = g(x)h(x)$, either $g(x)$ or $h(x)$ has the same degree as $f(x)$, so therefore $f(x)$ is irreducible.

Suppose $f(x)$ is irreducible. Then to show that $\langle f(x) \rangle$ is maximal, suppose that $\langle f(x) \rangle \subseteq I \subseteq F[x]$. Because $F[x]$ is a principal ideal domain, $I = \langle g(X) \rangle$ for some $g(x) \in F[x]$. Then $f(x) = g(x)q(x)$ for some $q(x) \in F[x]$. Because $f(x)$ is irreducible, either $g(x)$ or $q(x)$ is degree zero. In the first case, $\langle g(x) \rangle = I$. In the second case, $g(x)$ is a constant, so $\langle f(x) \rangle = \langle g(x) \rangle$.

Lecture 7: Unique Factorization of Polynomials

We're developing the ring theory we need to start studying field extensions. Two clases ago, we saw an example of how to construct the field $\mathbb{C}$ as the quotient ring $\mathbb{R}[x]/\langle x^2+1\rangle$. This motivated the idea of enlarging a field F by taking the quotient of $F[x]$ by a maximal ideal (the resulting field will indeed contain F as a subfield, but we haven't proved this yet).

So what are the maximal ideals of $F[x]$? Last class, we defined an *irreducible polynomial* in $F[x]$ as a polynomial which doesn't factor as a product of polynomials of lesser degree (in other words, the only factorizations are the stupid ones where one of the factors is a constant, like $x^2+1 = (\frac{1}{2})(2x^2+2)$). We learned that all ideals in $F[x]$ are principal (generated by a single element), and the maximal ideals are exactly the ones that are generated by an irreducible polynomial. Therefore, whenever you have an irreducible polynomial $f(x) \in F[x]$, the ring $F[x]/\langle f(x)\rangle$ will be a field.

Soon, we will see this new field will be one in which the original polynomial has roots (just like how x^2+1 has roots $\pm i$ in $\mathbb{C} \cong \mathbb{R}[x]/\langle x^2+1\rangle$). Before we do this, let's prove that all polynomials factor uniquely into a product of irreducible polynomials. In this way, "irreducible polynomials" are like the indivisible building blocks of all polynomials, just like how all prime numbers are the indivisible building blocks of all integers.

Theorem (unique factorization): Let F be a field. Any polynomial $f \in F[x]$ can be written as a product $p_1p_2\dots p_n$ of polynomials which are irreducible in $F[x]$. Moreover, if $q_1q_2\dots q_m$ is another way to write f as a product of polynomials which are irreducible in $F[x]$, then $m=n$ and you can re-order the q_i's so that $p_i = q_iu_i$ for units u_i.

We'll get to the proof soon, but here are some comments:

What's the meaning of the condition about units? Consider the factorizations of x^2-1 in $\mathbb{R}[x]$.

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

These are different ways to write x^2-1 as a product of polynomials which are irreducible in $\mathbb{R}[x]$, but they're not different in a meaningful way – all we've done is multiply one factor by two and divide the other factor by two. The theorem says that there's only one way to factor x^2-1 if you ignore these superficially different ways to factor.

Isn't this obvious? No. Consider the subring R of complex numbers consisting of elements of the form $a+b\sqrt{-5}$. In this ring, 21 can be factored in two very different ways:

$$21 = (7)(3)$$
$$21 = (1+2\sqrt{-5})(1-2\sqrt{-5})$$

I am not going to define what it means for an element of R to be "irreducible," so this example is very vague. All I want for you to learn from it is that there are situations where factorization happens differently than you're used to in $\mathbb{Z}$, and our proof of the unique factorization theorem will use the fact that our ring is $F[x]$, where F is a field (in particular, we'll use the fact that $F[x]$ is a principal ideal domain). Before the proof, a quick lemma.

Lemma: Let $f(x) \in F[x]$ be an irreducible polynomial. If $f(x)$ is a factor of $p(x)q(x)$, where $p(x), q(x) \in F[x]$, then either $f(x)$ is a factor of $p(x)$ or of $q(x)$.

Proof of Lemma: Let $I = \langle f(x), p(x) \rangle$. Because $F[x]$ is a principal ideal domain, $I = \langle r \rangle$ for some $r \in F[x]$.

If r is a unit: Then there is some g, h so that $gf + hp = 1$, so $qgf + qhp = q$. Because f is a factor of the left hand side, then f is a factor of q.

If r is not a unit: Then $f = rz$ for some z, and because f is irreducible and r is not a unit, z must be a unit. Then $\langle f \rangle = \langle r \rangle = I$. So f is a factor of p.

Example: The polynomial $x^2 + 1 \in \mathbb{R}[x]$ is a factor of $x^6 - x^5 + x^4 - x^2 + x - 1$ (you can verify this by long division), and

$$x^6 - x^5 + x^4 - x^2 + x - 1 = (x^3 - x^2 + x - 1)(x^3 + 1)$$

so the lemma proves that either $x^2 + 1$ is a factor of $x^3 - x^2 + x - 1$ or is a factor of $x^3 + 1$. Indeed, $x^3 - x^2 + x - 1 = (x^2 + 1)(x - 1)$.

Example: The polynomial $x^2 - 1 \in \mathbb{R}[x]$ is a factor of itself, $x^2 - 1 = (x + 1)(x - 1)$. But it is *not* true that either $x^2 - 1$ is a factor of $x + 1$ or of $x - 1$. This shows that the condition that $f(x)$ is irreducible in the lemma is necessary.

Proof (of unique factorization): To prove the existence of a factorization, start with some $f(x) \in F[x]$. If it is irreducible, we're done. If it's not, then $f = p_1 p_2$, where p_1 and p_2 have smaller degree than f, but greater than 0. If both p_1 and p_2 are both irreducible we're done; otherwise we factor whichever of p_1 and p_2 are reducible. We continue this process until all terms are irreducible (this process must terminate because the maximum degree of the *reducible* terms in the product decreases in each step).

To prove the uniqueness (up to units), let $p_1 p_2 \ldots p_n$ and $q_1 q_2 \ldots q_m$ be irreducible factorizations of f. Use induction: if $n = 1$, then f is irreducible and $n = m$ and $p_1 = q_1$. Assume the uniqueness statement is true for elements of $F[x]$ with a factorization having at most $n - 1$ elements. By the lemma, p_1 is a factor of at least one of $q_1, q_2, \ldots, q_m$. Re-order the elements so that p_1 is a factor of q_1. That is, $up_1 = q_1$ for some unit u (it must be a unit because q_1 is irreducible). Then

$$up_1 p_2 \ldots p_n = uq_1 q_2 \ldots q_m$$

But since $up_1 = q_1$, we can cancel the corresponding terms from the above equation

$$p_2 p_3 \ldots p_n = (uq_2) q_3 \ldots q_m$$

and apply the inductive hypothesis, which tells us that the terms on the left and right are the same (up to re-ordering and units). If you delete the u factor from uq_2, it is still true that the terms on the left and right are the same up to re-ordering and units; because p_1 and q_1 are also the same up to a unit, the inductive step follows.

Lecture 8: Fundamental Theorem of Field Theory

We're finally prepared to begin our systematic study of field extensions: the process of taking one field and enlarging it into a new field. In a few weeks, we'll see how field extensions can help us study geometric constructions. The process of drawing lines and circles on a paper with a straightedge and compass, and then intersecting these lines, can be translated into the process of constructing larger fields from a smaller field. We'll use field theory to solve problems that stumped the (ancient) Greeks.

Definition: A **subfield** of a field F is a subring S of F which is itself a field.

Definition: A field E is an **extension field** of a field F if F is a subfield of E.

Warning: A lot of mathematicians will say things like "E contains F as a subfield" when they really mean "E contains a subfield isomorphic to F." This is imprecise, but I might do it too sometimes.

Examples: $\mathbb{Q}$ is a subfield of $\mathbb{R}$, and $\mathbb{R}$ is a subfield of $\mathbb{C}$. Said differently, $\mathbb{R}$ is an extension field of $\mathbb{Q}$, and $\mathbb{C}$ is an extension field of $\mathbb{R}$.

Questions:

1. Can you think of any fields that contain $\mathbb{C}$? (Watch out: the quaternions are not commutative!)
2. Can you think of any fields between $\mathbb{Q}$ and $\mathbb{R}$?
3. Can you think of any fields between $\mathbb{Q}$ and $\mathbb{C}$ that neither contain nor are contained in $\mathbb{R}$?

Important Comment: If E is an extension field of F, any polynomial $f(x) \in F[x]$ is also an element of $E[x]$. In other words, if F is a subfield of E, then $F[x]$ is a subring of $E[x]$.

Theorem: Let F be a field and $f(x)$ be an irreducible polynomial in $F[x]$

1. $F[x]/\langle f(x)\rangle$ is a field that contains a subfield isomorphic to F.
2. The polynomial f has a root in $F[x]/\langle f(x)\rangle$.

The statement of the second part of this theorem can be confusing. In it, you should think of $f(x)$ as a function that you plug elements into and it spits out some output element. Don't get it confused with the element $[f(x)]$ of the field $F[x]/\langle f(x)\rangle$.

Proof:

(1): Because $\langle f(x)\rangle$ is a maximal ideal, $F[x]/\langle f(x)\rangle$ is a field. Consider the map

$$\begin{aligned}\varphi : F &\to F[x]/\langle f(x)\rangle\\ a &\mapsto [a].\end{aligned}$$

This map is a homomorphism, because

$$\begin{aligned}\varphi(a+b) &= [a+b] = [a]+[b] = \varphi(a)+\varphi(b)\\ \varphi(ab) &= [ab] = [a][b] = \varphi(a)\varphi(b)\end{aligned}$$

The kernel of this homomorphism is the ideal of all elements of F sent to the zero coset, i.e. $F \cap \langle f(x)\rangle$. But $F \cap \langle f(x)\rangle = 0$, so φ is injective. By the first isomorphism theorem, $\operatorname{im}(\varphi) \cong F$.

(2): Suppose $f(y) = a_0 + a_1 y + \cdots + a_n y^n$ (originally, we used the variable x in the polynomial f, but here I want you to think of f as a function and I don't want you to confuse the "input variable of f" with the element x in the polynomial ring $F[x]$, so I changed the name of the variable).

When we plug the element $[x] \in F[x]/\langle f(x)\rangle$ into f, we get the following element of $F[x]/\langle f(x)\rangle$

$$[a_0] + [a_1][x] + [a_2][x]^2 + \cdots + [a_n][x]^n = [a_0 + a_1x + a_2x^2 + \cdots + a_nx^n] = [0]$$

So $[x]$ is a root of f.

Corollary (Fundamental Theorem of Field Theory): Let F be a field and $f(x)$ be *any* nonconstant polynomial in $F[x]$. Then there is an extension field E of F in which $f(x)$ has a root.

Proof: If $f(x)$ has a root, then $E = F$. Otherwise, let $f(x) = p_1(x)p_2(x)\dots p_n(x)$ be a factorization of $f(x)$ into irreducible polynomials. For any $1 \le i \le n$, the field $F[x]/\langle p_i(x)\rangle$ is a field extension of F for which $p_i(x)$ has a root. Therefore, $f(x)$ also has a root in $F[x]/\langle p_i(x)\rangle$.

Notation: Let E be a field extension of F, and let $a_1, \dots, a_n \in E$. We write $F(a_1, \dots, a_n)$ (sometimes called "F adjoint a_1 through a_n") is the smallest subfield of E that contains both F and also the set $\{a_1, \dots, a_n\}$.

Example: When $E = \mathbb{R}, F = \mathbb{Q}$ and $a = \sqrt{2}$, the field $\mathbb{Q}(\sqrt{2})$ is $\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\} \subseteq \mathbb{R}$.

Example: When $E = \mathbb{C}, F = \mathbb{Q}$ and $a = i$, the field $\mathbb{Q}(i)$ is $\{a + bi \mid a, b \in \mathbb{Q}\} \subseteq \mathbb{C}$. This field is an extension of $\mathbb{Q}$ and a subfield of $\mathbb{C}$.

Example/Question: When $E = \mathbb{C}, F = \mathbb{Q}$ and $a_1 = \sqrt{2}, a_2 = \sqrt{5}$, the field $\mathbb{Q}(\sqrt{2}, \sqrt{5})$ is *not* equal to $\{a + b\sqrt{2} + c\sqrt{5} \mid a, b, c \in \mathbb{Q}\} \subseteq \mathbb{R}$. Why not? Can you describe what it *is* equal to?

Optional problems (don't hand in):

1. For any $a, b \in \mathbb{R}$ with $b \neq 0$, show that $\mathbb{R}(a + bi) = \mathbb{C}$.
2. Describe the field $\mathbb{Q}(\pi)$.
3. (Gallian, example 2) Find two different field extensions of $\mathbb{Z}_3$, one with 9 elements and another with 27 elements, in which the polynomial $f(x) = x^5 + 2x^2 + 2x + 2 \in \mathbb{Z}_3[x]$ has a root (hint: one factor of $f(x)$ is $x^2 + 1$).

Lecture 9: More on Field Extensions

Reminder of Last Class

Last class, we learned two ways to make a field F bigger.

Method 1: Start with an irreducible polynomial $f(x) \in F[x]$ of degree ≥ 2, and construct the field $E = F[x]/\langle f(x)\rangle$. We learned two important facts about this field:

1. The cosets represented by constant polynomials are a subfield of E which is isomorphic to F (this is the sense in which we "made the field F bigger").
2. The polynomial f has a root inside E, even though it doesn't have one in F.

Method 2: Start with a field extension E of F, and a collection of elements $\{\alpha_1, \dots, \alpha_n\}$ of E. We defined the extension $F(\alpha_1, \dots, \alpha_n)$ to be the smallest field that contains both F and the elements $\{\alpha_1, \dots, \alpha_n\}$.

Today, we will learn a relationship between these constructions. Also, we will describe how to measure the "size" of a field extension using the language of dimensions of vector spaces.

Relationship between $F(\alpha)$ and $F[x]/\langle f(x)\rangle$

Theorem: Let E be a field extension of F, and let $\alpha \in E$ be a root of an irreducible polynomial $f(x) \in F[x]$. Then

$$F(\alpha) \cong F[x]/\langle f(x)\rangle$$

Proof: For a polynomial $b_0 + b_1x + \cdots + b_nx^n \in F[x]$, plugging α into p gives

$$b_0 + b_1\alpha + \cdots + b_n\alpha^n$$

which is an element of $F(\alpha)$, since it is the sum of products of things in $F(\alpha)$. Consider the map

$$\begin{aligned} \varphi : F[x] &\to F(\alpha) \\ p(x) &\mapsto p(\alpha) \end{aligned}$$

This is a ring homomorphism. Its kernel contains $f(x)$, so $\langle f(x)\rangle \subseteq \ker(\varphi) \subseteq F[x]$. Since $\langle f(x)\rangle$ is maximal and $\ker(\varphi) \neq F[x]$, it must be the case that $\langle f(x)\rangle = \ker(\varphi)$. By the first isomorphism theorem, $\operatorname{im}(\varphi) \cong F[x]/\langle f(x)\rangle$, which shows that $\operatorname{im}(\varphi)$ is a field. Because the image is a field containing both F and α, it must be the entire field $F(\alpha)$. By the first isomorphism theorem, $F[x]/\langle f(x)\rangle \cong F(\alpha)$.

In the past, we've seen how every element of $F[x]/\langle f(x)\rangle$ can be represented by a polynomial of degree $< \deg(f)$. Let's revisit this result in the context of field extensions.

Theorem: Let E be a field extension of F, and $f(x)$ be an irreducible polynomial in $F[x]$ of degree n that has a root $\alpha \in E$. Every element of $F[x]/\langle f(x)\rangle$ has a unique representative of the form $b_0 + b_1x + \cdots + b_{n-1}x^{n-1}$, and every element of $F(\alpha)$ can be written as $b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1}$ in a unique way (in both cases, the b_i's are elements of F).

Proof: Consider $[p(x)] \in F[x]/\langle f(x)\rangle$. By the long division theorem, there are unique $q(x), r(x) \in F[x]$ with

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

and $\deg(r(x)) < n$. Because $f(x)q(x) \in \langle f(x)\rangle$, it follows that $[p(x)] = [r(x)]$. By the "uniqueness" part of the long division theorem, this $r(x)$ is unique. The isomorphism between $F[x]/\langle f(x)\rangle$ and $F(\alpha)$ constructed in the previous proof sends this coset to $r(\alpha)$.

Besides giving us an intuitive description of $F[x]/\langle f(x)\rangle$ and $F(\alpha)$, this result also sheds light on an interesting feature of this kind of field extension: depending on the degree of $f(x)$, you might need more (or fewer) powers of α to express every element of $F(\alpha)$. For example, if the degree of the irreducible polynomial is 2, everything in $F(\alpha)$ can be written as a combination of 1 and α. If the degree is 3, we need $1, \alpha$ and α^2. In general, the number of different powers of α necessary to represent all elements of $F(\alpha)$ is equal to the degree of the irreducible polynomial it satisfies. In other words, the elements $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ are a *basis* for the vector space $F(\alpha)$ over F, and we can measure the "size" of a field extension E of F as the dimension of E *as a vector space* over F. Let's review some abstract linear algebra.

Reminder of Abstract Linear Algebra

Definition/Reminder: A **vector space** over a field F is a set V with two operations: vector addition $v_1 + v_2$ and scalar multiplication kv_1 (where $v_1, v_2 \in V$ and $k \in F$) that satisfy the following properties:

- $(V, +)$ is an abelian group.
- Multiplication distributes with the addition operations in V *and* F:

$$(k_1 + k_2)v = k_1v + k_2v$$
$$k(v_1 + v_2) = kv_1 + kv_2$$

- Associativity: $(k_1k_2)v = k_1(k_2v)$.
- Scalar multiplication by $1 \in F$ is the identity: $1v = v$

In your linear algebra class, you probably studied the vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ so much that you might have forgotten the abstract definition of vector spaces and started thinking of them as "things isomorphic to $\mathbb{R}^n$ or $\mathbb{C}^n$".

Definition/Reminder: Vectors $v_1, \ldots, v_n$ in V are **linearly independent** if whenever $k_1v_1 + k_2v_2 + \cdots + k_nv_n = 0$ for some $k_1, \ldots, k_n \in F$, then $k_1 = k_2 = \cdots = k_n = 0$.

Definition/Reminder: A **basis** for V is a linearly independent subset B of V with the property that every element of V can be written as a linear combination of elements of B (that is, as $k_1b_1 + k_2b_2 + \cdots + k_nb_n$, where the k_i's are in F and b_i's are in B).

Definition/Reminder: The **dimension** of V is the number of elements of a basis of B. (fact: every basis of a vector space has the same number of elements, so this definition makes sense).

Example: The set $V = \mathbb{R}^n$ is a vector space over the field $\mathbb{R}$. We add vectors by

$$(a_1, \dots, a_n) + (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n)$$

and multiply by scalars $k \in \mathbb{R}$ by elements of V according to the rule

$$k(a_1, \dots, a_n) = (ka_1, \dots, ka_n)$$

A basis for V over $\mathbb{R}$ is

$$\begin{aligned} e_1 &= (1, 0, \dots, 0) \\ e_2 &= (0, 1, \dots, 0) \\ &\vdots \\ e_n &= (0, 0, \dots, 1) \end{aligned}$$

Example: The set V of all polynomials in $\mathbb{R}[x]$ with degree ≤ 2 is a vector space over $\mathbb{R}$. We add vectors by standard polynomial addition, and multiply elements of $\mathbb{R}$ by elements of V according to the rule

$$k(a_0 + a_1x + a_2x^2) = ka_0 + ka_1x + ka_2x^2$$

A basis for V over $\mathbb{R}$ is

$$\begin{aligned} e_1 &= 1 \\ e_2 &= x \\ e_3 &= x^2 \end{aligned}$$

Example: The set $\mathbb{C}$ of all complex numbers is a vector space over $\mathbb{R}$ using the standard addition of complex numbers, and the standard multiplication of a complex number by a real number. A basis for V over $\mathbb{R}$ is

$$\begin{aligned} e_1 &= 1 \\ e_2 &= i \end{aligned}$$

Example: The set $\mathbb{R}$ of all real numbers is a vector space over $\mathbb{Q}$ using the standard addition of real numbers, and the standard multiplication of a real number by a rational number. A basis for $\mathbb{R}$ as a vector space over $\mathbb{Q}$ is infinite; its construction also requires the axiom of choice.

In many of the examples above, we view a field as a vector space over a subfield of itself. More preciely, if E is a field extension of F (so F is a subfield of E), we can view E as a vector space over F in the following way:

- The addition rule for elements of the vector space E is the same as the addition rule in the field E

- The multiplication of an element of F with an element of E is the same as the multiplication rule inside E (because elements of F are also in E, this makes sense)

Degree of a Field Extension

Definition: Let E be a field extension of F. We say that E **has degree** n **over** F if the dimension of E over F (as a vector space) is equal to n. In this case, we write

$$[E : F] = n$$

If $[E : F]$ is finite, we say E is a **finite extension** of F. Otherwise, we say that E is an **infinite extension** of F.

Example: Referring to the examples above,

- $[\mathbb{C} : \mathbb{C}] = 1$
- $[\mathbb{C} : \mathbb{R}] = 2$
- $\mathbb{R}$ is an infinite extension of $\mathbb{Q}$
- $[\mathbb{Q}(2^{1/3}) : \mathbb{Q}] = 3$

Optional problems:

1. The polynomial $x^3 + 9x + 6 \in \mathbb{Q}[x]$ is irreducible. It has a root $\alpha \in \mathbb{R}$. Calculate the inverse of $1 + \alpha$ in the ring $\mathbb{Q}(\alpha)$ (your answer should be some linear combination of powers of α).
2. The polynomial $x^3 - 2x - 2 \in \mathbb{Q}[x]$ is irreducible. Let α be a root of it in some field extension. Compute $(1 + \alpha)(1 + \alpha + \alpha^2)$ and $\frac{1+\alpha}{1+\alpha+\alpha^2}$ in $\mathbb{Q}(\alpha)$ (as in the previous problem, your answers should be some linear combination of powers of α).

Lecture 10: Eisenstein's criterion and adjoining multiple elements

Let F be a field. So far, we've discussed two ways of constructing field extensions of F. First, if you have an irreducible polynomial $f(x) \in F[x]$, the field $F[x]/\langle f(x)\rangle$ is an extension of F. Second, if α is an element of a field extension E of F, then $F(\alpha)$ is defined to be the smallest subfield of E containing both F and α.

These two constructions are related to each other when $f(x)$ is an irreducible polynomial having α as a root. In this case, we proved that the two field extensions described above are actually isomorphic. Moreover, we can actually use the isomorphism between them to describe the elements of $F(\alpha)$: they are F-linear combinations of the elements $\{1, \alpha, \alpha^2, \dots, \alpha^{n-1}\}$, where n is the degree of the irreducible polynomial. This observation motivated the definition of the *degree* of a field extension as the dimension of the larger field when viewed as a vector space over the smaller field.

We're almost ready to apply everything we've learned to the topic of geometric constructions. There are only two more topics that need to be covered. First of all, we are going to be studying extensions of $\mathbb{Q}$ a lot, and for that it will be helpful to have ways to determine when a polynomial in $\mathbb{Q}[x]$ is irreducible. Next, we need to understand the structure of field extensions obtained by adjoining *multiple* elements to a field, not just one.

Eisenstein's Criterion

Theorem (Eisenstein's criterion): A polynomial

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x].$$

is irreducible as an element of $\mathbb{Q}[x]$ if there is a prime number $p \in \mathbb{Z}$ such that the following three conditions are satisfied.

- p divides each of of $a_{n-1}, a_{n-1}, \dots, a_0$
- p does not divide a_n
- p^2 does not divide a_0

Proof: I won't prove this, because it would take an entire 2-hour lecture of learning about irreducibility for polynomials whose coefficients *aren't* fields, and we have much more exciting ways to spend our time!

Example 1: The polynomial $2x^5 + 6x^3 + 300x^2 - 9x + 6$ satisfies the three criteria for $p = 3$, so it is irreducible as a polynomial in $\mathbb{Q}[x]$. (note: the coefficient "0" of x^4 is divisible by 3)

Remark 1: Notice that to apply Eisenstein's criteria, the coefficients must be in $\mathbb{Z}$, but the result is about irreducibility in $\mathbb{Q}[x]$. This isn't a typo.

Remark 2: The converse of Eisenstein isn't true. For example, the polynomial $x^2 + 1$ is irreducible as an element of $\mathbb{Q}[x]$ even though you can't find any primes p for which Eisenstein's criteria apply.

Question: Can you apply Eisenstein's criteria to the polynomial $x^6 + 30x^5 - 15x^3 + 6x - 120$?

Question: Can you apply Eisenstein's criteria to the polynomial $x^4 + 30x^2 + 6x - 36$?

Adjoining multiple elements to a field

Remember that if E is an extension of F, and $\alpha_1, \ldots, \alpha_n \in E$, the field $F(\alpha_1, \ldots, \alpha_n)$ is defined to be the smallest subfield of E that contains F and each α_i. We have a complete understanding of the structure of $F(\alpha)$ when α is the root of an irreducible polynomial in F. What if you adjoin more than one element? The next result says that "Adjoining two elements is the same as adjoining the elements one at a time"

Lemma: Let E be a field extension of F, and $\alpha, \beta \in E$.

$$F(\alpha, \beta) = (F(\alpha))(\beta)$$

Proof: Because $(F(\alpha))(\beta)$ contains F, α, and β, and $F(\alpha, \beta)$ is defined to be the *smallest* subfield of E containing F, α, and β, it follows that $F(\alpha, \beta) \subseteq (F(\alpha))(\beta)$.

To show the other inclusion, first observe that $F(\alpha)$ is a subfield of $F(\alpha, \beta)$ (because $F(\alpha, \beta)$ contains F and α, and $F(\alpha)$ is the *smallest* subfield of E containing F and α)

Therefore, $F(\alpha, \beta)$ contains both $F(\alpha)$ and β. But $(F(\alpha))(\beta)$ is the *smallest* subfield of E containing $F(\alpha)$ and β, so $(F(\alpha))(\beta) \subseteq F(\alpha, \beta)$.

Detailed example: $\mathbb{Q}(\sqrt{3}, \sqrt{5})$

The above theorem says that $\mathbb{Q}(\sqrt{3}, \sqrt{5}) = (\mathbb{Q}(\sqrt{5}))(\sqrt{3})$. Let's remind ourselves what we know about $\mathbb{Q}(\sqrt{5})$.

- $x^2 - 5 \in \mathbb{Q}[x]$ is an irreducible polynomial (by Eisenstein) having $\sqrt{5}$ as a root. Therefore, $\mathbb{Q}(\sqrt{5}) \cong \mathbb{Q}[x]/\langle x^2 - 5 \rangle$.
- $\mathbb{Q}(\sqrt{5})$ has basis $\{1, \sqrt{5}\}$ as a vector space over $\mathbb{Q}$
- $\mathbb{Q}(\sqrt{5})$ consists of elements of the form $\{a + b\sqrt{5}\}$, where $a, b \in \mathbb{Q}$.

Next, we want to adjoin $\sqrt{3}$ to the field $\mathbb{Q}(\sqrt{5})$ to get $\mathbb{Q}(\sqrt{5}, \sqrt{3})$. If we knew that $x^2 - 3$ is an irreducible polynomial in $\mathbb{Q}(\sqrt{5})[x]$, we could use the theorems from last week to describe $(\mathbb{Q}(\sqrt{5}))(\sqrt{3})$. But we can't use Eisenstein to conclude that $x^2 - 3 \in \mathbb{Q}(\sqrt{5})[x]$ is irreducible, because Eisenstein only applies to polynomials in $\mathbb{Q}[x]$, but we *can* use the irreducibility test that says a polynomial of degree 2 or 3 is irreducible if and only if it has no roots.

To show $x^2 - 3$ has no roots in $\mathbb{Q}(\sqrt{5})$, assume towards a contradiction that it does: $a + b\sqrt{5}$ is a root of $x^2 - 3$ for some $a, b \in \mathbb{Q}$. That is,

$$a^2 + 5b^2 + 2ab\sqrt{5} = 3$$

Case 1: If $ab \neq 0$, then $\sqrt{5} = (3 - a^2 + 5b^2)/(2ab)$ would be a rational number, which it's not.

Case 2: If $a = 0$, then $5b^2 = 3$, so $\sqrt{5}b = \sqrt{3}$, but multiplying through by $\sqrt{5}$ shows that $\sqrt{15}$ must be rational, which it's not.

Case 3: If $b = 0$, then $a = \sqrt{3}$ would be a rational number, which it's not.

So $x^2 - 3 \in \mathbb{Q}(\sqrt{5})$ is indeed an irreducible polynomial, so we can conclude

- $\mathbb{Q}(\sqrt{5}, \sqrt{3}) \cong \mathbb{Q}(\sqrt{5})[x]/\langle x^2 - 3 \rangle$.
- $\mathbb{Q}(\sqrt{5}, \sqrt{3})$ has basis $\{1, \sqrt{3}\}$ as a vector space over $\mathbb{Q}(\sqrt{5})$

- $\mathbb{Q}(\sqrt{5},\sqrt{3})$ consists of elements of the form $\{a+b\sqrt{3}\}$, where $a,b \in \mathbb{Q}(\sqrt{5})$. Because these a,b are themselves of the form $c+d\sqrt{5}$ for $c,d \in \mathbb{Q}$, we can write every element of $\mathbb{Q}(\sqrt{5},\sqrt{3})$ in the form $k_1+k_2\sqrt{3}+k_3\sqrt{5}+k_4\sqrt{15}$, where each $k_i \in \mathbb{Q}$.

Degrees of field extensions multiply

Theorem: Let K be an extension of E with basis $\{\alpha_1,\ldots,\alpha_n\}$, and E be an extension of F with basis $\{\beta_1,\ldots,\beta_k\}$. Then $B=\{\alpha_i\beta_j \mid 1\le i\le n, 1\le j\le k\}$ is a basis for K over F.

Proof: Let $a \in K$. Then

$$a = e_1\alpha_1 + e_2\alpha_2 + \cdots + e_n\alpha_n$$

for $e_1,\ldots,e_n \in E$. Since each e_i can be written as an F-linear combination of β_j's, say $\sum f_{ij}\beta_j$, we have

$$\begin{aligned} a = {} & (f_{11}\beta_1 + f_{12}\beta_2 + \cdots + f_{1k}\beta_k)\alpha_1 + \\ & (f_{21}\beta_1 + f_{22}\beta_2 + \cdots + f_{2k}\beta_k)\alpha_2 + \\ & \quad\vdots \\ & (f_{n1}\beta_1 + f_{n2}\beta_2 + \cdots + f_{nk}\beta_k)\alpha_n. \end{aligned}$$

by expanding the above expression, we have written a as a F-linear combination of elements of B. This proves that B *spans* K as a vector space over F.

It remains to show that there are no linear relations amongst the B's. Suppose that

$$\sum_{i,j} c_{ij}\alpha_i\beta_j = 0$$

Then by grouping all the terms with the same i-value together, we can re-write this as

$$\sum_i \left(\sum_j c_{ij}\beta_j\right)\alpha_i = 0$$

But because the $\{\alpha_i\}$ are linearly independent, this implies that each $\sum_j c_{ij}\beta_j$ is zero. Then using the fact that the $\{\beta_j\}$ are linearly independent, this means that each c_{ij} equals zero. This shows that the only F-linear combination of elements of B equal to zero is the all-zero combination, which proves the linear indpendence of elements of B, which complete the proof that B is a basis.

Corollary: If $F \subseteq E \subseteq K$ are fields, then $[K:F]=[K:E][E:F]$.

Proof: The number of elements in a basis for K as a vector space over F equals the product of the number of elements of a basis for K as a vector space over E times the number of elements of a basis for E as a vector space over F.

Example: We saw that $[\mathbb{Q}(\sqrt{3},\sqrt{5}):\mathbb{Q}]=4$, and $[\mathbb{Q}(\sqrt{3},\sqrt{5}):\mathbb{Q}(\sqrt{3})]=2$ and $[\mathbb{Q}(\sqrt{3}):\mathbb{Q}]=2$ in the last example.

Application: Let E be a field extension of F of degree p, where p is a prime number. What are the subfields of E containing F?

Application: Calculate $[\mathbb{Q}(\sqrt[7]{3},\ \sqrt[17]{5}):\mathbb{Q}]$.

Lecture 11: Constructible Numbers

We're finally ready for our second application of ring theory (the first was the Chinese remainder theorem). This week, we'll show how the theory of field extensions helps us solve basic questions about plane geometry that stumped the Greeks. Suppose you have a collection of points, lines, and circles on the plane $\mathbb{R}^2$, and you have a compass and straightedge. There are three basic operations you can perform using these tools.

Compass and Straightedge Operations

1. Draw an (infinite) straight line between two points.
2. Set your compass radius to the distance between two points, and draw a circle around a point. This third point may be the same as or may be different from your original two.
3. Draw a point at the intersection of two circles, of two lines, or of a circle and a line.

A natural question to ask is:

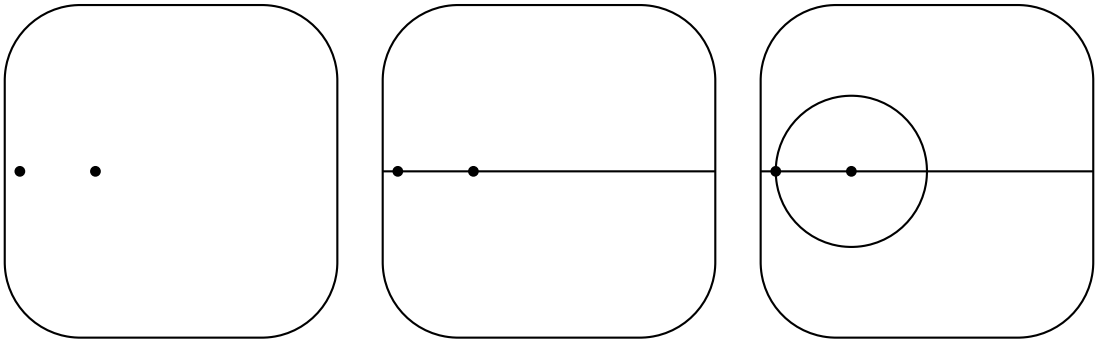
> What kinds of more complicated operations can I perform by a sequence of these basic operations?

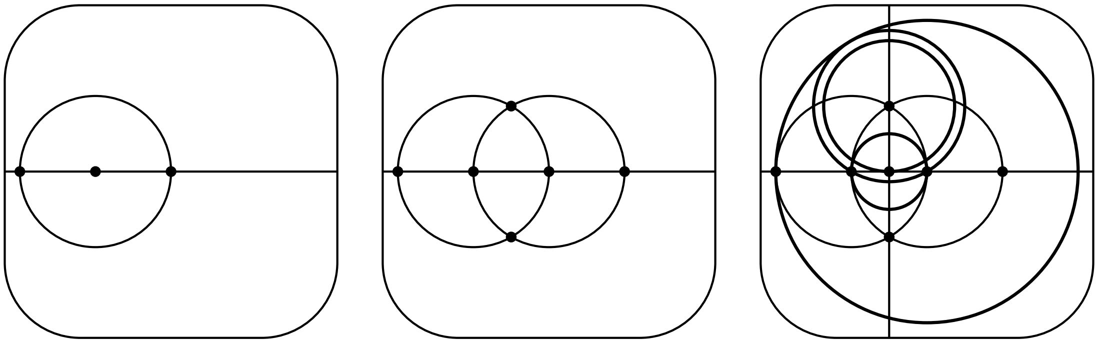
For example, if you have two points that are a certain distance d apart, can you construct a equilateral triangle with side lengths d? What about a square, pentagon, or hexagon? Can you bisect angles? Trisect angles? Can you construct parallel lines? Perpendicular lines?

Constructibility

Recall: Start with two points on the plane $\mathbb{R}^2$ that are distance 1 apart. A point, line, or circle on $\mathbb{R}^2$ is **constructible** if can be drawn in a finite number of steps using the operations above. A real number k is **constructible** if there are two constructible points that are distance $|k|$ apart.

Let's get a feel for constructible lines, points, circle, and numbers by seeing what we can draw just by starting with two points of distance 1 apart.





Clearly, you can get a lot of lines and circles and points just from these simple operations. Let's be more systematic in our study of constructible numbers. What properties do they have?

Claim: If a and b are constructible numbers, then so is $a + b$.

Proof: If either of a or b is zero, the claim is trivial. Otherwise, draw a line through the two points p, q that are distance $|a|$ apart, then a circle of length $|b|$ around q. One of the points where this circle intersects the line will be a distance of $|a + b|$ from p.

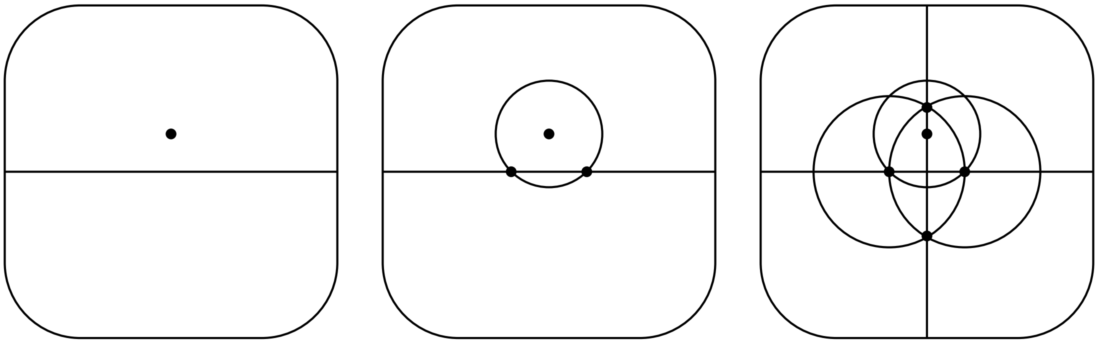
Claim: If a and b are constructible numbers, so is $a - b$

Proof: If b is constructible, then so is $-b$. So $a + (-b)$ is constructible.

As a consequence of all this, we know that every integer is constructible. We also want to show that ab and a/b are constructible, but first we show two general tricks for constructions.

Claim: If p is a constructible point and L a constructible line, then the line through p perpendicular to L is also constructible.

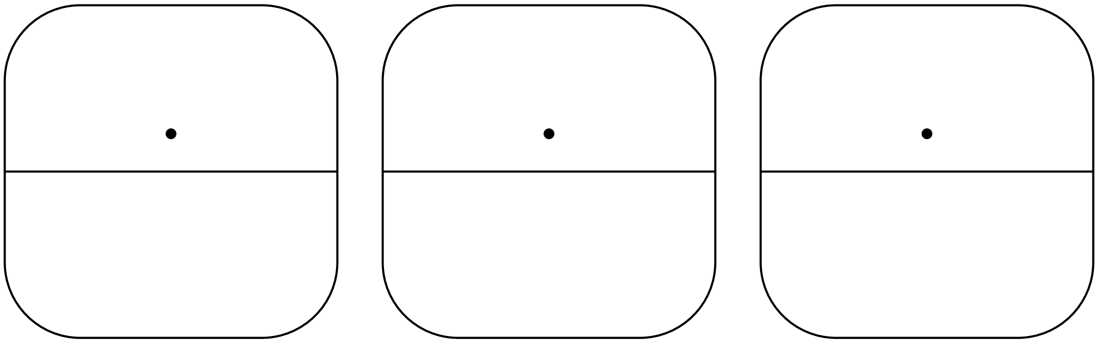
Proof:



Note that the above proof works even when p is on the line L.

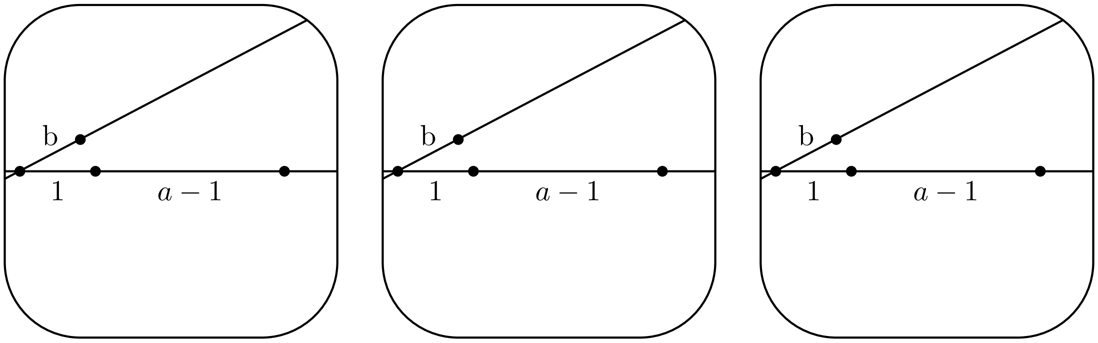
Claim: If p is a constructible point and L a constructible line, then the line through p parallel to L is also constructible.

Proof: (fill me in!)



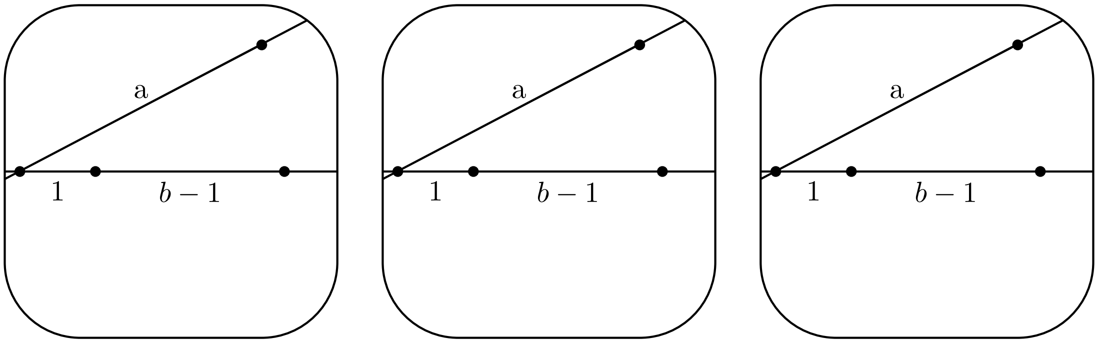
Claim: If $a, b \in \mathbb{R}$ are constructible, then so is ab.

Proof: (fill me in!)



Claim: If $a, b \in \mathbb{R}$ are constructible, then so is a/b.

Proof: (fill me in!)



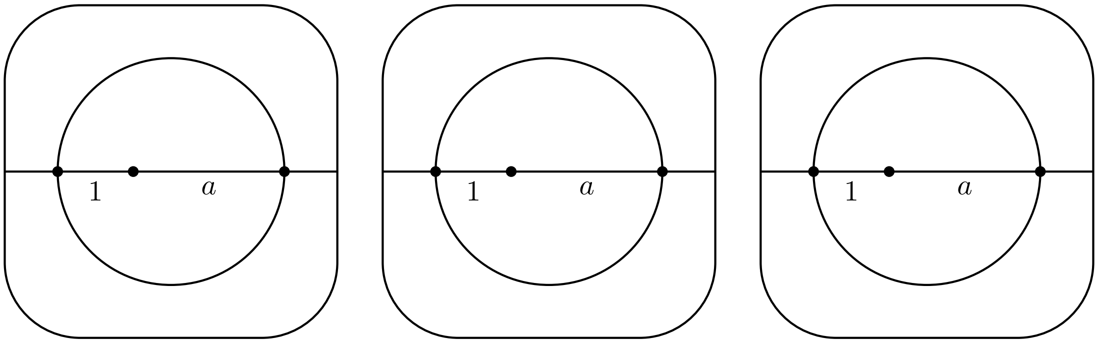
Theorem: The set of constructible numbers is a subfield of $\mathbb{R}$.

Proof: We just need to show that it is a subring in which every nonzero element has an inverse. The fact that it is a subring follows from the fact that it is closed under addition, subtraction, and multiplication. The fact that every nonzero element is a unit follows from the fact that 1 is constructible, and that $1/a$ is constructible if a is constructible.

We already know that there are some irrational constructible numbers (for example $\sqrt{2}$). The next two give some other examples of constructible numbers.

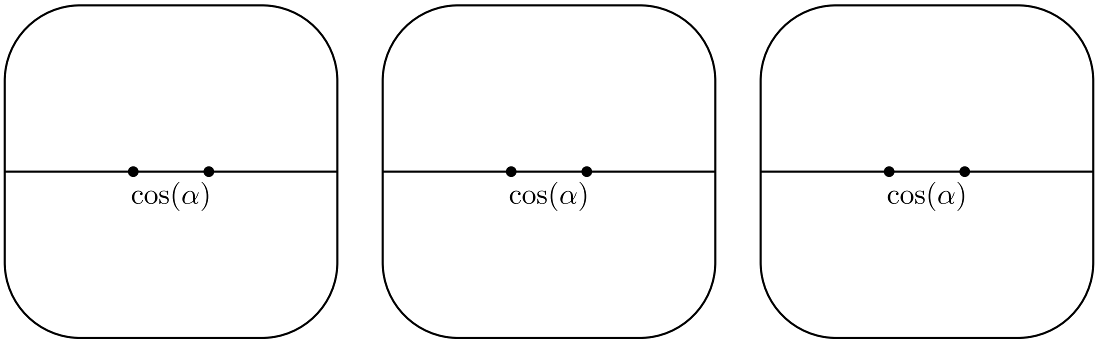
Claim: If $a > 0$ is constructible, so is $\sqrt{a}$.

Proof: (fill me in!)



Claim: Let α be an angle. If $\cos(\alpha)$ is constructible, then so is $\sin(\alpha)$.

Proof: (fill me in!)



Lecture 12: Impossible Constructions

Today, we'll prove that certain constructions are impossible with a straightedge and compass. We will begin by describing a general procedure for taking any sequence of straightedge and compass operations, and associating to it sequence of field extensions of $\mathbb{Q}$ – whenever we draw a new point as one of our operations, we adjoin the x and y coordinates of the point. We will show that these field extensions are always degree 2. However, some geometric constructions, such as trisecting certain angles, necessitate constructing a field extension whose degree is divisible by 3, which we will show cannot occur as a sequence of degree-2 extensions.

Let's remind outselves of the equations for lines and circles in $\mathbb{R}^2$. Let F be a subfield of $\mathbb{R}$.

Line: If (x_1, y_1) and (x_2, y_2) are two distinct points with $x_1, y_1, x_2, y_2 \in F$, then the line through them is given by

$$L = \{(y_1 - y_2)x + (x_2 - x_1)y + y_2x_1 - x_2y_1 = 0\}$$

If we simplify, we get

$$L = \{ax + by + c = 0\} \qquad \text{where } a, b, c \in F$$

Circle: If d is the distance between (x_1, y_1) and (x_2, y_2), and (x_3, y_3) is a third point, with $x_1, y_1, x_2, y_2, x_3, y_3 \in F$, then the circle through (x_3, y_3) of radius d is given by

$$C = \{(x - x_3)^2 + (y - y_3)^2 = d^2\}$$

If we simplify (and use the fact that $d^2 \in F$), we get

$$C = \{x^2 + y^2 + ax + by + c = 0\} \qquad \text{where } a, b, c \in F$$

Now let's study what happens when we take lines or circles of this form and intersect them. *What extension field of F will the coordinates of these new points live in?*

Intersecting Lines and Circles

Proposition: Let F be a subfield of $\mathbb{R}$, and suppose

$$L_1 = \{a_1x + b_1y + c_1 = 0 \mid a_1, b_1, c_1 \in F\}$$
$$L_2 = \{a_2x + b_2y + c_2 = 0 \mid a_2, b_2, c_2 \in F\}.$$

intersect at a point (p, q). Then $p, q \in F$.

Proof: Solve one of the equations for y in terms of x, and then substitute it into the other equation. This gives a linear polynomial with coefficients in F whose root equals p, so $p \in F$.

Proposition: Let F be a subfield of $\mathbb{R}$, and suppose

$$L = \{a_1x + b_1y + c_1 = 0 \mid a_1, b_1, c_1 \in F\}$$
$$C = \{x^2 + y^2 + a_2x + b_2y + c_2 = 0 \mid a_2, b_2, c_2 \in F\}$$

intersect in a point (p, q). Then either $p, q \in F$, or $F(p, q)$ is a degree-2 field extension of F.

Proof: First suppose $b_1 \neq 0$. Then solve the equation for L for y in terms of x, and then substitute it into the equation for C. This shows that the x-coordinate of intersection, p, is the root of a degree-2 polynomial with coefficients in F, so $[F(p) : F]$ is either 2 or 1. Plugging p into the equation for L and shows that $q \in F(p)$, so $F(p, q) = F(p)$. If $b_1 = 0$, then repeat the proof above with the roles of x and y reversed.

Proposition: Let F be a subfield of $\mathbb{R}$, and suppose

$$C_1 = \{x^2 + y^2 + a_1x + b_1y + c_1 = 0 \mid a_1, b_1, c_1 \in F\}$$
$$C_2 = \{x^2 + y^2 + a_2x + b_2y + c_2 = 0 \mid a_2, b_2, c_2 \in F\}$$

intersect in a point (p, q). Then either $p, q \in F$, or $F(p, q)$ is a degree-2 field extension of F.

Proof: By subtracting the equation for C_2 from the one for C_1, we see that (p, q) must be on the line

$$(a_1 - a_2)x + (b_1 - b_2)y + (c_1 - c_2) = 0$$

Then apply the previous proposition.

Associating field extensions to straightedge and compass operations

Suppose we start with a collection of points in $\mathbb{R}^2$ that have coordinates in $\mathbb{Q}$. If we perform a sequence of straightedge and compass operations on these points, we can construct a sequence of field extensions of $\mathbb{Q}$ – every time you draw a point (p, q), adjoin p and q to the current field F. By the above propositions, every time you draw a line or a circle, the equation for your line or circle will have coefficients in your current field; by the above three propositions, any point you draw at the intersection of these lines or circles will correspond to a field extension of degree 1 or 2.

Trisecting the Angle is Impossible

When people say "It is impossible to trisect an angle using a straightedge and compass," they mean "There exist constructible angles θ for which $\theta/3$ is not constructuble."

Proposition: If an angle θ is constructible, then so are the numbers $\cos(\theta)$ and $\sin(\theta)$.

Proof: Suppose the angle is formed by the lines L_1 and L_2 meeting at the point p. be the point at the angle. Draw a circle of radius 1 around p, and suppose it meets L_2 at the point q. The line perpendicular to L_1 through q will intersect L_1 at a point $\cos(\theta)$ away from p and $\sin(\theta)$ away from q.

Proposition: The angle $\pi/3$ is constructible, but $\pi/9$ is not.

Proof: The fact that $\pi/3$ is constructible is easy: take two points of distance 1 apart, and draw a circles of radius 1 around each of them.

The fact that $\pi/6$ is *not* constructible relies on a little-known trigonometric identity, the triple angle formula:

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

Applied to $\theta = \pi/9$, this becomes

$$1/2 = 4(\cos(\pi/9))^3 - 3(\cos(\pi/9))$$

So $\cos(\pi/9)$ is a root of the polynomial $4x^3 - 3x - 1/2 \in \mathbb{Q}[x]$, which is irreducible (we will prove this afterwards). This means that $[\mathbb{Q}(\cos(\pi/9)) : \mathbb{Q}] = 3$.

Now assume towards a contradiction that we can find a sequence of straightedge and compass operations that constructed the angle $\pi/9$. Then we could construct the number $\cos(\pi/9)$, and the point $(\cos(\pi/9), 0) \in \mathbb{R}^2$. This means that the sequence of field extensions corresponding to these operations

$$\mathbb{Q} \subseteq F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n$$

has $\cos(\pi/9) \in F_n$. But this is impossible, each extension has degree 2 and $[\mathbb{Q}(\cos(\pi/9)) : \mathbb{Q}] = 3$.

Constructing a regular 7-gon is impossible

Proposition: It is impossible to construct a regular 7-gon.

Proof: Assume towards a contradiction that is is possible to construct a regular 7-gon. The interior angle of a 7-gon is $5\pi/7$, so it is possible to construct the angle $\pi - 5\pi/7 = 2\pi/7$, and therefore it is possible to construct $\cos(2\pi/7)$. A true fact about the number $\cos(2\pi/7)$ (which we will not prove) is that it is a root of the polynomial

$$8x^3 + 4x^2 - 4x - 1 \in \mathbb{Q}[x].$$

Using the same technique as in the last section, we can check that this polynomial has no roots, so is irreducible. Therefore, $[\mathbb{Q}(\cos(2\pi/7)) : \mathbb{Q}] = 3$, so $\cos(2\pi/7)$ cannot be contained in any field extension obtained through straightedge and compass operations.

Squaring the circle is impossible

Definition: Let F be a field, and k be an element of an extension of F. If there is a polynomial in $F[x]$ that has k as a root, then k is called **algebraic over** F. Otherwise, k is **transcendental over** F.

Some real numbers that are algebraic over $\mathbb{Q}$ include all rational numbers, as well as the n^{th} roots of any rational number. Some real numbers that are transcendental over $\mathbb{Q}$ include π, e, e^a (where a is any nonzero algebraic number). The set of real numbers which are algebraic over $\mathbb{Q}$ is countable, so the set of real numbers which are transcendental over $\mathbb{Q}$ is uncountable. In this sense, there are "more" transcendental numbers than real numbers.

Proposition: If α is transcendental over F, then $F(\alpha)$ has infinite degree over F.

Proof: Assume towards a contradiction that $[F(\alpha) : F]$ is some finite number n. Then the set $\{1, \alpha, \ldots, \alpha^n\}$ has $n+1$ elements in it, so must have some equation of linear dependence

$$a_0 1 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n \qquad \text{where } a_0, \ldots, a_n \in F.$$

Then α is a root of the polynomial $f(x) = a_0 + a_1x + \cdots + a_nx^n \in F[x]$, a contradiction to the fact that α is transcendental.

We will use the fact (without proof) that π is transcendental in the next proposition.

Proposition: It is impossible to construct a square whose area equals the area of a circle of radius 1.

Proof: If it were possible, we could construct a sequence of field extensions

$$\mathbb{Q} \subseteq F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n$$

such that F_n contains the number $\sqrt{\pi}$ (this is the side length of a square with area π). But then F_n contains the field $\mathbb{Q}(\pi)$, and $[\mathbb{Q}(\pi) : \mathbb{Q}]$ is infinite. But $[F_n : \mathbb{Q}]$ is not infinite.

Lecture 13: Splitting Fields

We ended class last Wednesday with the definition of an algebraic element of a field extension: an element α of an extension of F is *algebraic over* F if it is a root of a polynomial in $F[x]$, and *transcendental over* F otherwise. Recall also that if α is algebraic, then $[F(\alpha) : F]$ is finite; if is it transcendental, then $[F(\alpha) : F] = \infty$. Some of the most well-understood field extensions are the ones in which *all* the elements are algebraic.

Definition: A field extension E of F is **algebraic** if every element of E is algebraic over F.

Notice that if a field extension is has finite degree, then it must be algebraic. The converse isn't true, as you'll see in the homework. There are three important ways of constructing algebraic field extensions, only the first of which we've studied.

Important algebraic field extensions

Adjunction: Adjoin a single algebraic element α from an extension of F. Equivalently, take an irreducible polynomial $f(x) \in F[x]$ that has α as a root, and take $F[x]/\langle f(x)\rangle$.

Splitting Fields: Take any single polynomial $f(x) \in F[x]$ and adjoin *all* its roots to F.

Algebraic Closure: Adjoin *all* the roots of *all* the polynomials in $F[x]$ to F.

So far we've only studied adjunction because I wanted to get to geometric constructions as soon as possible, and we didn't need splitting fields or algebraic closures for geometric constructions. Now we need to go back and learn about these other two constructions today and tomorrow so that we'll be prepared to study Galois theory after the exam.

Splitting fields: definitions and examples

Definition: Let $f(x) \in F[x]$. An extension field E of F is called a **splitting field for $f(x)$ over F** if the following two conditions are satisfied:

1. $f(x)$ factors into linear polynomials ("splits completely") in $E[x]$.
2. $f(x)$ does *not* split completely in $K[x]$ for any $F \subsetneq K \subsetneq E$.

Example: $\mathbb{Q}(\sqrt{2})$ is a splitting field for $x^2 - 2$ over $\mathbb{Q}$. Let's check that the two conditions are satisfied

1. The polynomial $x^2 - 2$ splits as $(x + \sqrt{2})(x - \sqrt{2})$ in $\mathbb{Q}(\sqrt{2})[x]$.
2. Because $\mathbb{Q}(\sqrt{2})$ is a degree-2 field extension of $\mathbb{Q}$, there are no fields K for which $F \subsetneq K \subsetneq E$, so the second condition is vacuously true.

Non-Example: $\mathbb{Q}(\sqrt[3]{2})$ is *not* a splitting field for $x^3 - 2$ over $\mathbb{Q}$. This is because the polynomial $x^3 - 2$ does *not* split in $\mathbb{Q}(\sqrt[3]{2})[x]$. Certainly it has a root $\sqrt[2]{2}$, and therefore it has a linear factor, $(x - \sqrt[3]{2})$, but if we divide by this linear factor, we find that

$$x^3 - 2 = (x - \sqrt[3]{2})(x^2 + \sqrt[3]{2}x + (\sqrt[3]{2})^2)$$

and that the second factor in the above decomposition is irreducible in $\mathbb{Q}(\sqrt[3]{2})$ (the roots of it are complex, and everything in $\mathbb{Q}(\sqrt[3]{2})$ is real).

The second example raises the question: What *is* a splitting field for $\mathbb{Q}(\sqrt[3]{2})$ over $\mathbb{Q}$? Well, we could take $\mathbb{Q}(\sqrt[3]{2})$ and adjoin a root of $(x^2+\sqrt[3]{2}x+(\sqrt[3]{2})^2)$ (or, equivalently, take $\mathbb{Q}(\sqrt[3]{2})[x]/\langle(x^2+\sqrt[3]{2}x+(\sqrt[3]{2})^2)\rangle$). This idea of "keep adding roots of irreducible factors" is the core idea in the proof that every polynomial has a splitting field.

Proposition: For any field F and any $f(x) \in F[x]$, there is an extension E of F which is a splitting field for $f(x)$ over F.

Proof: Let $f(x) = p_1(x)p_2(x)\dots p_k(x)$ be the irreducible decomposition of $f(x)$. If each p_i is linear, then $f(x)$ already splits completely in $F[x]$ and F is itself a splitting field for $f(x)$ over F. Otherwise, let $p_i(x)$ be a non-linear irreducible factor of $f(x)$, and let $F_1 = F[x]/\langle p_i(x)\rangle$. Then F_1 is a field extension of F in which $p_i(x)$ has a root α_1, so $f(x)$ has more linear terms in $F_1[x]$ than in $F[x]$. If $f(x) \in F_1[x]$ still does not completely split, then we repeat the process. Eventually, $f(x)$ will split in $F_k[x]$, where $F_k = F(\alpha_1, \alpha_2, \dots, \alpha_k)$ for roots α_i.

Example: To form a splitting field for $x^3 - 2$ over $\mathbb{Q}$, the first step is to adjoin $\alpha_1 = \sqrt[3]{2}$ to get the extension $\mathbb{Q}(\sqrt[3]{2})$. In $\mathbb{Q}(\sqrt[3]{2})[x]$, the polynomial $x^3 - 2$ factors as

$$x^3 - 2 = (x - \sqrt[3]{2})(x^2 + \sqrt[3]{2}x + (\sqrt[3]{2})^2)$$

so we can adjoin a root of the non-linear term, $\alpha_2 = \sqrt[3]{2}(\frac{-1}{2} + \frac{\sqrt{3}}{2}i)$. Then the polynomial splits in

$$F(\sqrt[3]{2}, \sqrt[3]{2}(\frac{-1}{2} + \frac{\sqrt{3}}{2}i))$$

We can slightly simplify the above description of the splitting field by noticing that it's equal to

$$F(\sqrt[3]{2}, \sqrt{3}i)$$

Example: The splitting field for $(x^2-3)(x^2-5)$ is the field $\mathbb{Q}(\sqrt{3}, \sqrt{5})$. We saw that this has degree 4 over $\mathbb{Q}$ earlier in the course (Oct. 15).

Example: The polynomial $x^4 + 4 \in \mathbb{Q}[x]$ has irreducible factorization

$$x^4 + 4 = (x^2 + 2x + 2)(x^2 - 2x + 2)$$

The roots of the first factor in $\mathbb{C}$ are $-1+i$ and $-1-i$, and the roots of the second factor are $1+i$ and $1-i$. Adjoining all these roots to $\mathbb{Q}$ is the same as $\mathbb{Q}(i)$. The splitting field for $x^4 + 4$ over $\mathbb{Q}$ is $\mathbb{Q}(i)$.

Splitting fields are unique up to isomorphism

In the definition of a splitting field, it is not clear how many splitting fields there are for a single fixed polynomial over a single fixed field. In this section, we prove that any two such splitting fields are isomorphic.

Observation: If $\varphi : F \to F'$ is a homomorphism, then

$$F[x] \to F'[x]$$
$$a_0 + a_1x + a_2x^2 + \ldots \mapsto \varphi(a_0) + \varphi(a_1)x + \varphi(a_2)x^2 + \ldots$$

is a homomorphism of rings. We will call this homomorphism φ also (sorry, I know it's confusing, but this is common notation in maths). If $\varphi : F \to F'$ is an isomorphism, then the corresponding homomorphism $\varphi : F[x] \to F'[x]$ is also an isomorphism, and sends the ideal $\langle f(x)\rangle$ to the ideal $\langle \varphi(f(x))\rangle$.

Proposition: Let $\varphi : F \to F'$ be an isomorphism, and $f(x)$ be an irreducible polynomial in $F[x]$. If α is a root of f in some extension of F, and β is a root of $\varphi(f(x))$ in some extension of F', then there is an isomorphism $\tilde{\varphi} : F(\alpha) \to F'(\beta)$ such that $\tilde{\varphi}(\alpha) = \beta$, and $\tilde{\varphi}\big|_F = \varphi$.

Proof: Because $f(x)$ is irreducible, $\varphi(f(x))$ is irreducible too. Because φ sends $\langle f(x)\rangle$ to $\langle \varphi(f(x))\rangle$, the map $F[x] \mapsto F'[x]/\langle \varphi(f(x))\rangle$ will have kernel equal to $\langle f(x)\rangle$, so we have an isomorphism

$$\frac{F[x]}{\langle f(x)\rangle} \to \frac{F'[x]}{\langle \varphi(f(x))\rangle}$$

that sends a coset $[q(x)]$ to $[\varphi(q(x))]$. The composition of isomorphisms

$$F(\alpha) \to \frac{F[x]}{\langle f(x)\rangle} \to \frac{F'[x]}{\langle \varphi(f(x))\rangle} \to F'(\beta)$$

gives the desired isomorphism.

Theorem: Let $\varphi : F \to F'$ be an isomorphism, and $f(x)$ be *any* polynomial in $F[x]$. If E is a splitting field for $f(x)$ in F, and E' is a splitting field for $\varphi(f(x))$ in F', then there is an isomorphism $\tilde{\varphi} : E \to E'$ such that $\tilde{\varphi}\big|_F = \varphi$.

Proof: Let $p(x)$ be an irreducible factor of $f(x)$ of degree ≥ 2. Let $\alpha_1 \in E$ be a root for $p(x)$ and $\beta_1 \in E'$ be root for $\varphi(p(x))$. By the previous proposition, there is an isomorphism $F(\alpha_1) \to F'(\beta_1)$ that restricts to φ on F. If we repeat this process (until $f(x)$ no longer has any irreducible factors of degree ≥ 2, then we have a isomorphism $F(\alpha_1, \ldots, \alpha_k) \to F'(\beta_1, \ldots, \beta_k)$ that restricts to φ on F. Because $f(x)$ splits in $F(\alpha_1, \ldots, \alpha_k)$, and $\varphi(f(x))$ splits in $F'(\beta_1, \ldots, \beta_k)$, we have $E = F(\alpha_1, \ldots, \alpha_k)$ and $E' = F'(\beta_1, \ldots, \beta_k)$, which completes the proof.

Corollary: Let F be a field, and $f(x) \in F[x]$. Any two splitting fields for $f(x)$ over F are isomorphic.

Proof: Apply the previous theorem to the case when $F' = F$ and φ is the identity map.

Practice problems: Determine the splitting field over $\mathbb{Q}$ and their degrees over $\mathbb{Q}$ of the following polynomials

- $x^4 - 2 \in \mathbb{Q}[x]$
- $x^4 + 2 \in \mathbb{Q}[x]$
- $x^4 + x^2 + 1 \in \mathbb{Q}[x]$

Lecture 14: Algebraic Extension and Characteristic of a Field

Last time we discussed the existence of splitting fields of a polynomial. Today, we'll prove that every splitting field is unique, then we'll talk about algebraic closures and the characteristic of a field.

Splitting fields are unique up to isomorphism

In the definition of a splitting field, it is not clear how many splitting fields there are for a single fixed polynomial over a single fixed field. In this section, we prove that any two such splitting fields are isomorphic.

Observation: If $\varphi : F \to F'$ is a homomorphism, then

$$F[x] \to F'[x]$$
$$a_0 + a_1x + a_2x^2 + \ldots \mapsto \varphi(a_0) + \varphi(a_1)x + \varphi(a_2)x^2 + \ldots$$

is a homomorphism of rings. We will call this homomorphism φ also (sorry, I know it's confusing, but this is common notation in maths). If $\varphi : F \to F'$ is an isomorphism, then the corresponding homomorphism $\varphi : F[x] \to F'[x]$ is also an isomorphism, and sends the ideal $\langle f(x)\rangle$ to the ideal $\langle\varphi(f(x))\rangle$.

Proposition: Let $\varphi : F \to F'$ be an isomorphism, and $f(x)$ be an irreducible polynomial in $F[x]$. If α is a root of f in some extension of F, and β is a root of $\varphi(f(x))$ in some extension of F', then there is an isomorphism $\tilde{\varphi} : F(\alpha) \to F'(\beta)$ such that $\tilde{\varphi}(\alpha) = \beta$, and $\tilde{\varphi}\big|_F = \varphi$.

Proof: Because $f(x)$ is irreducible, $\varphi(f(x))$ is irreducible too. Because φ sends $\langle f(x)\rangle$ to $\langle\varphi(f(x))\rangle$, the map $F[x] \mapsto F'[x]/\langle\varphi(f(x))\rangle$ will have kernel equal to $\langle f(x)\rangle$, so we have an isomorphism

$$\frac{F[x]}{\langle f(x)\rangle} \to \frac{F'[x]}{\langle\varphi(f(x))\rangle}$$

that sends a coset $[q(x)]$ to $[\varphi(q(x))]$. The composition of isomorphisms

$$F(\alpha) \to \frac{F[x]}{\langle f(x)\rangle} \to \frac{F'[x]}{\langle\varphi(f(x))\rangle} \to F'(\beta)$$

gives the desired isomorphism.

Example: We can apply the above proposition to the identity isomorphism $\mathbb{Q} \to \mathbb{Q}$, the polynomial $f(x) = x^3 - 2 \in \mathbb{Q}[x]$, and the roots $\alpha = \sqrt[3]{2} \in \mathbb{R}$ and $\beta = \xi\sqrt[3]{2} \in \mathbb{C}$, where $\xi = e^{2\pi i}3$.

The isomorphism $\tilde{\varphi}$ from

$$\mathbb{Q}(\alpha) = \mathbb{Q}(\sqrt[3]{2}) = \{a_0 + a_1\sqrt[3]{2} + a_2(\sqrt[3]{2})^2 \mid a_i \in \mathbb{Q}\}$$

to

$$\mathbb{Q}(\beta) = \mathbb{Q}(\xi\sqrt[3]{2}) = \{b_0 + b_1\xi\sqrt[3]{2} + b_2(\xi\sqrt[3]{2})^2 \mid b_i \in \mathbb{Q}\}$$

is the identity on $\mathbb{Q}$ and sends $\sqrt[3]{2}$ to $\xi\sqrt[3]{2}$, so by the properties of homomorphisms it sends

$$a_0 + a_1\sqrt[3]{2} + a_2(\sqrt[3]{2})^2 \mapsto a_0 + a_1\xi\sqrt[3]{2} + a_2(\xi\sqrt[3]{2})^2$$

Notice also that the irreducible factorization of $f(x)$ in $\mathbb{Q}(\sqrt[3]{2})$ is given by

$$f(x) = (x - \sqrt[3]{2})(x^2 - \sqrt[3]{2}x + (\sqrt[3]{2})^2)$$

while the irreducible factorization of $f(x)$ in $\mathbb{Q}(\xi\sqrt[3]{2})$ is given by

$$f(x) = (x - \xi\sqrt[3]{2})(x^2 - \xi\sqrt[3]{2}x + (\xi\sqrt[3]{2})^2)$$

If we wanted to, we could apply the proposition again, applied to the isomorphism $\tilde{\varphi} : \mathbb{Q}(\sqrt[3]{2}) \to \mathbb{Q}(\xi\sqrt[3]{2})$ and the polynomial $(x^2 - \sqrt[3]{2}x + (\sqrt[3]{2})^2)$. The proposition would give us new field extensions of $\mathbb{Q}(\sqrt[3]{2})$ and $\mathbb{Q}(\xi\sqrt[3]{2})$, and an isomorphism between these two extensions that agrees with $\tilde{\varphi}$ (and therefore equals the identity on $\mathbb{Q}$). This logic is the key idea in the proof that splitting fields are unique.

Theorem: Let $\varphi : F \to F'$ be an isomorphism, and $f(x)$ be *any* polynomial in $F[x]$. If E is a splitting field for $f(x)$ in F, and E' is a splitting field for $\varphi(f(x))$ in F', then there is an isomorphism $\tilde{\varphi} : E \to E'$ such that $\tilde{\varphi}\big|_F = \varphi$.

Proof: Let $p(x)$ be an irreducible factor of $f(x)$ of degree ≥ 2. Let $\alpha_1 \in E$ be a root for $p(x)$ and $\beta_1 \in E'$ be root for $\varphi(p(x))$. By the previous proposition, there is an isomorphism $F(\alpha_1) \to F'(\beta_1)$ that restricts to φ on F. If we repeat this process (until $f(x)$ no longer has any irreducible factors of degree ≥ 2, then we have a isomorphism $F(\alpha_1, \ldots, \alpha_k) \to F'(\beta_1, \ldots, \beta_k)$ that restricts to φ on F. Because $f(x)$ splits in $F(\alpha_1, \ldots, \alpha_k)$, and $\varphi(f(x))$ splits in $F'(\beta_1, \ldots, \beta_k)$, we have $E = F(\alpha_1, \ldots, \alpha_k)$ and $E' = F'(\beta_1, \ldots, \beta_k)$, which completes the proof.

Corollary: Let F be a field, and $f(x) \in F[x]$. Any two splitting fields for $f(x)$ over F are isomorphic.

Proof: Apply the previous theorem to the case when $F' = F$ and φ is the identity map.

Algebraic Extensions

Recall that all finite fields are algebraic. Remember that whenever we adjoin a root of an irreducible polynomial to a field, the extension has degree equal to the degree of the polynomial, so this extension is algebraic. Also, since the construction of a splitting field consists of adjoining a finite number of elements to a field, this extension will also be finite and hence algebraic.

Proposition: Let $F \subseteq E$ and $E \subseteq K$ be algebraic field extensions. Then $F \subseteq K$ is an algebraic field extension.

Proof: It suffices to prove that $[F(\alpha) : F] < \infty$ for every $\alpha \in K$. Since α is algebraic over E, it is the root of some polynomial $e_0 + e_1x + \cdots + e_nx^n$, where $e_i \in E$. Then

$$[F(\alpha, e_0, e_1, \ldots, e_n) : F] = [F(\alpha, e_0, e_1, \ldots, e_n) : F(e_0, e_1, \ldots, e_n)][F(e_0, e_1, \ldots, e_n) : F]$$

and both terms on the right are finite (the first because α is algebraic over $F(e_0, \ldots, e_n)$, the second because each e_i is algebraic over F), so the term on the left is finite. Since $F(\alpha)$ is a subfield of $F(\alpha, e_0, \ldots, e_n)$, it must have finite degree over F also.

Corollary: Let E be a field extension of F. If $a, b \in E$ are algebraic over F, then so are $a+b, a-b, ab$, and a/b (assuming $b \neq 0$). Hence the set of elements of E that are algebraic over F is a field.

Proof: Because $F(a,b) = F(a)(b)$ has finite degree over F (by the previous proposition), then the subfields $F(a+b), F(a-b), F(ab), F(a/b)$ of $F(a,b)$ must also have finite degree over F, hence be algebraic extensions. Therefore, the elements $a+b$, $a-b$, ab, and a/b must all be algebraic over F.

Definition: Let E be a field extension of F. The **algebraic closure of F in E** is the subfield of E consisting of all elements of E that are algebraic over F. It is an algebraic extension of F.

Characteristic of a Field

Many of the fields that we study are subfields of $\mathbb{R}$ or $\mathbb{C}$ and definitely have infinitely many elements, while other fields we study like $\mathbb{Z}_5$ or $\mathbb{Z}_2/\langle x^2+x+1\rangle$ definitely have finitely many elements. There are some fields, such as the field of fractions of $\mathbb{Z}_{\ntrianglelefteq}[x]$, that have infinitely many elements, but still sort of share many similarities with finite fields, like the property that if you keep adding an element to itself, you'll eventually get zero. The definition of the *characteristic* of a field helps distinguish, conceptually, between "infinite fields that really behave like infinite fields" and "fields that may or may not be finite, but in some ways behave like finite fields."

Definition: The **characteristic** of a field is the smallest positive number n such that $1+1+\cdots+1=0$, where there are n copies of 1 written. If no such number exists (for example, in $\mathbb{Q}$) then the characteristic is zero.

Claim: The characteristic of a field is either zero or a prime number.

Proof: If the characteristic were composite, say $n = ab$, then the product of $(1+1+\cdots+1)$ (written a times) with $(1+1+\cdots+1)$ (written b times) would be zero. Since fields have no zero divisors, and n is defined to be the *smallest* positive integer such that n copies of 1 added together equals zero, this gives the contradiction.

Lecture 15: The Galois Group

Today, we're beginning the final third of the course, which studies automorphisms of fields.

Definition: An **automorphism** of a ring R is an isomorphism from R to itself.

Definition: If E is a field extension of F, an **F-automorphism of E** is an automorphism of E that fixes F. In other words, $\varphi(a) = a$ for all $a \in F$. The **Galois group Gal**(F/E) is the group whose elements are the F-automorphisms of E, and whose group operation is composition of automorphisms.

Ultimately, studying field automorphisms will show us that some quintic (degree 5) polynomials in $\mathbb{Q}[x]$ have roots that cannot be expressed using radicals (i.e. square roots or cube roots, or fourth roots, etc.). In other words, some quintics are "not solvable by radicals". The idea is similar to the proof that certain geometric straightedge-and-compass constructions are impossible. In those proofs, we associated a sequence of field extensions to a geometric construction, and by studying the *degree* of these extensions, we proved that certain constructions are impossible. To prove that certain quintics are not solvable by radicals, we will associate field extensions to the process of "solving something with radicals", and by studying *automorphisms* of these extensions, we can show that certain polynomials cannot be solved by radicals. Today, we'll just do lots of examples.

Example: Describe the Galois group $\mathrm{Gal}(\mathbb{Q}(\sqrt{2})/\mathbb{Q})$.

Remember that all elements of $\mathbb{Q}(\sqrt{2})$ can be written in the form $a+b\sqrt{2}$, where $a, b \in \mathbb{Q}$. Using the fact that any $\mathbb{Q}$-automorphism φ of $\mathbb{Q}(\sqrt{2})$ must send every element of $\mathbb{Q}$ to itself, and properties of homomorphisms, we see that

$$\varphi(a + b\sqrt{2}) = \varphi(a) + \varphi(b)\varphi(\sqrt{2}) = a + b\varphi(\sqrt{2}).$$

So knowing what φ does to $\sqrt{2}$ tells us what φ does to *every* element. But $\varphi(\sqrt{2})$ can't be just *any* element of $\mathbb{Q}(\sqrt{2})$: notice that

$$\varphi(\sqrt{2}\sqrt{2} - 2) = \varphi(\sqrt{2})\varphi(\sqrt{2}) - \varphi(2) = \varphi(\sqrt{2})^2 - 2$$

and

$$\varphi(\sqrt{2}\sqrt{2} - 2) = \varphi(0) = 0.$$

This means that $\varphi(\sqrt{2})$ must be a root of the equation $x^2 - 2$, so $\varphi(\sqrt{2})$ is $\sqrt{2}$ or $-\sqrt{2}$.

Therefore, the only possible $\mathbb{Q}$-automorphisms of $\mathbb{Q}(\sqrt{2})$ are

$$\begin{aligned} \mathrm{id} : \mathbb{Q}(\sqrt{2}) &\to \mathbb{Q}(\sqrt{2}) \\ a + b\sqrt{2} &\mapsto a + b\sqrt{2} \end{aligned}$$

and

$$\begin{aligned} \tau : \mathbb{Q}(\sqrt{2}) &\to \mathbb{Q}(\sqrt{2}) \\ a + b\sqrt{2} &\mapsto a - b\sqrt{2} \end{aligned}$$

It's straightforward to verify that both of these are indeed $\mathbb{Q}$-automorphisms, and that $\tau \circ \tau = \mathrm{id}$. Therefore, $\mathrm{Gal}(\mathbb{Q}(\sqrt{2})/\mathbb{Q})$ consists of two elements id and τ, with group operation given by $\mathrm{id} \cdot \tau = \tau \cdot \mathrm{id} = \tau$, and $\mathrm{id} \cdot \mathrm{id} = \tau \cdot \tau = \mathrm{id}$.

Two things that helped us determine the splitting field in the above case. The first was that the whole automorphism was determined by where $\sqrt{2}$ was sent. And the second was that $\sqrt{2}$ could only get sent to $\sqrt{2}$ or $-\sqrt{2}$, which are the roots of $x^2 - 2$. These properties are formalized in the next two propositions.

Proposition: Let $\alpha_1, \dots, \alpha_n$ be algebraic over F. An F-automorphism of $F(\alpha_1, \alpha_2, \dots, \alpha_n)$ is completely determined by where it sends $\alpha_1, \dots, \alpha_n$. In other words, if φ and ψ are two F-automorphisms such that $\varphi(\alpha_i) = \psi(\alpha_i)$ for all i, then φ and ψ are the same F-automorphism.

Proof: Recall that the field $F(\alpha_1, \dots, \alpha_n)$ is spanned (as a vector space) over F by elements of the form

$$\alpha_1^{k_1} \alpha_2^{k_2} \dots \alpha_n^{k_n}$$

If you write an element in terms of this basis and apply the properties of homomorphisms and the fact that any F-automorphism must take every element of F to itself, you learn that the image of an F-automorphism is determined by the elements $\varphi(\alpha_i)$.

Proposition: Let E be an extension of F and $f(x) \in F[x]$. Any F-automorphism φ of E will send roots of $f(x)$ to other roots of $f(x)$.

Proof: Suppose $f(x) = a_0 + a_1 x + \cdots + a_n x^n$, where $a_i \in F$. If $f(\alpha) = 0$, then

$$\begin{aligned} f(\varphi(\alpha)) &= a_0 + a_1\varphi(\alpha) + a_2\varphi(\alpha)^2 + \cdots + a_n\varphi(\alpha)^n \\ &= \varphi(a_0) + \varphi(a_1)\varphi(\alpha) + \varphi(a_2)\varphi(\alpha)^2 + \cdots + \varphi(a_n)\varphi(\alpha)^n \\ &= \varphi(a_0) + \varphi(a_1)\varphi(\alpha) + \varphi(a_2)\varphi(\alpha^2) + \cdots + \varphi(a_n)\varphi(\alpha^n) \\ &= \varphi(a_0) + \varphi(a_1\alpha) + \varphi(a_2\alpha^2) + \cdots + \varphi(a_n\alpha^n) \\ &= \varphi(a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n) \\ &= \varphi(0) \\ &= 0 \end{aligned}$$

Example: Describe the Galois group $\mathrm{Gal}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$.

Fill in the blanks! What information do you need to describe the a $\mathbb{Q}$-automorphism φ of $\mathbb{Q}(\sqrt[3]{2})$ completely?

What could $\varphi(\sqrt[3]{2})$ possibly be?

List all elements of $\mathrm{Gal}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$

In the previous example, we saw that the Galois group was very small because the other roots of $x^3 - 2$ were not contained in the field $\mathbb{Q}(\sqrt[3]{2})$, so there was nowhere that a $\mathbb{Q}$-automorphism could send $\sqrt[3]{2}$ except to $\sqrt[3]{2}$ itself! In the next example, we'll see that the Galois group is much bigger if we take the *splitting field* of $x^3 - 2$, so that all the roots of $x^3 - 2$ are available as images of $\sqrt[3]{2}$.

Example: Let E be the splitting field for $x^3 - 2$ over $\mathbb{Q}$. Describe the Galois group $\mathrm{Gal}(E/\mathbb{Q})$.

Let $\xi = e^{\frac{2\pi i}{3}}$, and recall that the splitting field for $x^3 - 2$ can be obtained by adjoining the roots of $x^3 - 2$ to $\mathbb{Q}$, so $E = \mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{2}\xi, \sqrt[3]{2}\xi^2)$. Equivalently, $E = \mathbb{Q}(\sqrt[3]{2}, \xi)$. What information do you need to describe the a $\mathbb{Q}$-automorphism φ of E completely?

What could $\varphi(\sqrt[3]{2})$ possibly be? What could $\varphi(\xi)$ possibly be?

List all elements of $\mathrm{Gal}(E/\mathbb{Q})$

In the above example, we used the fact that the element ξ is a root of the polynomial x^2+x+1 to help determine its possible images under $\mathbb{Q}$-automorphisms of E. In any algebraic field extension, *every* element α of the larger field is the root of some polynomial with coefficients in the smaller field. In fact, there will be a unique monic irreducible polynomial for which α is a root.

Proposition: Let α be algebraic over F. There is a unique monic irreducible polynomial in $F[x]$ for which α is a root.

Proof: Let $I \subseteq F[x]$ be the ideal of polynomials for which α is a root. Because $F[x]$ is a principal ideal domain, $I = \langle p(x) \rangle$ for some polynomial $p(x)$, which must be irreducible (otherwise one of its irreducible factors would have α as a root, but this factor would not be in $\langle p(x) \rangle = I$, a contradiction). By dividing $p(x)$ by its leading coefficient, we may assume that $p(x)$ is monic. Every polynomial that has α as a root is a multiple of $p(x)$, so either must be reducible or must not be monic, so $p(x)$ is the unique irreducible polynomial having α as a root.

Definition: Let α be algebraic over F. The **minimal polynomial of α over F** is the unique monic irreducible polynomial in $F[x]$ having α as a root.

Example: Describe the Galois group $\mathrm{Gal}(\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q})$.

Fill in the blanks! What information do you need to describe the a $\mathbb{Q}$-automorphism φ of E completely?

What could $\varphi(\sqrt{2})$ possibly be? What could $\varphi(\sqrt{3})$ possibly be?

List all elements of $\mathrm{Gal}(E/\mathbb{Q})$

Practice problems: .

1. Describe the Galois group of the splitting field of x^4+1 over $\mathbb{Q}$.
2. Describe the Galois group of the splitting field of x^2+9 over $\mathbb{Q}$.
3. Describe the Galois group of the splitting field of $x^2-10x+21$ over $\mathbb{Q}$.

Lecture 16: The Fundamental Theorem of Galois Theory

Last Wednesday we defined the concept of a Galois group, and studied two examples of Galois groups. Today, we'll learn the Fundamental theorem of Galois theory, which describes a bijection between subfields of certain field extensions and the subgroups of the corresponding Galois group. To motivate it, we begin by revisiting an example from last Wednesday.

Example

Let E be the splitting field for $x^3 - 2$ over $\mathbb{Q}$. We saw last week that $E = \mathbb{Q}(\sqrt[3]{2}, \xi)$, where $\xi = e^{\frac{2\pi i}{3}}$, and that every $\mathbb{Q}$-automorphism of E could be described by where it sent $\sqrt[3]{2}$ and ξ, and that the only places where $\sqrt[3]{2}$ could be sent are roots of $x^3 - 2$, and that the only places where ξ could be sent are roots of $x^2 + x + 1$. This logic produced the following list of elements of $\mathrm{Gal}(E/\mathbb{Q})$. Fill in the blanks!

$$\mathrm{id} : \begin{cases} \xi \mapsto \xi \\ \sqrt[3]{2} \mapsto \sqrt[3]{2} \end{cases} \qquad\qquad \tau : \begin{cases} \xi \mapsto \xi^2 \\ \sqrt[3]{2} \mapsto \sqrt[3]{2} \end{cases}$$

$$\sigma : \begin{cases} \xi \mapsto \xi \\ \sqrt[3]{2} \mapsto \xi\sqrt[3]{2} \end{cases} \qquad\qquad \sigma^2 : \begin{cases} \xi \mapsto \\ \sqrt[3]{2} \mapsto \end{cases}$$

$$\tau\sigma : \begin{cases} \xi \mapsto \\ \sqrt[3]{2} \mapsto \end{cases} \qquad\qquad \tau\sigma^2 : \begin{cases} \xi \mapsto \\ \sqrt[3]{2} \mapsto \end{cases}$$

Draw a lattice for the subfields of E and a lattice for the subgroups of $\mathrm{Gal}(E/F)$.

The lattices look similar, but upside down. For every subfield K of E, there is a corresponding subgroup of $\text{Gal}(E/\mathbb{Q})$ consisting of the automorphisms that fix every element of K (that is, $\text{Gal}(E/K)$). Going the other way, for every subgroup H of $\text{Gal}(E/\mathbb{Q})$, there is a subfield of E consisting of the elements of E that are fixed by every automorphism in H.

Definition: Let E be an extension of F, and let H be a subgroup of $\text{Gal}(E/F)$. The **fixed field of** H is the set

$$E_H := \{x \in E \mid \varphi(x) = x \text{ for all } \varphi \in H\}$$

This correspondence between subfields of E and subgroups of $\text{Gal}(E/\mathbb{Q})$ is inclusion reversing: if J and H are subgroups of $\text{Gal}(E/\mathbb{Q})$ and $J \subseteq H$, then every field element fixed by all the automorphisms in H will also be fixed by all the automorphisms in J. This means the fixed field of J must be bigger than or equal to the fixed field of H, i.e. $E_J \supseteq E_H$. This is why the subgroup lattice is an "upside down" version of the subfield lattice. The fundamental theorem of Galois theory says that this inclusion-reversing bijection between subfields and subgroups always happens, as long as the fields in question aren't too weird (they're finite or characteristic zero), and the extension you're studying is a splitting field of some polynomial.

Fundamental Theorem of Galois Theory

Let F be a finite field or a field of characteristic zero. Let E be the splitting field over F of some polynomial in $F[x]$. Then the following two maps are inverses (and hence define a bijection)

$$\left\{\begin{array}{c}\text{Subfields of}\\ \text{E containing F}\end{array}\right\} \leftrightarrow \left\{\begin{array}{c}\text{Subgroups of}\\ \text{Gal(E/F)}\end{array}\right\}$$

$$K \mapsto \text{Gal}(E/K)$$

$$E_H \leftarrow\!\shortmid H$$

Furthermore, for any subfield K of E containing F, the following are true

1. $[E:K] = |\text{Gal}(E/K)|$ and $[K:F] = |\text{Gal}(E/F)|/|\text{Gal}(E/K)|$.
2. If K is the splitting field over F of some polynomial, then $\text{Gal}(E/K)$ is a *normal* subgroup of $\text{Gal}(E/F)$, and $\text{Gal}(K/F) \cong \text{Gal}(E/F)/\text{Gal}(E/K)$.

Example

Let E be the splitting field for x^7-1, and suppose we know the (true) fact that $x^6+x^5+\cdots+x+1$ is irreducible in $\mathbb{Q}[x]$. Let's study the subfields of E over $\mathbb{Q}$ by studying the subgroups of $\text{Gal}(E/\mathbb{Q})$ and using the fundamental theorem of Galois theory.

First, let $\xi = e^{\frac{2\pi i}{7}}$. The roots of x^7-1 are the elements $\{1, \xi, \xi^2, \ldots, \xi^6\}$, so $E = \mathbb{Q}(\xi)$, and $[E:\mathbb{Q}] = 6$. Because the minimal polynomial for 1 is $x-1$, and the minimal polynomial for all other roots is $x^6+x^5+\cdots+x+1$, it follows that every $\mathbb{Q}$-automorphism must take 1 to itself, and permute the roots of $x^6+x^5+\cdots+x+1$. But be careful: not *every* permutation of these roots represents a valid $\mathbb{Q}$-automorphism – for example, by the properties of homomorphisms, if $\xi \mapsto \xi^2$, then that forces $\xi^2 \mapsto (\xi^2)(\xi^2) = \xi^4$.

Draw the $\mathbb{Q}$-automorphism of E given by $\xi \mapsto \xi^2$ by showing how it permutes the roots of $x^7 - 1$. What is the order of the element of $\mathrm{Gal}(E/F)$ that it represents?

Draw the $\mathbb{Q}$-automorphism of E given by $\xi \mapsto \xi^3$ by showing how it permutes the roots of $x^7 - 1$. What is the order of the element of $\mathrm{Gal}(E/F)$ that it represents?

Notice that this second $\mathbb{Q}$-automorphism, τ, has order 6, so since $|\mathrm{Gal}(E/\mathbb{Q})| = [E : \mathbb{Q}] = 6$, it must be the case that τ generates the entire group $\mathrm{Gal}(E/F) \cong \mathbb{Z}_6$. Draw the lattice of subgroups of $\mathrm{Gal}(E/F)$ and the lattice of subfields of E containing F.

Practice problems: .

1. Draw the lattice of subfields for the splitting field E of $x^5 - 1$ over $\mathbb{Q}$ and the lattice of subgroups of $\mathrm{Gal}(E/\mathbb{Q})$. You may use the fact that $x^4 + x^3 + x^2 + x + 1$ is irreducible in $\mathbb{Q}[x]$.
2. Draw the lattice of subfields for the splitting field E of $x^4 - 2$ over $\mathbb{Q}$ and the lattice of subgroups of $\mathrm{Gal}(E/\mathbb{Q})$.

Lecture 17: FTOGT Examples

Monday, we stated the fundamental theorem of Galois theory:

Fundamental Theorem of Galois Theory

Let F be a finite field or a field of characteristic zero. Let E be the splitting field over F of some polynomial in $F[x]$. Then the following two maps are inverses (and hence define a bijection)

$$\left\{\begin{array}{c}\text{Subfields of}\\ \text{E containing F}\end{array}\right\} \leftrightarrow \left\{\begin{array}{c}\text{Subgroups of}\\ \text{Gal(E/F)}\end{array}\right\}$$

$$K \mapsto \mathrm{Gal}(E/K)$$

$$E_H \leftarrow\!\shortmid H$$

Furthermore, for any subfield K of E containing F, the following are true

1. $[E:K] = |\mathrm{Gal}(E/K)|$ and $[K:F] = |\mathrm{Gal}(E/F)|/|\mathrm{Gal}(E/K)|$.
2. If K is the splitting field over F of some polynomial, then $\mathrm{Gal}(E/K)$ is a *normal* subgroup of $\mathrm{Gal}(E/F)$, and $\mathrm{Gal}(K/F) \cong \mathrm{Gal}(E/F)/\mathrm{Gal}(E/K)$.

Today, we do two examples to illustrate the theory, in which we compare the lattice of subfields of an extension with the lattice of subgroups of its Galois group.

Example

Let E be the splitting field for x^7-1, and suppose we know the (true) fact that $x^6+x^5+\cdots+x+1$ is irreducible in $\mathbb{Q}[x]$. Let $\xi = e^{\frac{2\pi i}{7}}$. The roots of x^7-1 are the elements $\{1, \xi, \xi^2, \ldots, \xi^6\}$, so

$$\begin{aligned} E &= \mathbb{Q}(\xi, \xi^2, \ldots, \xi^6) \\ &= \mathbb{Q}(\xi) \end{aligned}$$

Because the minimal polynomial for ξ is $x^6+x^5+\cdots+x+1$, it follows that $[E:\mathbb{Q}] = 6$.

How many elements are there of $\mathrm{Gal}(E/\mathbb{Q})$?

Any $\mathbb{Q}$-automorphism of E is determined by where it sends ξ. Where could a $\mathbb{Q}$-automorphism of E possibly send ξ?

Draw the $\mathbb{Q}$-automorphism of E given by $\xi \mapsto \xi^2$ by showing how it permutes the roots of $x^7 - 1$. What is the order of the element of $\mathrm{Gal}(E/F)$ that it represents?

Draw the $\mathbb{Q}$-automorphism of E given by $\xi \mapsto \xi^3$ by showing how it permutes the roots of $x^7 - 1$. What is the order of the element of $\mathrm{Gal}(E/F)$ that it represents?

Notice that this second $\mathbb{Q}$-automorphism, τ, has order 6, so it must generates the entire group $\mathrm{Gal}(E/F) \cong \mathbb{Z}_6$. Draw the lattice of subgroups of $\mathrm{Gal}(E/F)$ and the lattice of subfields of E containing F.

Example

Let E be the splitting field for $x^4 - 2$ over $\mathbb{Q}$. Let's study the subfields of E and the subgroups of $\mathrm{Gal}(E/\mathbb{Q})$.

First, we know that the roots of $x^4 - 2$ in $\mathbb{C}$ are $\{\sqrt[4]{2}, i\sqrt[4]{2}, -\sqrt[4]{2}, -i\sqrt[4]{2}\}$, so

$$\begin{aligned} E &= \mathbb{Q}(\sqrt[4]{2}, i\sqrt[4]{2}, -\sqrt[4]{2}, -i\sqrt[4]{2}) \\ &= \mathbb{Q}(\sqrt[4]{2}, i) \end{aligned}$$

What is the degree $[E : \mathbb{Q}]$? What does this tell you about $\mathrm{Gal}(E/\mathbb{Q})$?

Describe the elements of $\mathrm{Gal}(E/\mathbb{Q})$ in terms of where they send the elements $\sqrt[4]{2}$ and i. Draw the permutation of the roots of $x^4 - 2$ given by each.

What familiar group is $\mathrm{Gal}(E/\mathbb{Q})$ isomorphic to? Draw the lattice of subgroups of $\mathrm{Gal}(E/\mathbb{Q})$.

Solvability

The fundamental theorem of Galois theory gives us the understanding about Galois groups that we need to start proving things about solvability of quintics. Recall that we were able to show that certain geometric constructions are impossible by translating facts about geometric constructions into facts about degrees of field extensions. Similarly, we can also translate facts about "having roots that can be expressed just in terms of rational numbers and symbols $\sqrt[n]{}$" into facts about Galois groups. To study this problem, we need to define rigorously the concept of "having roots that can be expressed in terms of rational numbers and $\sqrt[n]{}$."

Definition: Let F be a field. A polynomial $f(x) \in F[x]$ is **solvable by radicals over** F if it splits in some extension field of the form $F(a_1, \dots, a_n)$, where the elements a_i have the property that there are positive integers $k_1, \dots, k_n$ such that

$$\begin{aligned} a_1^{k_1} &\in F \\ a_2^{k_2} &\in F(a_1) \\ a_3^{k_3} &\in F(a_1, a_2) \\ &\vdots \\ a_n^{k_n} &\in F(a_1, a_2, \dots, a_{n-1}) \end{aligned}$$

Soon, we will prove that whenever a polynomial is solvable by radicals, the Galois group of its splitting field must be a *solvable* group.

Definition: A group G is **solvable** if there are subgroups $H_0, H_1, \dots, H_k$ such that

$$\{e\} = H_0 \subseteq H_1 \subseteq H_2 \subseteq \cdots \subseteq H_n = G$$

with the property that each H_i is normal in H_{i+1}, and H_{i+1}/H_i is abelian.

Next class, we'll see examples of solvable groups and study their properties.

Lecture 18: More on Solvable Groups

Last class, we interrupted our discussion of Galois groups to define two concepts: the notion of a polynomial being *solvable by radicals*, and the definition of a *solvable group.* Today we'll learn some facts about solvable groups, then give some examples of Galois groups that are solvable. We'll prove on Monday that whenever a polynomial is solvable by radicals, the Galois group of its splitting field is a solvable group.

Proposition: Let N be a normal subgroup of a group G. If G is solvable, then G/N is solvable.

Proof: Because G is solvable, there are subgroups

$$\{\mathrm{id}\} = H_0 \subseteq H_1 \subseteq \cdots \subseteq H_n = G$$

such that H_i is normal in H_{i+1} and H_{i+1}/H_i is abelian. Let $\varphi : G \to G/N$ be the quotient homomorphism. We will verify that the sequence of subgroups

$$\{\mathrm{id}\} = \varphi(H_0) \subseteq \varphi(H_1) \subseteq \cdots \subseteq \varphi(H_n) = G/N$$

satisfies the conditions needed to prove that G/N is solvable.

$\varphi(H_i)$ is normal in $\varphi(H_{i+1})$: Let $y \in \varphi(H_i)$ and $x \in \varphi(H_{i+1})$, so $y = \varphi(\tilde{y})$ and $x = \varphi(\tilde{x})$, where $\tilde{y} \in H_i$ and $\tilde{x} \in H_{i+1}$. Then

$$xyx^{-1} = \varphi(\tilde{x})\varphi(\tilde{y})\varphi(\tilde{x})^{-1} = \varphi(\tilde{x}\tilde{y}\tilde{x}^{-1}) = \varphi(\tilde{y}')$$

for some $\tilde{y}' \in H_i$ (using the fact that H_i is normal in H_{i+1}). This means that $xyx^{-1} \in \varphi(H_i)$, so H_i is normal in H_{i+1}.

$\varphi(H_{i+1})/\varphi(H_i)$ is abelian: Consider the homomorphism

$$H_{i+1} \to \varphi(H_{i+1})/\varphi(H_i)$$

given by following φ with the quotient homomorphism. This is a surjective homomorphism that contains H_i, so it defines a homomorphism $H_{i+1}/H_i \to \varphi(H_{i+1})/\varphi(H_i)$. By the first isomorphism theory of group theory, $\varphi(H_{i+1})/\varphi(H_i)$ is the quotient of an abelian group, hence is itself abelian.

Proposition: Let N be a normal subgroup of a group G. If N and G/N are solvable, then G is solvable.

Proof: Let

$$\{\mathrm{id}\} = H_0 \subseteq H_1 \subseteq \cdots \subseteq H_n = N \text{ and}$$
$$\{\mathrm{id}\} = J_0 \subseteq J_1 \subseteq \cdots \subseteq J_k = G/N$$

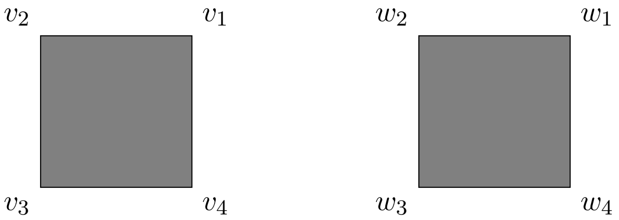
be sequences of subgroups showing that N and G/N are solvable. To find such a sequence of subgroups of G, let $\varphi : G \to G/N$ be the quotient homomorphism and consider the sequence.

$$\{\mathrm{id}\} = H_0 \subseteq H_1 \subseteq \cdots \subseteq H_n = N = \varphi^{-1}(J_0) \subseteq \varphi^{-1}(J_1) \subseteq \cdots \subseteq \varphi^{-1}(J_k) = G$$

The fact that the solvability conditions are satisfied for the H_i subgroups follows from the fact that they arise from the sequence of subgroups that shows N is solvable. Then, consider the surjective homomorphism $\varphi^{-1}(J_{i+1}) \to J_{i+1}/J_i$ given by following φ by the quotient homomorphism. Its kernel is $\varphi^{-1}(J_i)$, so $\varphi^{-1}(J_i)$ must be normal, and the image is J_{i+1}/J_i, so by the first isomorphism theorem of Group theory, $\varphi^{-1}(J_{i+1})/\varphi^{-1}(J_i) \cong J_{i+1}/J_i$ which is abelian.

Example

We showed last time that the dihedral group D_4 is solvable. What about the subgroup of the symmetries of *two* squares?



To be precise, by "symmetry of two squares", I mean any permutation σ of the 8 verticies above such that drawing the lines

$$\sigma(v_1) - \sigma(v_2) - \sigma(v_3) - \sigma(v_4) - \sigma(v_1) \quad \text{and} \quad \sigma(w_1) - \sigma(w_2) - \sigma(w_3) - \sigma(w_4) - \sigma(w_1)$$

still makes the two squares in the figure above. Let G be this group of symmetries, H be the subgroup of G consisting of the symmetries that don't exchange the two squares, and let K be the subgroup of H that consists the symmetries that fix every one of the w verticies. Clearly, $K \cong D_4$, and $H \cong D_4 \oplus D_4$. We have the sequence of subgroups $K \subseteq H \subseteq G$. $K \cong D_4$ is solvable, and $H/K \cong D_4$ is also solvable, so $H \cong D_4 \oplus D_4$ is solvable. Similarly, since H is solvable and $G/H \cong \mathbb{Z}_2$ is solvable, then G is solvable.

Solvable Galois Groups

Proposition: Let F be any subfield of $\mathbb{C}$, and n be a positive integer, and E be the splitting field for $x^n - 1$ over F. Then $\text{Gal}(E/F)$ is abelian.

Proof: Let $\varphi, \psi \in \text{Gal}(E/F)$. We will prove the claim by verifying that $\varphi\psi = \psi\varphi$.

Let $\xi = e^{\frac{2\pi i}{n}}$. The roots of $x^n - 1$ are $\{1, \xi, \xi^2, \ldots, \xi^{n-1}\}$, and $E = F(\xi)$. Any F-automorphism is determined by its image of ξ, and each F-automorphism sends ξ to a power of ξ. Suppose $\varphi(\xi) = \xi^a$ and $\psi(\xi) = \xi^b$. Then

$$\psi\varphi(\xi) = \psi(\xi^a) = \psi(\xi)^a = \xi^{ab} \text{ and}$$
$$\varphi\psi(\xi) = \varphi(\xi^b) = \varphi(\xi)^b = \xi^{ab}$$

so $\psi\varphi(\xi) = \varphi\psi(\xi)$. Since any F-automorphism is determined by where it sends ξ, this proves that $\psi\varphi = \varphi\psi$, so $\text{Gal}(E/F)$ is abelian.

Proposition: Let F be any subfield of $\mathbb{C}$, n be a positive integer, $c \in \mathbb{R}$, and let E be the splitting field for $x^n - c$ over F. The Galois group $\text{Gal}(E/F)$ is solvable.

Proof: Let $\xi = e^{\frac{2\pi i}{n}}$, so the splitting field E for $x^n - c$ over F is given by $E = F(\xi, \sqrt[n]{c})$. By the fundamental theorem of Galois theory, the field extensions $F \subseteq F(\xi) \subseteq E = F(\xi, \sqrt[n]{c})$ correspond to the subgroups

$$\{\text{id}\} \subseteq \text{Gal}(E/F(\xi)) \subseteq \text{Gal}(E/F)$$

Because $F(\xi)$ is the splitting field for $x^n - 1$ over F, it follows that $\text{Gal}(E/F(\xi))$ is a normal subgroup of $\text{Gal}(E/F)$ and that $\text{Gal}(E/F)/\text{Gal}(E/F(\xi)) \cong \text{Gal}(F(\xi)/F)$, which is normal by the previous proposition.

It remains only to show that $\mathrm{Gal}(E/F(\xi))$ is abelian. To show this, note that the zeros of $x^n - c$ are $\{\sqrt[n]{c}, \sqrt[n]{c}\xi, \dots, \sqrt[n]{c}\xi^{n-1}\}$, so $E = F(\xi)(\sqrt[n]{c})$, and every $F(\xi)$-automorphism of E is determined by which zero of $x^n - c$ it sends $\sqrt[n]{c}$.

Suppose $\varphi(\sqrt[n]{c}) = \sqrt[n]{c}\xi^a$ and $\psi(\sqrt[n]{c}) = \sqrt[n]{c}\xi^b$. Then

$$\begin{aligned} \psi\varphi(\sqrt[n]{c}) &= \psi(\sqrt[n]{c}\xi^a) = \psi(\sqrt[n]{c})\xi^a = \sqrt[n]{c}\xi^{a+b} \text{ and} \\ \varphi\psi(\sqrt[n]{c}\xi) &= \varphi(\sqrt[n]{c}\xi^b) = \varphi(\sqrt[n]{c})\xi^b = \sqrt[n]{c}\xi^{a+b} \end{aligned}$$

so $\psi\varphi = \varphi\psi$. Therefore, $\mathrm{Gal}(E/F(\xi))$ is abelian.

Lecture 19: Solvable by Radicals implies Solvable Galois Group

We're finally ready to prove the relationship between the notion of a polynomial being "solvable by radicals" and a group being "solvable".

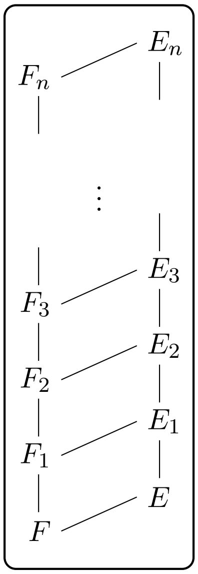
Theorem: Let F be a subfield of $\mathbb{C}$, let E be the splitting field of $f(x) \in F[x]$. If $f(x)$ is solvable by radicals over F, then $\text{Gal}(E/F)$ is solvable.

Proof: Suppose $f(x)$ is solvable by radicals. Then there are elements $a_1, \ldots, a_n$ such that $F(a_1, \ldots, a_n)$ contains E and for all $i \leq n$, there is some positive integer k_i such that $a_i^{k_i} \in F(a_1, \ldots, a_{i-1})$.

Let $c_i \in F(a_1, \ldots, a_{i-1})$ be the element $a_i^{k_i}$.

The proof involves studying the two towers of field extensions described below.

$$\begin{aligned}
F_1 &= \text{spl. field of } x^{k_1} - c_1 \text{ over } F \\
&= F(\text{roots of } x^{k_1} - c_1) \\
F_2 &= \text{spl. field of } x^{k_2} - c_2 \text{ over } F_1 \\
&= F(\text{roots of } x^{k_1} - c_1 \text{ and roots of } x^{k_2} - c_2) \\
&\vdots \\
F_n &= \text{spl. field of } x^{k_n} - c_n \text{ over } F_{n-1} \\
&= F(\text{roots of } x^{k_1} - c_1 \text{ and } x^{k_2} - c_2 \ldots \text{ and } x^{k_{n-1}} - c_{n-1})
\end{aligned}$$



and

$$\begin{aligned}
E_1 &= \text{spl. field of } x^{k_1} - c_1 \text{ over } E \\
&= E(\text{roots of } x^{k_1} - c_1) \\
E_2 &= \text{spl. field of } x^{k_2} - c_2 \text{ over } E_1 \\
&= E(\text{roots of } x^{k_1} - c_1 \text{ and roots of } x^{k_2} - c_2) \\
&\vdots \\
E_n &= \text{spl. field of } x^{k_n} - c_n \text{ over } E_{n-1} \\
&= E(\text{roots of } x^{k_1} - c_1 \text{ and } x^{k_2} - c_2 \ldots \text{ and } x^{k_{n-1}} - c_{n-1})
\end{aligned}$$

First, let's notice that E_i is the splitting field for $f(x) \in F_i[x]$ over F_i. Because F_n contains each of the elements a_i, it contains all of $F(a_1, \ldots, a_n)$ and hence contains all roots of $f(x)$ already. So $F_n = E_n$.

If we study any one of the regions made of $F_i, E_i, F_{i+1}, E_{i+1}$ (including $i = 0$, where $F_0 = F$ and $E_0 = E$), we notice that

- E_i is the splitting field for $f(x)$ over F_i
- F_{i+1} is the splitting field for $x^{k_{i+1}} - c_{i+1}$ over F_i

- E_{i+1} is the splitting field for $f(x)(x^{k_{i+1}} - c_{i+1})$ over F_i

Therefore, the fundamental theorem of Galois theory tells us that

$$\mathrm{Gal}(F_{i+1}/F_i) \cong \frac{\mathrm{Gal}(E_{i+1}/F_i)}{\mathrm{Gal}(E_{i+1}/F_{i+1})} \quad \text{and} \quad \mathrm{Gal}(E_i/F_i) \cong \frac{\mathrm{Gal}(E_{i+1}/F_i)}{\mathrm{Gal}(E_{i+1}/E_i)}$$

from which we can apply our knowledge about solvable groups to deduce that

$$\mathrm{Gal}(E_{i+1}/F_{i+1}) \text{ solvable} \Rightarrow \mathrm{Gal}(E_{i+1}/F_i) \text{ solvable} \quad \text{and}$$
$$\mathrm{Gal}(E_{i+1}/F_i) \text{ solvable} \Rightarrow \mathrm{Gal}(E_i/F_i) \text{ solvable}$$

which, when combine, give

$$\mathrm{Gal}(E_{i+1}/F_{i+1})\text{solvable} \Rightarrow \mathrm{Gal}(E_i/F_i)\text{solvable}$$

Because we know that $F_n = E_n$, it follows that $\mathrm{Gal}(E_n/F_n)$ is the trivial group (and hence is solvable). Then, the composition of implications

$$\mathrm{Gal}(E_n/F_n)\text{solvable} \Rightarrow \mathrm{Gal}(E_{n-1}/F_{n-1})\text{solvable} \Rightarrow \cdots \Rightarrow \mathrm{Gal}(E/F)\text{solvable}$$

proves the claim.

The converse of the statement is also true: if $\mathrm{Gal}(E/F)$ is solvable, then $f(x)$ is solvable by radicals. We will not prove it. However, it has some nice consequences.

Claim: Let $f(x) \in \mathbb{Q}[x]$ be a polynomial of degree ≤ 4. Then $f(x)$ is solvable by radicals.

Proof: Let E be the splitting field of $f(x)$. We know (from the homework assignment) that $\mathrm{Gal}(E/F)$ is a subgroup of S_n, where n is the degree of f.

- $S_2 \cong \mathbb{Z}_2$ is abelian, so it is solvable.
- S_3 has a normal subgroup A_3 (the *alternating group*) consisting of the even permutations. The chain of subgroups $\{\mathrm{id}\} \subseteq A_3 \subseteq S_3$ satisfies the conditions necessary to show that S_3 is solvable.
- S_4 has a normal subgroup A_4 consisting of the even permutations. A_4 has a normal subgroup isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$, which is abelian. The chain of subgroups $\{\mathrm{id}\} \subseteq \mathbb{Z}_2 \times \mathbb{Z}_2 \subseteq A_4 \subseteq S_4$ satisfies the conditions necessary to show that S_4 is solvable.

Because every subgroup of a solvable group is solvable, and $\mathrm{Gal}(E/F)$ is a subgroup of either S_2, S_3, or S_4, this shows that $f(x)$ is solvable by radicals.

For example, consider the cubic equation $ax^3 + bx^2 + cx + d \in \mathbb{Q}[x]$. Let

$$C = \sqrt[3]{\frac{2b^3 - 9abc + 27a^2d + \sqrt{(2b^3 - 9abc + 27a^2d)^2 - 4(b^2 - 3ac)^3}}{2}}$$

Then the roots of $f(x)$ are given by

$$x_1 = \frac{-1}{3a}\left(b + C + \frac{b^2 - 3ac}{C}\right)$$
$$x_2 = \frac{-1}{3a}\left(b + \frac{-1+\sqrt{-3}}{2}C + \frac{-1-\sqrt{-3}}{2}\frac{b^2 - 3ac}{C}\right)$$
$$x_3 = \frac{-1}{3a}\left(b + \frac{-1-\sqrt{-3}}{2}C + \frac{-1+\sqrt{-3}}{2}\frac{b^2 - 3ac}{C}\right)$$

Lecture 20: Some Quintics are not Solvable by Radicals

We're ready for the last result of the course, that some quintic (degree five) polynomials in $\mathbb{Q}[x]$ are not solvable by radicals. First, we'll use the concept of *conjugacy classes* to prove that the group S_5 is not solvable. Then, we'll find a quintic polynomial in $\mathbb{Q}[x]$ for which the Galois group of the splitting field over $\mathbb{Q}$ is S_5. Using the theorem we proved on Monday, this will prove that $f(x)$ is not solvable by radicals.

Conjugacy classes, normal subgroups, and S_5

Definition: Let a, b be elements of the group G. We say that a is **conjugate** to b if there is some $g \in G$ such that $gag^{-1} = b$.

Proposition: The "conjugacy" property is an equivalence relation.

Proof: Fill me in!

Reflexivity:

Symmetry:

Transitivity:

An equivalence relation on a set always defines a partition of the set into equivalence classes. In this case, the equivalence class containing a is called the *conjugacy class* of a.

Definition: The **conjugacy class** of $a \in G$ is $cl(a) = \{gag^{-1} \mid g \in G\}$.

Example: What are the conjugacy classes of S_3? Fill me in!

Remarks about conjugacy classes:

1. If a conjugacy class contains just one element a, then $gag^{-1} = a$ for all $g \in G$. In other words, $cl(a) = \{a\}$ if and only if $ga = ag$ for *all* elements of G. Recall that the **center** of a group G, written $Z(G)$, is the subgroup of G consisting of all such elements of G. The identity of a group commutes with all other elements, so $\mathrm{id} \in Z(G)$ and $cl(\mathrm{id}) = \{\mathrm{id}\}$.
2. Conjugacy classes are *not* subgroups! (except, of course, $cl(\mathrm{id}) = \{\mathrm{id}\}$)
3. If H is a normal subgroup of a group G, then for all $g \in G$ and $h \in H$, the normality of H means that $ghg^{-1} \in H$. Therefore, for any $h \in H$, all elements of the conjugacy class $cl(h)$ must *also* be inside H, so H must be a union of conjugacy classes.

Question: What are all normal subgroups of S_3?

Answer: We saw that S_3 has one conjugacy class of size 1, one of size 2, and one of size 3. Any subgroup of G must contain $\{\mathrm{id}\}$ and also must have a number of elements dividing 6. The only possible ways to take unions of conjugacy classes that meets these requirements gives the subgroups $\{\mathrm{id}\}$, A_3, and S_3.

We can use similar logic to find the normal subgroups of A_5. This time, I'll tell you the conjugacy classes.

Proposition: The only normal subgroups of A_5 are $\{\mathrm{id}\}$ and A_5.

Proof: The conjugacy classes of A_5 are the following:

- $\{\mathrm{id}\}$, containing 1 element
- $\{$all permutations of the form $(a\ b)(c\ d)\}$, containing 15 elements.
- $\{$all permutations of the form $(a\ b\ c)\}$, containing 20 elements.
- $\{(1\ 2\ 3\ 4\ 5), (1\ 2\ 4\ 5\ 3), (1\ 2\ 5\ 3\ 4), (1\ 3\ 5\ 4\ 2), (1\ 3\ 2\ 5\ 4), (1\ 3\ 4\ 2\ 5),$
  $(1\ 4\ 3\ 5\ 2), (1\ 4\ 5\ 2\ 3), (1\ 4\ 2\ 3\ 5), (1\ 5\ 4\ 3\ 2), (1\ 5\ 2\ 4\ 3), (1\ 5\ 3\ 2\ 4)\}$
  which has 12 elements.
- $\{(1\ 2\ 3\ 5\ 4), (1\ 2\ 4\ 3\ 5), (1\ 2\ 5\ 4\ 3), (1\ 3\ 4\ 5\ 2), (1\ 3\ 2\ 4\ 5), (1\ 3\ 5\ 2\ 4),$
  $(1\ 4\ 5\ 3\ 2), (1\ 4\ 2\ 5\ 3), (1\ 4\ 3\ 2\ 5), (1\ 5\ 3\ 4\ 2), (1\ 5\ 4\ 2\ 3), (1\ 5\ 2\ 3\ 4)\}$
  which has 12 elements.

The only way to take unions of these conjugacy classes in such a way that $\{\mathrm{id}\}$ is included and the total number of elements divides $|A_5| = 60$ gives the subgroups $\{\mathrm{id}\}$ and A_5.

Corollary: S_5 is not solvable.

Proof: If S_5 were solvable, then A_5 must also be solvable, since subgroups of solvable groups are themselves solvable. But then there would be subgroups

$$\{id\} = H_0 \subseteq H_1 \subseteq H_2 \subseteq \cdots \subseteq H_{n-1} \subseteq H_n = A_5$$

where H_{n-1} is a proper normal subgroup of A_5 and A_5/H_{n-1} is abelian. This is impossible because the only normal subgroups of A_5 are $\{\mathrm{id}\}$ and A_5.

A Galois group isomorphic to S_5

Our goal for today is to find a quintic polynomial in $\mathbb{Q}[x]$ which is not solvable by radicals. We know that if we can find a polynomial for which the Galois group of its splitting field is isomorphic to S_5, we'll be done. But there's a problem. So far in our study of Galois theory, we've considered splitting fields of polynomials that have pretty easy-to-write-down roots, like $\sqrt{3}$ or $\sqrt[3]{2}$ or $i\sqrt[4]{2}$. This makes it relatively easy to analyze the Galois group of the splitting field, because we can explicitly write down the action of the $\mathbb{Q}$-automorphism on the roots to check whether it is an automorphism.[1] When we're looking for a quintic in $\mathbb{Q}[x]$ that *isn't* solvable by radicals, we're guaranteed that our roots won't have such nice expressions! As a consequence, we need to work harder to analyze its Galois group. In particular, we'll need to have a theorem that says "whenever the order of a group is divisible by a prime number p, it has an element of order p".

Definition: The **centralizer** of an element $a \in G$, written $C(a)$, consists of all $g \in G$ such that $gag^{-1} = a$. That is, it consists of all g such that $ga = ag$.

Proposition: The map of sets

$$\begin{aligned} \{\text{left cosets of } C(a)\} &\to \{\text{elements of } cl(a)\} \\ [g] &\mapsto gag^{-1} \end{aligned}$$

is well-defined and is a bijection.

Proof: Fill me in!

Well-defined:

Injective:

Surjective: .

Corollary: For any group G,

$$\begin{aligned} |G| &= \sum_{a \in A} |cl(a)| && \left(\begin{array}{c} A \text{ consists of a single element} \\ \text{from each conjugacy class of } G \end{array}\right) \\ &= |Z(G)| + \sum_{a \in A_2} |cl(a)| && \left(\begin{array}{c} A_2 \text{ consists of a single element} \\ \text{from each conjugacy class of } G \text{ of size} \geq 2 \end{array}\right) \\ &= |Z(G)| + \sum_{a \in A_2} \frac{|G|}{|C(a)|} \end{aligned}$$

Proposition: Let G be a group. If p is a prime number that divides $|G|$, then G has an element of order p.

[1] remember: not *all* permutations of the roots define an element of the Galois group! Sometimes algebraic relations between the roots impose restrictions on what the permutation could be.

Proof: Certainly it's true if $|G| = p$ (the only group whose order is p is $\mathbb{Z}_p$, which has a element of order p). Assume $|G| = n$, and that we have proven the claim is true for all groups of size $\leq n-1$. Consider the formula

$$|G| = |Z(G)| + \sum_{a \in A_2} \frac{|G|}{|C(a)|}$$

and consider the following two cases

p **divides the order of some** $C(a)$**:** Each $C(a)$ is a *proper* subgroup of G (if $C(a)$ were all of G, then the conjugacy class containing a would have just one element). By the inductive hypothesis, $C(a)$ has an element of order p, so therefore G has an element of order p.

p **does not divide the order of any** $C(a)$**:** In this case, p divides each of the numbers $|G|/|C(a)|$, and we already know that p divides $|G|$, so therefore p must also divide $|Z(G)|$. Then $Z(G)$ is an abelian subgroup of G whose order is divisible by p, so by the classification theorem of finite abelian groups, it has an element of order p.

Theorem: The quintic $f(x) = 3x^5 - 15x + 5$ is not solvable by radicals over $\mathbb{Q}$.

Proof: Let E be the splitting field of $f(x)$ over $\mathbb{Q}$. It suffices to prove that $\text{Gal}(E/\mathbb{Q}) \cong S_5$. The graph of $f(x)$ shows that there are five roots: three real roots and two complex roots which are complex conjugates of one another. We know that $\text{Gal}(E/\mathbb{Q})$ is a subgroup of S_5 where the element of S_5 corresponding to a $\mathbb{Q}$-automorphism is described by the permutation it induces on the five roots of $f(x)$.

The complex conjugation map is an element of $\text{Gal}(E/\mathbb{Q})$ which corresponds, under the identification $\text{Gal}(E/\mathbb{Q}) \cong S_5$, to a cycle of length 2 (since it fixes three elements and permutes the other two).

Let α be any root of $f(x)$. Because $f(x)$ is irreducible (by Eisenstein), $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 5$, so by the fundamental theorem of Galois theory, $|\text{Gal}(E/\mathbb{Q})|$ is divisible by 5. Therefore, $\text{Gal}(E/\mathbb{Q})$ has an element of order 5, which must correspond under the identification $\text{Gal}(E/\mathbb{Q}) \cong S_5$ to a cycle of length 5 (cycles of length 5 are the only elements of order 5 in S_5).

If a subgroup of S_5 contains a 5-cycle and also a 2-cycle, then it must be the entire group S_5.

Therefore, $\text{Gal}(E/\mathbb{Q}) \cong S_5$ is not solvable, hence $f(x)$ is not solvable by radicals over $\mathbb{Q}$.