250A: Groups, Rings, and Fields
With Worked Examples

Nir Elber

Fall 2021

Contents

How strange to actually have to see the path of your journey in order to make it.

—Neal Shusterman, [Shu16]

Theme 1
Group Grind

Groups, as men, will be known by their actions.

—Guillermo Moreno

1.1 August 26

1.1.1 Logistics

Course website is bcourses.berkeley.edu/courses/1504926.

1.1.2 Group Talk

Recall the definition.

Definition 1.1 (Group, concrete)**.** A group G is the set of symmetries of something.

"Something" here is quite vague, but it's generally something like a graph or vector space or group.

Also, "symmetry" is quite vague, but it's somewhat intuitive: we are more or less asking for structure-preserving mappings. Namely, our "something" is a set X, we are asking for our group G to be a structure-preserving maps in $\mathrm{Sym}(X)$. In practice, what "structure-preserving" means is clear.

There is also an abstract definition of groups.

Definition 1.2 (Group, abstract)**.** A group is a set G with an operation $* : G \times G \to G$ which satisfies the following properties for any $a, b, c \in G$.

- Associative: $a * (b * c) = (a * b) * c$.
- Identity: there is an identity $e \in G$ such that $a * e = e * a = a$.
- Inverses: there is an a^{-1} such that $a * a^{-1} = a^{-1} * a = e$.

We might ask how the abstract and concrete definitions interplay.

For example, suppose we have a concrete group G. Then we can recover our abstract group by having the binary operation be composition. Association holds because function application is associative; id is our identity; inverses exists because symmetries are bijective.

In the other direction, it's less obvious how we take an abstract group to a concrete one.

Question 1.3 (Cayley)**.** Given an abstract group G, can we realize G as the symmetries of some object X?

To make this rigorous, we should talk about group actions so that "some object" can be rigorized.

Definition 1.4 (Group action)**.** Fix G a(n abstract) group and S a set. We say that G (left) acts on a set S if we have a map $\cdot : G \times S \to S$ which satisfies the following.

- Identity: $es = s$ for any $s \in S$.
- Associativity: $(gh) \cdot s = g \cdot (h \cdot s)$ for any $g, h \in G$.

Above we have technically defined a left group action; right group actions are defined analogously.

To answer Question 1.3, we let G act on $S := G$ (as the set) where the group action is defined by left multiplication. This is indeed an action, where the identity and associative laws follow from the corresponding laws in a group. This implies that we have a map

$$G \to \mathrm{Sym}(G).$$

In fact, this is injective because its kernel is trivial: the only map taking $e \mapsto e$ is e itself by the identity law.

How can we restrict $\mathrm{Sym}(S)$ so that this is injection is also surjective? To add extra structure to S, we equip S with a right G-action.[1] The key observation, now, is that the left action on S by G preserves the right action. Namely, if we have $g_\ell, g_r \in G$ and $s \in G$, then

$$(g_\ell s)g_r = g_\ell(sg_r)$$

by associativity. In other words, we can multiply on the left or right in any order.

Warning 1.5. We do not need to have that the left action preserves the left action. Namely, this is asking for $g \cdot (hs) = h \cdot (gs)$, which need not be true for non-abelian groups.

So in fact we have the restriction that

$$G \hookrightarrow \mathrm{Sym}_{G\text{-right}}(S).$$

We claim this is surjective. Indeed, suppose $\sigma : S \to S$ is a bijection such that $\sigma(sg) = (\sigma s)g$ for any $g \in G$. Then we claim σ is multiplication by $\sigma e \in G$. Indeed,

$$\sigma s = \sigma(es) = (\sigma e)s,$$

which is what we wanted. So we have the following.

Theorem 1.6 (Cayley)**.** Any abstract group G is the group of symmetries of some mathematical object.

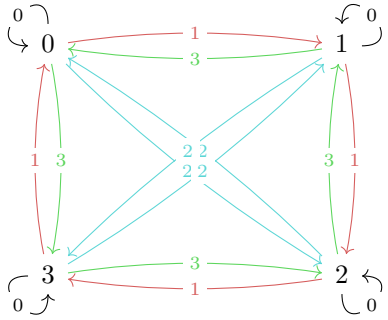
1.1.3 Representation Talk

Perhaps we would like a more natural object than the group acting on literally itself. It turns out that we can also realize any group G as the symmetry group of a graph. We'll do this for finite groups.

[1] For these keeping score, we now have three copies of G: we have $S = G$, as well as actions of G on S on the left and right.

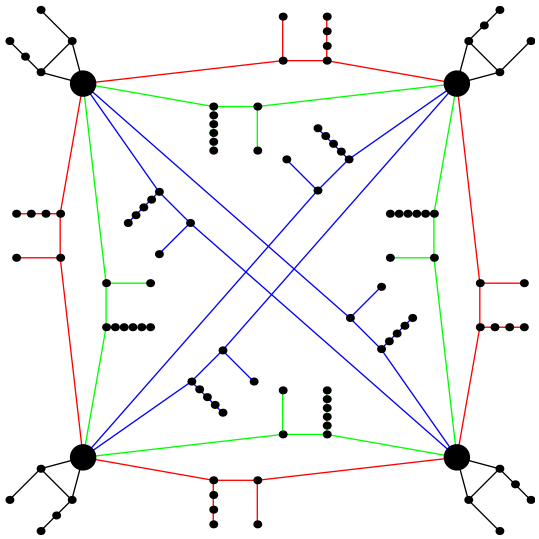
Theorem 1.7. Any finite group G is the symmetry group of some finite graph.

Proof. We again do a little bit of cheating. We set our graph S to have vertices labeled by G. Next we color the edges of the graph according to the group action. Here is an example graph for $\mathbb{Z}/4\mathbb{Z}$.



So we have a colored graph, and we can check that the only symmetries of this graph corresponds to the action of G itself: once we decide which vertex we should take 0 to, the preservation of each colored arrow forces the other vertices. (For example, if we take 0 to 2, there is only one red edge which was going out from 0, and there is only one edge currently going out from 2, so we have to send 1 to 3 as well. The other vertices are similar.)

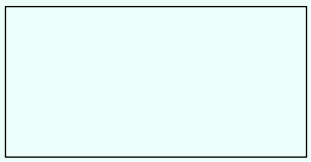
Now we would like to get rid of the colors and directions of the graph. For example, we might take a directed edge and add markers along the edge to ensure the symmetries are well-defined. For example, here is one way we could add markers to the graph of $\mathbb{Z}/4\mathbb{Z}$.



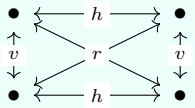
Here the colors are added for clarity though the graph is actually uncolored. The point is that we should need to send the big vertices to other big vertices and fake "colored and directed" edges to ones that match. ■

These are called Cayley graphs. Let's do some examples now.

Example 1.8. Here is a rectangle.



Our symmetries are do-nothing e, flip vertically v, flip horizontally h, and rotate 180 degrees r. Our Cayley graph looks like this.



To round things out, we note that group theory is roughly about the following.

- We want to classify all groups. For example, what kinds of groups act on crystals?
- Given a group G, we want to see what are the interesting things that groups act on. In general, these are permutation representations, but we are often just interested in linear representations acting on vector spaces. For example, there is some story here in physics.

  (We can also represent groups via their multiplication table. Professor Borcherds does not like these.)

1.1.4 Maps of Groups

Here is our motivating question.

Question 1.9. When are two groups the same?

For example, we might have G_1 be the symmetry group of a rectangle and G_2 the set of elements $\langle a, b : a^2 = b^2 = ab^{-1}b^{-1}\rangle$, then these are in fact the same: name a the horizontal flip and b the vertical flip. Then we see we have a bijective, structure-preserving map from $G_1 \to G_2$.

Definition 1.10 (Homomorphism)**.** A map of groups $\varphi : G_1 \to G_2$ is a *homomorphism* if and only if

$$\varphi(gh) = \varphi(g)\varphi(h)$$

for any $g, h \in G_1$. We can check this implies $\varphi(e_1) = e_2$ and $\varphi(g^{-1}) = \varphi(g)^{-1}$.

Then an isomorphism is a bijective homomorphism.

Definition 1.11. An map of groups $\varphi : G_1 \to G_2$ is an isomorphism if and only if it is a bijective homomorphism.

Let's give some examples.

Example 1.12. Consider $\exp : (\mathbb{R}, +) \to (\mathbb{R}^\times, \times)$. This is a homomorphism because $\exp(a+b) = \exp(a) \times \exp(b)$. However, this is not an isomorphism because it does not hit negative elements.

Example 1.13. Fix $G = \mathbb{Z}/4\mathbb{Z}$ and $H = (\mathbb{Z}/5\mathbb{Z})^\times$. Then we have the isomorphism by sending

$$1 \in G \longmapsto 2 \in H.$$

In other words, $\varphi(k) = 2^k$. We can check this really is a bijection.

1.1.5 Lagrange's Theorem

Let's list all groups.

1. There is only one group of order 1: it has to be trivial.
2. There is only one group of order 2: we need an identity and a non-identity element, which has to square to the identity.
3. For order 3, we introduce Lagrange's theorem.

Theorem 1.14 (Lagrange)**.** The order of an element g of a group G divides $\#G$.

We won't prove this yet.

3. Now, with Lagrange's theorem, we note that a non-identity element needs to have order bigger than one but dividing 3 and so must be three. So there is an element of order 3, so it must be cyclic.

We remark that this same argument gives the following.

Proposition 1.15. Suppose G is a group of prime order. Then $G \cong \mathbb{Z}/p\mathbb{Z}$.

Let's prove Lagrange's theorem then.

Proof of Theorem 1.14. The point is that the order of g is the size of the subgroup $\langle g \rangle$. So we show the more general statement as follows.

Theorem 1.16 (Lagrange, II)**.** Fix $H \subseteq G$ a subgroup of a group. Then $\#H \mid \#G$.

Proof. We need to study the geometric meaning of a subgroup H. Well, suppose G is the group of symmetries of some object S and pick up a point $p \in S$. Then we could set $\mathrm{Stab}(s)$ to be the set of elements fixing $s \in S$. For example, for an icosahedron, there is a $\mathbb{Z}/5\mathbb{Z}$ fixing a vertex, there is a $\mathbb{Z}/3\mathbb{Z}$ fixing a face, and so on.

So we can realize subgroups as stabilizers of subsets. We would like the converse: given a subgroup H, we would like a set S with a G-action such that H is the stabilizer of some subset of S.

To make the problem easier, suppose that the G-action is transitive so that it lives in one orbit. Namely, fixing $s_0 \in S$, we have a function $G \to S$ by

$$g \longmapsto gs_0.$$

We would like for $gs_0 = s_0$ to be equivalent to $g \in H$ for our particular subgroup H. Quickly, we note that $gs_0 = g's_0$ if and only if $g(g')^{-1}s_0 = s_0$ if and only if $g(g')^{-1} \in H$ if and only if $g \in g'H$ if and only if $gH = g'H$.

This suggests a construction of our set S as G/H, the set of *cosets* $\{gH : g \in G\}$, or the equivalence classes as given above. We do have to check that $g \in g'H$ is an equivalence relation (say $\sim$), however. We will not be detailed about this.

- Note $g \sim g$ because $e \in H$.
- Note $g \sim g'$ implies $g' \sim g$ because H has inverses.
- Note $g \sim g'$ and $g' \sim g''$ implies $g \sim g''$ because H is associative.

Remark 1.17. If we work with monoids, this is no longer an equivalence relation because of the lack of inverses.

In fact, we can check that equivalence classes have the same size: if we have two cosets g_1H and g_2H, then we have a bijection $g_1H \to g_2H$ by $g_1h \mapsto g_2g_1^{-1} \cdot g_1h$. (We will not check that this is bijective here, but it is at least injective, and it has an inverse, so it is.)

So we have G act on G/H by left multiplication. Any two of these equivalence classes have the same size, so they all have size $\#(eH) = \#H$, so we see that the order of G is $\#H$ times the number of classes $\#(G/H) =: [G : H]$. So indeed, $\#H \mid \#G$. ■

This completes the proof. ■

We remark that we also have the following.

Proposition 1.18. If G acts transitively on a set S, then we see $\#S = \#G/\#\operatorname{Stab}(s_0)$ for any chosen $s_0 \in S$,

Proof. This follows from the above proof: consider the (surjective) map $G \twoheadrightarrow S$ defined by $g \mapsto gs_0$. This is actually defined up to coset of $\operatorname{Stab}(s_0)$ because we have that $g_1s_0 = g_2s_0$ if and only if $g_2^{-1}g_1 \in \operatorname{Stab}(s_0)$ if and only if $g_1 \operatorname{Stab}(s_0) = g_2 \operatorname{Stab}(s_0)$. So we actually have a bijection $G/\operatorname{Stab}(s_0) \twoheadrightarrow S$, which is the result. ■

Example 1.19. How many rotations of an icosahedron are there? Well, take H to be the subgroup fixing a vertex. By spinning along a vertex, there are 5 such rotations fixing the subgroup, and there are 12 total vertices, so there are 60 total rotations here.

Let's see some other applications of Lagrange's theorem.

Proposition 1.20 (Fermat's little)**.** Fix $x \in (\mathbb{Z}/p\mathbb{Z})^\times$. Then $x^{p-1} \equiv 1$.

Proof. Well, the order of $(\mathbb{Z}/p\mathbb{Z})^\times$ is $p-1$, so the order of x divides $p-1$, from which the result follows. ■

More generally, we have the following.

Proposition 1.21 (Euler's totient)**.** Fix $x \in (\mathbb{Z}/m\mathbb{Z})^\times$. Then $x^{\varphi(m)} \equiv 1$.

Proof. The point is that the order of x has to divide the order of $\#(\mathbb{Z}/m\mathbb{Z})^\times = \varphi(m)$. So the result follows. ■

1.2 August 31

1.2.1 Groups of Order Four

Let's continue our list of groups. Let's work with groups G of orders 4. All elements must have order dividing 4.

- If there's an element of order 4, we are cyclic.

- Otherwise, all non-identity elements have order 2. Note that we know this group is abelian already! Indeed, for $a, b \in G$, we see $(ab)^2 = e$ implies that $abab = e$, so

$$ab = aababb = ba.$$

  Note that this is very special for 4; it is not the case that if all groups have order dividing 3.

  Well, now that we are abelian, we see G is a vector space over $\mathbb{F}_2$, which we can check by hand, and size reasons force us to have $G \cong (\mathbb{F}_2)^2$ by choosing a suitable basis.

So we have the following.

Theorem 1.22. We have exactly two groups of order 4, up to isomorphism.

Proof. Above we showed that all groups of order 4 are isomorphic to either $\mathbb{Z}/4\mathbb{Z}$ or $(\mathbb{Z}/2\mathbb{Z})^2$. Note that these are different because $\mathbb{Z}/4\mathbb{Z}$ has an element of order 4, though $(\mathbb{Z}/2\mathbb{Z})^2$ does not. ■

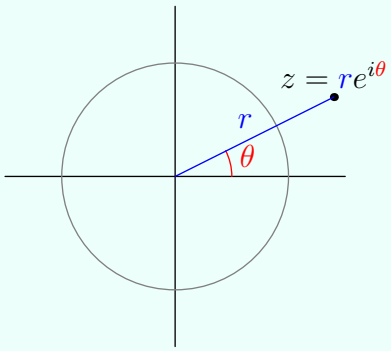
1.2.2 Product Groups

We remark that $\mathbb{F}_2^2$ is an example of a product.

Definition 1.23 (Product groups)**.** Given two groups G, H we can define the *product* group $G \times H$ of pairs (g, h) where $g \in G$ and $h \in H$. Here, multiplication is defined componentwise.

Example 1.24. For any field k, we have that k^n is a product group, for any positive integer n.

Example 1.25. We have that $\mathbb{C}^\times \cong \mathbb{R}_{>0} \times S^1$. This is merely saying that we can represent nonzero complex numbers uniquely by $z = re^{i\theta} \mapsto (r, \theta)$. Here is the image.



Example 1.26. We have that $\mathbb{Z}/6\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$, which is an instance of the Chinese remainder theorem.

We can generalize the previous example.

Proposition 1.27. More generally, we have that $\mathbb{Z}/mn\mathbb{Z} \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ when $\gcd(m, n) = 1$.

Remark 1.28. This does not hold for $m = n = 2$.

Proof of Proposition 1.27. This follows from the mapping

$$\mathbb{Z}/mn\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$$

by taking $[k]_{mn} \to ([k]_m, [k]_n)$. We can check that this is homomorphic by hand. This is injective because if $k \equiv 0 \pmod{m}$ and $k \equiv 0 \pmod{n}$, then $m, n \mid k$, so $mn \mid k$ because $\gcd(m, n) = 1$, so $k \equiv 0 \pmod{mn}$. Then this map is surjective for size reasons, giving our isomorphism. ■

Example 1.29. Consider the group of rotations of the various platonic solids. They have orders as follows.

- Tetrahedron: 12.
- Cube: 24.
- Octahedron: 24.
- Icosahedron: 60.
- Dodecahedron: 60.

If we add in reflections, the number of these objects doubles, and in fact the bottom four are a product of $\mathbb{Z}/2\mathbb{Z}$ times the group of rotations. As for why, the added $\mathbb{Z}/2\mathbb{Z}$ comes from the reflection which inverts the entire figure, sending a vertex to its opposite. (This inversion is not a rotation because it has determinant -1, when thought of as a matrix over $\mathbb{R}^3$.)

Example 1.30. Consider the set of all roots of unity in $\mathbb{C}$. These can be written explicitly as

$$U_\infty = \left\{e^{2i\pi q} : q \in \mathbb{Q}\right\}.$$

We can decompose this into

$$U_\infty \cong \{z : z \text{ has order a power of } 2\} \times \{z : z \text{ has odd order}\}.$$

Note that we can also take infinite products of groups, but sometimes that's too strong.

Definition 1.31 (Sum)**.** Given an infinite collection of groups $\{G_\alpha\}_{\alpha\in\lambda}$, we define the sum group

$$\bigoplus_{\alpha\in\lambda} G_\alpha = \left\{\{g_\alpha\}_{\alpha\in\lambda} \in \prod_{\alpha\in\lambda} G_\alpha : g_\alpha = e_\alpha \text{ with finitely many exceptions}\right\}.$$

This is a subgroup of the big product group.

Example 1.32. We can check that

$$U_\infty = \bigoplus_{p \text{ prime}} \{z : z \text{ has order a power of } p\}.$$

This proof essentially boils down to the Chinese remainder theorem.

Example 1.33. By unique prime factorization, we see that

$$\mathbb{Q}^\times \cong \{\pm 1\} \oplus \bigoplus_{p \text{ prime}} \langle p \rangle,$$

where $\langle p \rangle = p^{\mathbb{Z}}$ consists of the powers of p.

1.2.3 Groups of Orders Five and Six

For groups of order 5, they are cyclic. Here is an exercise, for fun.

Question 1.34. Find a graph whose automorphism groups is $\mathbb{Z}/5\mathbb{Z}$.

For groups of order 6, we note that we we have two obvious groups already:

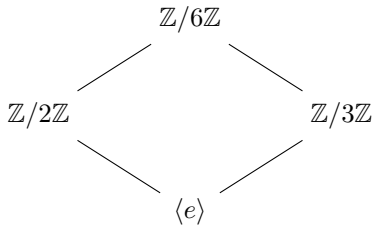
- We have $\mathbb{Z}/6\mathbb{Z}$, which is $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$.
- We have S_3, the permutation group on three letters.

Remark 1.35. Additionally, we see that S_3 is our first example of a nonabelian group! We see that

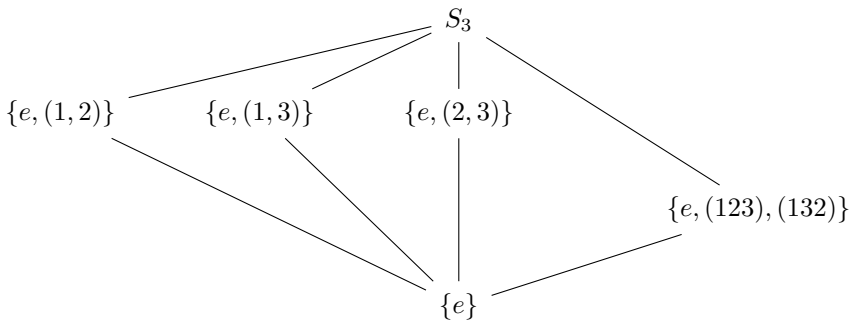
$$(12)(23) = (123) \qquad \text{but} \qquad (23)(12) = (132).$$

So this also shows that S_3 is not abelian.

These are not isomorphic because S_6 is not cyclic. Alternatively, we can draw out our subgroup chart; here is the chart for $\mathbb{Z}/6\mathbb{Z}$.



And we could write out the subgroup table of S_3, and find that there are lots of subgroups of order 2.



In fact, the subgroup table of S_3 has "lots" of subgroups of order 2. What's going on? These are an instance of "non-normal subgroups."

1.2.4 Normal Subgroups and Quotients

Our motivation here is the following question.

Question 1.36. Given groups $H \subseteq G$, can we define a group G/H? more precisely? Can we have a (surjective) homomorphism $\varphi : G \to G/H$ with kernel exactly H?

We can write this as a short exact sequence

$$1 \to H \to G \to G/H \to 1.$$

In general, we can define an exact sequence.

Definition 1.37 (Exactness)**.** Given a sequence of maps

$$A \to B \to C,$$

we say that this is *exact at* B if the image of $A \to B$ is the kernel of $B \to C$.

This lets us define short exact.

Definition 1.38 (Short exact sequence)**.** We define a short exact sequence as an exact sequence of the form

$$1 \to A \to B \to C \to 1.$$

Namely, $A \to B$ is injective, $B \to C$ is surjective, and the image of $A \to B$ is the kernel of $B \to C$.

Anyways, let's return to talking about our question. We are hoping that we have a well-defined map. So suppose that g_1 and g_2 have the same image in G/H: this is equivalent to $g_1 h = g_2$ for some $h \in H$ by rearranging $\varphi(g_1) = \varphi(g_2)$. So we define left cosets.

Definition 1.39 (Left cosetes)**.** We define G/H as the set of *left cosets* $\{gH : g \in G\}$. Note that we are not claiming this is a group in general.

We hope that our group law is

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

However, this might not be well-defined! The issue is that, for any $h \in H$, we also need

$$g_1 h H \cdot g_2 H = (g_1 h g_2) H.$$

Note that this is free for abelian groups, for $hg_2 = g_2 h$, so we can move the h over. However, we can weaken this condition to merely requiring $hg_2 = g_2 h'$ for some h', which is equivalent to $g_2 h g_2^{-1} \in H$ for each $h \in H$.

Definition 1.40 (Normal)**.** We say that a subgroup $H \subseteq G$ is normal if, for each $g \in G$, we have that $gHg^{-1} = H$. (Actually, $gHg^{-1} \subseteq H$ is good enough here.)

Proposition 1.41. Fix H a normal subgroup of G. Then G/H is a group.

Proof. The main check is that G/H has well-defined multiplication. Indeed, if $g_1 H = g_1' H$ and $g_2 H = g_2' H$, then $g_1 = g_1' h_1$ and $g_2 = g_2' h_2$ for some $H_1, h_2 \in H$ so that

$$g_1 H \cdot g_2 H = (g_1 g_2) H = (g_1' h_1 g_2' h_2) H = g_1' H \cdot g_2' \underbrace{(g_2')^{-1} h_1 g_2' h_2}_{\in H} H = g_1' H \cdot g_2' H.$$

From here, checking that G/H is actually a group is inherited more or less directly from G because $G \to G/H$ is homomorphic and surjective. ∎

Example 1.42. The subgroup $\{e, (123), (132)\} \subseteq S_3$ is normal. For example, for any $\sigma \in S_3$, we can check that

$$\sigma(123)\sigma^{-1} = (\sigma 1, \sigma 2, \sigma 3\} \in \{e, (123), (132)\}$$

because σ is a permutation. This normal subgroup gives us the exact sequence

$$1 \to \mathbb{Z}/3\mathbb{Z} \to S_3 \to \mathbb{Z}/2\mathbb{Z} \to 1$$

because the quotient $S_3/\{e, (123)(132)\}$ has order 2 and must be $\mathbb{Z}/2\mathbb{Z}$.

Non-Example 1.43. The subgroup $\{e, (12)\} \subseteq S_3$ is *not* a normal subgroup. Indeed, we can just check that

$$(13)(12)(13) = (23) \notin \{e, (12)\}.$$

However, we can check that conjugating $H = \{e, (12)\}$ by $g \in (23)$, we have that H is conjugate to $gHg^{-1} = \{e, (23)\}$.

As a side remark, we note that the left cosets equal the right ones for normal subgroups: any coset gH can be written as a right coset by writing it as $gH = gHg^{-1}g = Hg$ by normality.

However, for non-normal subgroups, there are dangers.

Example 1.44. Again take $H \coloneqq \{e, (12)\} \subseteq S_3$. Our left cosets are

$$\begin{cases} H = \{e, (12)\}, \\ (123)H = \{(123), (13)\}, \\ (132)H = \{(132), (23)\}. \end{cases}$$

However, our right cosets are

$$\begin{cases} H = \{e, (12)\}, \\ H(123) = \{(123), (23)\}, \\ H(132) = \{(132), (13)\}. \end{cases}$$

1.2.5 Cauchy's Theorem

Let's use this an excuse to introduce some theorems. Here is a motivating question.

Question 1.45. Suppose that $d \mid \#G$ for a group G. Is there an element of order d?

Well, of course not: $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ has order 4 but does not have an element of order 4. However, we have the following.

Theorem 1.46 (Cauchy)**.** Suppose that p is a prime dividing the order of a group G. Then there is an element of order p.

Proof. We do casework on if G is abelian.

Remark 1.47. Trying to prove something for groups G by doing casework on G abelian vs. G nonabelian is like trying to prove something for objects O in the universe by doing casework on if O is a banana or O is not a banana. But here we go.

- If G is abelian, we start by picking up $a \in G \setminus \{e\}$. (If $G = \{e\}$, there is nothing to show.) Then we can raise a to a power to kill all the primes in its order except for, say, q. If $p = q$, then we are done.

  Otherwise, we can look at $G/\langle a \rangle$, where this quotient is good because our groups is abelian. Then this has order $\#G/q$, which is still divisible by p because $q \neq p$. So induction can give us a coset $b\langle a \rangle \in G/\langle a \rangle$ of order p.

  However, $b^p \in \langle a \rangle$ is either the identity or some generic element of $\langle a \rangle$, but certainly $b^{pq} = e$. The order cannot be 1 ($b\langle a \rangle \in G/\langle a \rangle$ has order p), nor can it be q (this would force $b^q \in \langle a \rangle$, but $p \nmid q$), so the order is either p or pq. If p, we are done; if pq, then b^q has order p.

- If G is nonabelian, then we again have two cases: if G has a proper subgroup of order divisible by p, then we can do induction to finish. Otherwise, all proper subgroups have order not divisible by p with index $[G : H]$ always divisible by p.

  Now the trick is to look at the action of G on G by conjugation, and split up the action into orbits, which are conjugacy classes. Explicitly,

  $$Gg = \left\{ghg^{-1} : h \in G\right\}.$$

  Now we check that the size of any orbit Gg is $\#G/\#\operatorname{Stab}(g)$ by the Orbit-stabilizer theorem. But this is always divisible by p, except when $\#\operatorname{Stab}(g) = G$ because I said so.

  To finish, we do the class equation by hand. We see that

  $$G = \bigcup_{Gg} Gg \tag{$*$}$$

  because we are partitioning by the action. The left-hand side has size divisible by p, and the right-hand orbits are all divisible by p except for elements $h \in G$ such that $ghg^{-1} = h$ for all $g \in G$. This gives us the following definition.

Definition 1.48 (Center)**.** For g a group, we define $Z(G) = \{g \in G : ghg^{-1} = h\}$. In other words, $gh = hg$ for each $g \in Z(G)$ and $h \in G$, so $Z(G)$ commutes with everyone.

Finishing up the proof, we see that $(*)$ reads as

$$\#G = Z(G) + \sum_{\substack{Gg \\ \#Gg > 1}} \#Gg$$

after taking sizes, and everything here is divisible by p except $Z(G)$, requiring that $Z(G)$ has size divisible by p. But now $Z(G)$ is an abelian subgroup (everything commutes by definition), so it has an element of order p, finishing.

Alternatively, $Z(G)$ is a proper subgroup (proper because G is nonabelian) with order divisible by p, which is a contradiction to our assumption that G has no proper subgroups with order divisible by p. ∎

Remark 1.49. The above argument actually shows that if all proper subgroups have index divisible by p, then $Z(G)$ is divisible by p.

Remark 1.50. There are many ways for a group G to act on itself.

- There is a left action, by $g \cdot h = gh$..
- There is a trivial action: $g \cdot h = h$.
- There is a right action: $g \cdot h = hg^{-1}$. (Note the inverse is required for associativity reasons.)
- There is the conjugacy action: $g \cdot h = ghg^{-1}$.

Then there are the corresponding right actions.

Let's use this to classify groups of order 6.

Proposition 1.51. There are only two non-isomorphic groups of order 6, which are $\mathbb{Z}/6\mathbb{Z}$ and S_3.

Proof. Fix G of order 6. Then we are promised an element a of order 3 and an element b of order 2. Well, we claim that $\langle a \rangle$ is normal. More generally, we have the following.

Lemma 1.52. Fix $H \subseteq G$ a subgroup of index 2. Then H is normal.

Proof. Indeed, for any $g \in H$, we see that $gHg^{-1} = H$ for free. Otherwise, when $g \in G \setminus H$, we have that gH and Hg must both be disjoint from H while having size H (recall $[G : H] = 2$), so $gH = Hg = G \setminus H$. In particular, $gH = Hg$ implies $gHg^{-1} = H$ still. ∎

Thus, we have a short exact sequence

$$1 \to \underbrace{\mathbb{Z}/3\mathbb{Z}}_{\langle a \rangle} \to G \to \mathbb{Z}/2\mathbb{Z} \to 1.$$

Remark 1.53. Filling in the middle here need not be unique, even in basic cases. For example, we have a short exact sequence

$$1 \to \mathbb{Z}/2\mathbb{Z} \to G \to \mathbb{Z}/2\mathbb{Z} \to 1$$

where $G = (\mathbb{Z}/2\mathbb{Z})^2$ or $\mathbb{Z}/4\mathbb{Z}$.

Regardless, we simply do this by hand. We have the following definition.

Definition 1.54 (Split short exact sequence)**.** The short exact sequence

$$1 \to A \to B \to C \to 1$$

splits if B has a subgroup C_B isomorphic to C lifting $B \to C$.

In particular, we see that

$$1 \to \underbrace{\mathbb{Z}/3\mathbb{Z}}_{\langle a \rangle} \to G \to \mathbb{Z}/2\mathbb{Z} \to 1.$$

splits because G does have a subgroup $\langle b \rangle$ isomorphic to $\mathbb{Z}/2\mathbb{Z}$. The point is that $\langle b \rangle$ acts on $\langle a \rangle$ by conjugation because $\langle a \rangle$ is normal (this is the restriction of $G \to \mathrm{Aut}(\langle a \rangle)$ to from G to $\langle b \rangle$). So we have induced an action of $\mathbb{Z}/2\mathbb{Z}$ on $\mathbb{Z}/3\mathbb{Z}$, but we only have a few automorphisms of $\mathbb{Z}/3\mathbb{Z}$, so we are forced to have one of

$$\begin{cases} bab^{-1} = a, \\ bab^{-1} = a^2. \end{cases}$$

So we have the group presentations

$$\begin{cases} G = \langle a^3 = 1, b^2 = 1, bab^{-1} = a \rangle \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \\ G = \langle a^3 = 1, b^2 = 1, bab^{-1} = a^2 \rangle \cong S_3. \end{cases}$$

The last group is isomorphic to S_3 by taking $a = (123)$ and $b = (12)$, say. ■

1.2.6 Semidirect Products

What's happening with our split short exact sequences is semidirect products.

Definition 1.55 (Semidirect products)**.** Suppose that A and C are groups such that A has a C-action. (In other words, there is a homomorphism $C \to \mathrm{Aut}(A)$.) Then we define G as the *semidirect product* if we can form the short exact sequence

$$1 \to A \to G \to C \to 1$$

such that G has (isomorphic copies of) A as a normal subgroup and C as another subgroup.

We should actually exhibit our semidirect product. We have the following.

Proposition 1.56. Fix A and C as above. We define the semidirect product $G = A \times C$ as a set, with multiplication defined by

$$(a_1, c_1)(a_2, c_2) = (a_1(c_1 \cdot a_2), c_1c_2),$$

well $c_1 \cdot a_2$ refers to the C-action on A.

Remark 1.57. Let's try to motivate this multiplication. Informally, we want the action of C on A to be conjugation so that A stands a pretty good chance of being normal, and we want to be able to think of pairs (a, c) as actual products ac. These forces combine to let us write

$$\begin{aligned} (a_1, c_1)(a_2, c_2) &= a_1c_1a_2c_2 \\ &= a_1c_1a_2 \left(c_1^{-1}c_1\right) c_2 \\ &= a_1 \left(c_1a_2c_1^{-1}\right) (c_1c_2) \\ &= (a_1(c_1 \cdot a_2), c_1c_2)\,. \end{aligned}$$

Proof of Proposition 1.56. We have to check that this is a group, which can be checked by force. We run down the properties because some of this more subtle than it appears.

- Associativity in the second coordinate is inherited from C. Associativity in the first coordinate comes from writing

  $$\big((a_1, c_1)(a_2, c_2)\big)(a_3, c_3) = (a_1(c_1 \cdot a_2), c_1c_2)(a_3, c_3) = (a_1(c_1 \cdot a_2)(c_1c_2 \cdot a_3), \bullet),$$

  and comparing it with

  $$(a_1, c_1)\big((a_2, c_2)(a_3, c_3)\big) = (a_1, c_1)(a_2(c_2 \cdot a_3), \bullet) = (a_1c_1 \cdot (a_2(c_2 \cdot a_3)), \bullet).$$

  These are equal because our C-action is inducing a homomorphism $C \to \mathrm{Aut}(A)$.
- Our identity element is (e, e).
- Our inverse element is $(a, c)^{-1} = \left(c^{-1} \cdot a^{-1}, c^{-1}\right)$. On one side,

  $$(a, c)\left(c^{-1} \cdot a^{-1}, c^{-1}\right) = \left(a\left(c \cdot c^{-1}a^{-1}\right), e\right) = (e, e).$$

On the other side,

$$\left(c^{-1}\cdot a^{-1}, c^{-1}\right)(a,c) = \left(\left(c^{-1}\cdot a^{-1}\right)\left(c^{-1}a\right), e\right) = \left(c^{-1}\cdot\left(a^{-1}a\right), e\right) = (e,e),$$

where we again used that the C-action is inducing a homomorphism $C \to \mathrm{Aut}(A)$.

Now we will check that the short exact sequence

$$1 \to A \to G \to C \to 1$$

splits, as well as that A is normal in G. We have the following to check.

- Exact at A: the map $A \to G$ is injective, defined by $a \mapsto (a, e)$. It's not hard to see that this is homomorphic.
- Exact at C: the map $G \to C$ is surjective, defined by $(a, c) \mapsto c$. This is homomorphic because the second coordinate of $A \times C$ is merely multiplication.
- Exact at G: the map $A \to G$ surjects onto points of the form $\{(a, e) : a \in A\}$, and the kernel of $G \to C$ is exactly the points such that $(a, c) \mapsto c = e$, which is again $\{(a, e) : a \in A\}$. So $\mathrm{im}(A \to G) = \ker(G \to C)$.
- We split: The subgroup $\{(e, c) : c \in C\}$ is isomorphic to C and lifts our $G \to C$ projection, so the given short exact sequence splits.
- A is normal: We need to show that $A_G \coloneqq \{(a, e) : a \in A\}$ is normal in G. It is enough to note that, for any $(a_0, e) \in A_G$ and $(a, c) \in G$, we have

$$\begin{aligned}(a,c)(a_0,e)(a,c)^{-1} &= (a,c)(a_0,e)\left(c^{-1}a^{-1}, c^{-1}\right)\\ &= (\text{garbage}, c)\left(c^{-1}a^{-1}, c^{-1}\right)\\ &= \left(\text{more garbage}, cc^{-1}\right)\\ &= (\text{more garbage}, e).\end{aligned}$$

■

Example 1.58. We have that S_3 is the semidirect product of $\mathbb{Z}/3\mathbb{Z}$ by $\mathbb{Z}/2\mathbb{Z}$, notated $\mathbb{Z}/\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$. Notice that the construction of "semidirect" takes more data than is provided by $\mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z}$: we also need to know the action.

Let's do some more examples.

Example 1.59. Take the set of all linear functions $x \mapsto ax + b$, where our multiplication is composition. We can check that we have a normal subgroup $x \mapsto x+b$, and its quotient group is isomorphic to $x \mapsto ax$.

Example 1.60. The Poincaré group consists of the automorphisms of space-time. It has a normal subgroup consisting of translations through space-time, and the quotient is the "Lorentz group" all rotations of space-time which preserve the metric $t^2 - x^2 - y^2 - z^2 = 0$.

1.3 September 2

Why do I hear boss music?

1.3.1 Groups of Order 8

Last time we classified all groups of order 6. Note that groups of order 7 are cyclic because 7 is prime.

So let's look at order 8. Fix G a group of order 8. Note that our orders are all in $\{1, 2, 4, 8\}$. If there's an element of order 8, are cyclic, so we may ignore this order. So we have two possibilities.

- If all elements have order 2, then we see that all elements commute (again, $abab = e$ implies $ab = ba$), so G is a vector space over $\mathbb{F}_2$, so we are $G \cong \mathbb{F}_2^3 \cong (\mathbb{Z}/2\mathbb{Z})^3$ by size reasons.
- Otherwise there is at least one element of order 4. Calling this element $a \in G$, then we have an order-4 subgroup $\langle a \rangle$, which is index 2 and hence normal. So, as usual, we get a short exact sequence

$$1 \to \underbrace{\mathbb{Z}/4\mathbb{Z}}_{\langle a \rangle} \to G \to \mathbb{Z}/2\mathbb{Z} \to 1.$$

  So we have another extension problem to fill in G. Some possibilities for G include $\mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ or $\mathbb{Z}/8\mathbb{Z}$ (even though we don't care about this case currently), but perhaps there are others.

The point of our short exact sequence is that we have a $\mathbb{Z}/2\mathbb{Z}$-action on $\langle a \rangle$ by conjugation because $\langle a \rangle$ is abelian: given any coset $b\langle a \rangle \in G/\langle a \rangle$, the action of b on $\langle a \rangle$ only depends on the coset.

So we need to understand the actions of $\mathbb{Z}/2\mathbb{Z}$ on $\langle a \rangle$. Well, $\langle a \rangle \cong \mathbb{Z}/4\mathbb{Z}$ only has the automorphisms id and $a^k \mapsto a^{-k}$. Now, fix $b \in G \setminus \langle a \rangle$ so that we know

$$\begin{cases} bab^{-1} = a, & \text{or} \\ bab^{-1} = a^{-1}. \end{cases}$$

However, we note that for $b \in G/\langle a \rangle$, we see that b^2 needs to be in $\langle a \rangle$, so $b^2 \in \{1, a, a^2, a^3\}$, but in fact $b^2 = a$ and $b^2 = a^3$ are the same by taking $a \mapsto a^{-1}$. This gives us lots of cases, which we tabulate.

	$bab^{-1} = a$	$bab^{-1} = a^{-1}$
$b^2 = e$	?	?
$b^2 = a$	?	?
$b^2 = a^2$	?	?

We note that $bab^{-1} = a$ forces our group to be abelian because a and b generate. We now go through these in sequence.

- The case of $b^2 = e$ gives us $G \cong \langle b \rangle \times \langle a \rangle \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$.
- The case $b^2 = a$ gives $\mathbb{Z}/8\mathbb{Z}$ (b has order 8).
- In the last case, we see that $(ba)^2 = 1$, so $ba \mapsto b$ throws us into the abelian with $b^2 = e$ case, so we have $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ again.

Remark 1.61. The case of $b^2 = e$ makes the short exact sequence

$$1 \to \langle a \rangle \to G \to G/\langle a \rangle \to 1$$

split with $\langle b \rangle$ as our lift of $G/\langle a \rangle$.

So here is the table so far.

	$bab^{-1} = a$	$bab^{-1} = a^{-1}$
$b^2 = 1$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$	?
$b^2 = a$	$\mathbb{Z}/8\mathbb{Z}$	?
$b^2 = a^2$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$	?

Now we start looking at our nonabelian groups.

- The case of $b^2 = e$ is our *split* case, which is $\mathbb{Z}/4\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$. This turns out to be the symmetries of the square, which we name D_8. (Here, a is a rotation by $90°$, and b is a reflection.)
- In the case of $b^2 = a$, we have a problem because the order of b looks like 8. In particular, we supposed that we have no element of order 8, so $a^2 = b^4 = e$, which violates the order of a.
- The last case is the most interesting: it gives us the quaternion group. Renaming our elements to i, j, we have the group presentation

$$Q_8 \coloneqq \left\langle i, j : i^4 = j^4 = iji^{-1}j = e \right\rangle.$$

  So does this group actually exist? Well, let's realize Q_8 as an action on a vector space. It turns out we can write

$$i = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \qquad j = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad k = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}.$$

  We can check that i and j satisfy the relations needed of them from Q_8 and that they generate a group of order 8.

So we have the following table.

	$bab^{-1} = a$	$bab^{-1} = a^{-1}$
$b^2 = 1$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$	D_8
$b^2 = a$	$\mathbb{Z}/8\mathbb{Z}$	impossible
$b^2 = a^2$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$	Q_8

In total, we have the following proposition.

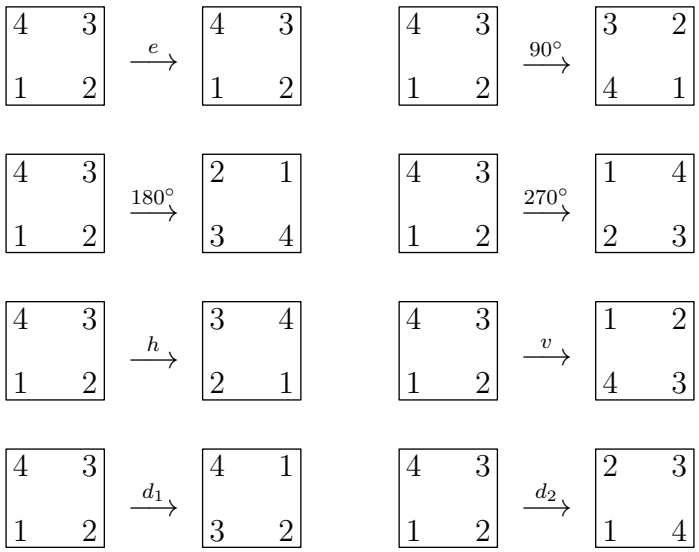
Proposition 1.62. We have the following classification of groups of order 8.

- Abelian: $\mathbb{Z}/8\mathbb{Z}$, $\mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, and $(\mathbb{Z}/2\mathbb{Z})^3$.
- Nonabelian: D_8, Q_8.

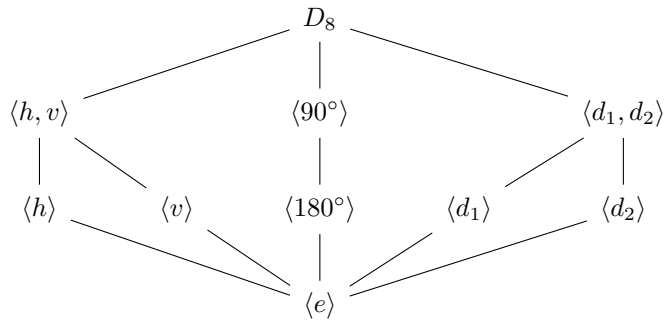
Proof. Given above. ∎

1.3.2 Quaternion Talk

Let's study D_8 and Q_8 a bit more closely by studying their subgroups. Before giving the subgroup lattice for D_8, we name our elements more concisely. They are as follows.

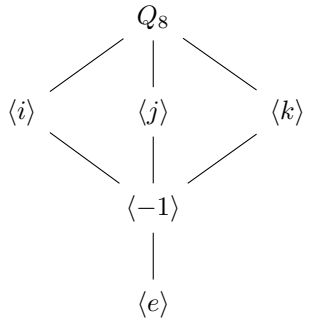


And so here is the subgroup lattice for D_8.



Note in particular that all of our order-4 subgroups ($\langle h, v\rangle$, $\langle 90°\rangle$, and $\langle d_1, d_2\rangle$) are normal because they are index-2, but not all of the order-2 subgroups are normal. For example, conjugating $\langle h\rangle$ by $90°$ gives $\langle v\rangle$. (However, $\langle 180°\rangle$ is our center and hence normal.)

And here is the lattice for Q_8.



Again, our subgroups $\langle i\rangle, \langle j\rangle$, and $\langle k\rangle$ are all normal because they are index-2, but in fact all of our subgroups are normal! Indeed, we only have one element of order 2 (which can be checked by hand), which is -1, and $\langle -1\rangle$ is our center and hence normal.

Also, the Q_8 group also creates a group ring, which is called $\mathbb{H}$, the *Hamiltonians*.

Definition 1.63 (Hamiltonians)**.** The *Hamiltonians* $\mathbb{H} := \mathbb{Z}[Q_8]$ is a noncommutative ring satisfying the relations $i^2 = j^2 = k^2 = ijk = -1$ and $ij = -ji = k^2$ and $jk = -kj = i$ and $ki = -ik = j$.

Can we go further? There are octonians, but their multiplication isn't even associative, so we don't care much about them.

Remark 1.64. For some reason, crackpots spend a long time trying to invent new $\mathbb{R}$-algebras like the above.

One reason that the quaternions are not too terrible to work with is that we were able to represent them inside of $\mathbb{C}^{2\times 2}$ as given above, so we have a pretty physical realization of these numbers. Also, quaternions are very good at describing rotations. The idea is to embed $\mathbb{R}^3$ into $\mathbb{H}$ by

$$\langle x, y, z\rangle \longmapsto xi + yj + zk.$$

Then a quaternion $g \in \mathbb{H}$ acts on $\langle x, y, z\rangle$ by conjugation: $v \mapsto ghg^{-1}$. We can check that this is a rotation of $\mathbb{R}^3$, which can be done by hand. And we can see that we can achieve all rotations by restricting our view to the elements with norm 1. In fact, the norm has the nice properties that $g = a + bi + cj + dk$ has

$$g\overline{g} = (a + bi + cj + dk)(a - bi - cj + dk) = a^2 + b^2 + c^2 + d^2,$$

so in fact our norm is nicely multiplicative. In other words, we get a surjective homomorphism from $S^3 = \{(a, b, c, d) \in \mathbb{R}^4 : a^2 + b^2 + c^2 + d^2 = 1\}$ to rotations $\mathrm{SO}_3(\mathbb{R})$. It turns out that there is nontrivial kernel here,

and in fact we have the short exact sequence

$$1 \to \{\pm 1\} \to S^3 \to \mathrm{SO}_3(\mathbb{R}) \to 1,$$

and this sequence turns out to be non-split!

Remark 1.65. Quaternions only require 4 numbers to represent a rotation, which is much nicer than representing these as 3×3 matrices, which requires more than twice as many numbers. As far as making money is concerned, this is probably the most useful fact you'll learn in this course.

Our non-split short exact sequence gives us ideas.

Definition 1.66 (Binary rotation groups)**.** Given a rotation group $G \subseteq \mathrm{SO}_3(\mathbb{R})$, we can check what happens when we pull it back into $\mathrm{SO}_3(\mathbb{R})$. For example, we can make G the rotations of a cube or the pentagon. The pullback will have twice that size because of the kernel $S^3 \to \mathrm{SO}_3(\mathbb{R})$, which are called the *binary rotation groups*.

1.3.3 Philosophy

Our work above more or less classifies all extension problems

$$1 \to \mathbb{Z}/4\mathbb{Z} \to G \to \mathbb{Z}/2\mathbb{Z} \to 1.$$

Doing this in general is hard, but there are tools. For example, the following theorem exists.

Theorem 1.67 (Schur–Zassenhaus)**.** Fix

$$1 \to A \to B \to C \to 1$$

a short exact sequence such that $\#A$ and $\#C$ are coprime. Then the short exact sequence splits, so $B \cong A \rtimes C$.

This isn't that terrible to prove, but the following theorem is very hard.

Theorem 1.68. Fix as above. Then all liftings of C into B are conjugate.

This turns out to be very hard, which requires maybe 300 pages to prove. This happens in group theory, where simple statements turn out to have very long and difficult proofs; roughly speaking, this is because it requires a proof of the Feit–Thompson theorem, which is also notoriously hard (and has been computer-verified!).

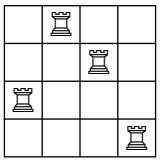
Remark 1.69 (Nir)**.** In fact, this is a general property of math: simple statements can have complicated proofs, and in fact, some simple statements must have complicated proofs. Roughly speaking, this is because determining if a given statement is true is uncomputable.

1.3.4 Rooks on a Chessboard

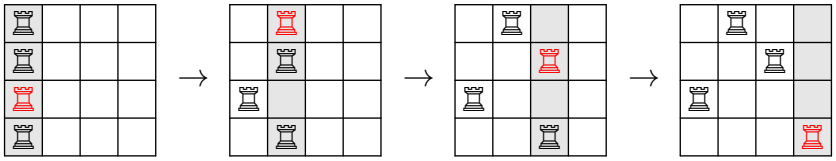
We have the following classical problem, which we'll talk about.

Question 1.70. How many ways can we place 8 rooks on a chessboard?

In other words, we are placing 8 objects on an 8×8 grid, none of which are in the same row or column. For example, the following is valid arrangement of rooks in a 4×4 grid.



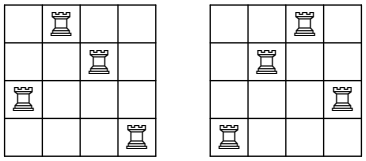
The answer to Question 1.70 turns out to be not that hard: it's just $8 \times 7 \times \cdots \times 1$ because we can just move from each column, going left to right, choosing a row that hasn't been chosen before to place our new rook. The first column has 8 options for row, then 7 options, then 6 options, and so on, totaling to $8!$. Here is an example of the process for the 4×4 case.



Let's make Question 1.70 more difficult.

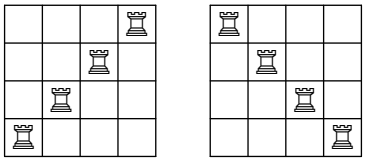
Question 1.71. How many ways, up to symmetry, can we place 8 rooks on a chessboard?

As an example of what we mean, here are two ways to place rooks on a 4×4 board, which are the same "up to symmetry," the symmetry here being the horizontal flip h.



For this, we need to understand the symmetries of chessboard, which is simply D_8, which acts on the set of all $8!$ arrangements of rooks on a chessboard. We want to know how many orbits of this D_8-action there are.

A first approximation is that any given arrangement gives rise to 8 different arrangements in its orbit, yielding $8!/8 = 5040$ total arrangements, but this is not the case. For example, the following two arrangements are a single orbit.



Namely, the problem is that some arrangements are more symmetric than others: the above arrangement only has an orbit of size 2 because it is fixed by 4 symmetries. So this appears very hard because we would have to check each individual arrangement of rooks and then check their symmetries. This seems very hard.

1.3.5 (Not) Burnside's Lemma

To solve this problem, there is Burnside's lemma.

Remark 1.72. Burnside's lemma is the Lemma which is not Burnside's. It was called Burnside's lemma by pure incompetence, and the name has stuck.

Theorem 1.73 (Not Burnside's)**.** The number of orbits of G on a set S is the average number of fixed points of elements of g. Namely,

$$\#(S/G) = \frac{1}{\#G}\sum_{g\in G} \#\{x \in S : gx = x\}.$$

This is much better because summing over the number of elements of G is much more tractable than summing over all possible arrangements of the rooks.

Proof of Theorem 1.73. The idea is to look at pairs $(g, x) \in G \times S$ such that $gx = x$. We count these pairs in two ways. In one direction, we can write

$$\{(g, x) \in G \times S : gx = x\} = \sum_{g\in G} \#\{x \in S : gx = x\}.$$

Alternatively, we can sum over S, which looks like

$$\{(g, x) \in G \times S : gx = x\} = \sum_{x\in S} \{g \in G \times S : gx = x\} = \sum_{x\in S} \#\operatorname{Stab}(x).$$

However, because we have a G-action, we may group the sum by orbits $Gx_0 \in S/G$. Indeed, for each orbit $Gx_0 \in S/G$, we see that the size of the stabilizer $\{g \in G : gx = x\}$ is the same for any $x \in Gx_0$. (Namely, g fixes x_0 if and only if hgh^{-1} fixes $hx_0 \in Gx_0$, so $\operatorname{Stab}(hx_0) = h\operatorname{Stab}(x_0)h^{-1}$.) Thus, we see

$$\{(g, x) \in G \times S : gx = x\} = \sum_{Gx_0\in S/G} \#(Gx_0) \cdot \#\operatorname{Stab}(x_0).$$

However, by the Orbit-stabilizer theorem, we see that $\#(Gx_0) = [G : \operatorname{Stab}(x_0)$, so

$$\{(g, x) \in G \times S : gx = x\} = \#G \sum_{Gx_0\in S/G} 1 = \#G \cdot \#(S/G).$$

It follows that

$$\#G \times \#(S/G) = \sum_{g\in G} \{x \in S : gx = x\},$$

which is what we wanted. ∎

Remark 1.74. If we look at the group element $g = e$, then we see that $\{x \in S : ex = x\} = S$, so we get $\#(S/G) \approx \#S/\#G$, which was our first-order approximation.

1.3.6 Back to the Rooks

Let's use this for our rooks. Take $G = D_8$. Our elements, as before, are id, h, v, d_1, and d_2. Then we also have $90°, 180°$, and $270°$. However, there is some repetition here because h and v have the same number of fixed points; similarly, d_1 and d_2 or $90°$ and $270°$ also have the same number of fixed points.

Remark 1.75. Note that these are the conjugacy classes of D_8. More generally in a group G, if two elements of our group g_1 and g_2 are conjugate, then they have the same number of fixed points. Namely, if we have $g \in G$ such that $g_1 = gg_2g^{-1}$, then

$$\{x \in S : g_1x = x\} \xrightarrow{\times g} \{x \in S : g_2x = x\}.$$

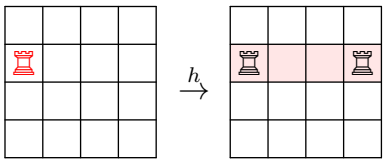
In words, $x \in S$ is a fixed point of g_1 if and only if gx is a fixed point of g_2.

Warning 1.76. Note that the $90°$ and $270°$ rotations, though they "look the same," are conjugate only in D_8 but *not* in the group of rotations $\langle 90° \rangle$. In Rhea's words, we need a reflection to make this work.

We now go down the list of D_8.

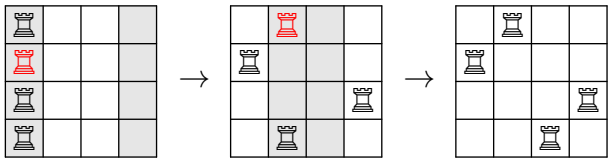
- For e, everything is a fixed point, so there are $8!$ arrangements here.
- For h and v, nothing is a fixed point because a rook in a particular row (respectively, column) would get moved somewhere else in the same row (respectively, column), which violates the conditions of placing rooks.

  Here is the image for the 4×4 case.

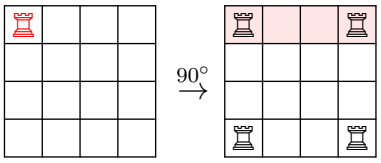


- For $180°$, there are $8 \times 6 \times 4 \times 2$ arrangements because placing one rook forces its inverse as well. So we place one rook and lose two options simultatenously.

  Here is the image for the 4×4 case, where we only make two choices.

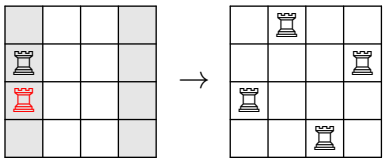


- For $90°$ and $270°$, we get 6×2. This is because, placing a rook in the first row, we can only place rooks outside the corners (or else they run into each other), which gives 6 options. Here is the image of this in the 4×4 case.



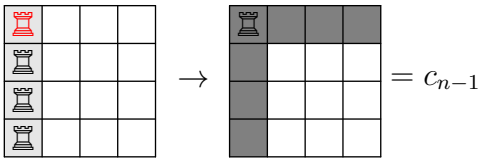
  After placing a rook in the first row, we lose four options because we need to place four rooks from the first one, which gives 2 options afterwards because we still cannot place in corners.

  Here is the image for the 4×4 case, where we only make one choice.

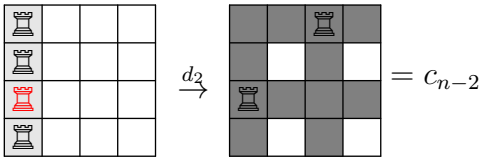


- We have d_1 and d_2 are the hardest. For concreteness, we count for d_2. We do this by a recursion: let c_n be the number of arrangements fixed by d_1 in a $n \times n$ board. We claim that $c_n = c_{n-1} + (n-1)c_{n-2}$. We have two cases.

  If we place the first rook in the top left corner, then we reduce this problem to the $(n-1) \times (n-1)$ board. Here is the image for that in the 4×4 case.



Otherwise, if we place our rook somewhere else in the first row, then we lose both a row and a column from the d_1 symmetry, reducing to the $(n-2) \times (n-2)$ case. Here is the image for that in the 4×4 case.



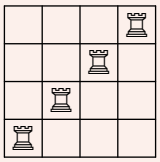
So indeed, $c_n = c_{n-1} + (n-1)c_{n-2}$, and we can compute that $c_8 = 764$.

In total, we get

$$\frac{40320 + 2 \cdot 0 + 2 \cdot 764 + 2 \cdot 12 + 8 \cdot 6 \cdot 4 \cdot 2}{8} = \boxed{5282}$$

This is a bit bigger than our first guess, which was $8!/8 = 5040$.

Remark 1.77. We can bypass Burnside's lemma by cheating a bit. The idea is to weight each orbit in S/G we are counting so that we don't have to look directly at the group: we weight an orbit by the reciprocal of its symmetry group. (This is contrast to weighting the orbits equally to count them.) For example, the following arrangement is weighted $1/4$ because of its four symmetries.



Why do we do this? Well, it turns out that the number of weighted orbits is $\#S/\#G$ exactly: write

$$\sum_{Gx \in S/G} \frac{1}{\#\{g \in G : gx = x\}} = \sum_{Gx \in S/G} \frac{1}{\#\operatorname{Stab}(x)} = \sum_{Gx \in S/G} \frac{\#Gx}{\#G} = \frac{\#S}{\#G}.$$

1.3.7 Groups of Order Nine

For groups of order 9, the obvious groups are $\mathbb{Z}/9\mathbb{Z}$ and $(\mathbb{Z}/3\mathbb{Z})^2$. These are the only abelian ones: if there's an element of order 9, then we are cyclic; otherwise, all elements have order 3, then we are an $\mathbb{F}_3$-vector space, forcing us to be $(\mathbb{Z}/3\mathbb{Z})^2$ for the usual size reasons.

What about non-abelian groups? We claim there are no nonabelian groups.

Proposition 1.78. All groups of order p^2 are abelian, for p prime.

Proof. Fix G of order p^2. Note that all proper subgroups have index divisible by p (the index is either p or p^2). In particular, the center has order divisible by p, borrowing the class equation logic from the proof of Cauchy's theorem.

We claim that $Z(G) = G$. Well, suppose for the sake of contradiction that there is $b \in G \setminus Z(G)$ so that $\#Z(G) = p$. The problem is that the set of elements $C(b)$ of G which commute with b is a subgroup of G, which contains $\{b\} \sqcup Z(G)$ and hence has order exceeding p. Thus, $C(b)$ has order p^2, implying that b commutes with all elements, violating $b \notin Z(G)$. ∎

1.4 September 7

You are filled with determination.

1.4.1 Groups of Order 10

Last time we classified groups of order 9. So let's do groups G of order 10. Well, Cauchy's theorem promises subgroups $\langle a\rangle$ and $\langle b\rangle$ of order 5 and 2 respectively. But now $[G:\langle a\rangle] = 2$ is normal, so we have the short exact sequence

$$1 \to \langle a\rangle \to G \to \langle b\rangle \to 1,$$

so in particular, it follows that G is the semidirect product of $\mathbb{Z}/5\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$. So we can determine G entirely based off of the $\mathbb{Z}/2\mathbb{Z}$-actions on $\mathbb{Z}/5\mathbb{Z}$.

This isn't very hard because $\operatorname{Aut}(\mathbb{Z}/5\mathbb{Z}) = (\mathbb{Z}/5\mathbb{Z})^\times \cong \mathbb{Z}/4\mathbb{Z}$, but $\mathbb{Z}/4\mathbb{Z}$ only has two elements of order two, so the only $\mathbb{Z}/2\mathbb{Z}$-actions on $\mathbb{Z}/5\mathbb{Z}$ are id and $x \mapsto x^{-1}$. So we get two groups of order 10, defined by

$$\begin{cases} bab^{-1} = a, \\ bab^{-1} = a^{-1}. \end{cases}$$

The first case is abelian and hence is $\mathbb{Z}/5\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \cong \mathbb{Z}/10\mathbb{Z}$. The second case is nonabelian, and we know a nonabelian group of order 10, namely D_{10}, so this must be that group.

Definition 1.79 (Dihedral group)**.** The dihedral group D_{2n} is the group of symmetries of a regular n-gon.

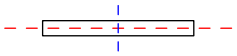
We remark that this logic can be generalized.

Proposition 1.80. Let G be a group of order $2p$ for p prime. Then $G \cong \mathbb{Z}/(2p)\mathbb{Z}$ or $G \cong D_{2p}$.

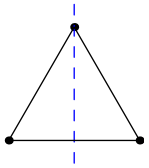
1.4.2 Dihedral Groups and Involutions

Let's look at some dihedral groups.

- D_4 is the group of symmetries of a line, which is $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. This should really be imagined as the symmetries of a rectangle, where one of the sides is very thin. Here we have highlighted the horizontal and vertical flips.

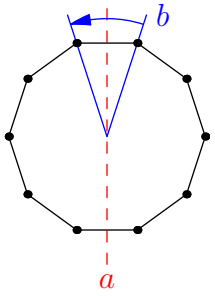


- D_6 is the group of symmetries of a triangle, which is S_3 because reflections can transpose any two vertices. For example, the following reflection transposes the bottom two vertices.



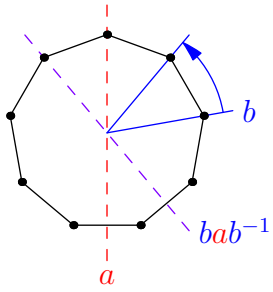
- D_8 is the group of symmetries of a square.
- In general D_{2n} is the group of symmetries of a regular n-gon.

In general $D_{2n} = \mathbb{Z}/n\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$, where $\mathbb{Z}/n\mathbb{Z}$ is a rotation and $\mathbb{Z}/2\mathbb{Z}$ is a reflection. Here's the picture for $D_{20} = \left\langle a^2 = b^{10} = e, aba^{-1} = b^{-1}\right\rangle$.

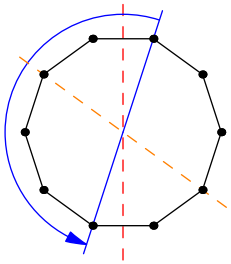


Also observe that each D_4, D_8, D_{12}, and so on all have nontrivial center, namely rotation by $180°$. (In the group presentation $D_{4k} = \left\langle a^2 = b^{2k} = e, aba^{-1} = b^{-1} \right\rangle$, this is $Z(D_{4k}) = \{e, b^k\}$.)

Let's continue talking about order-2 elements, or "involutions." In D_2, D_6, D_{10}, and so on all have their order-2 are conjugate. Indeed, all order-2 elements are reflections (there is no $180°$ rotation), which can all be rotated into each other, and this rotation corresponds to conjugation. Here is the image of rotating a translation for D_{18}, the symmetries of a nonagon.



Regardless, D_{4n} only has three types of involutions: a $180°$ rotation, reflection where the line goes through a vertex, and reflection where the line goes through the midpoint of a side. Here's a picture for the three types in D_{20}, the symmetries of a decagon.



(To reiterate, in D_{4n+2}, there is no $180°$ rotation, and reflections all go through both a vertex and a side.) In particular, having the $180°$ rotation in the center implies that there is a nontrivial element which commutes with all of our order-2 elements.

This property turns out to answer the following question.

Question 1.81. Is there a general property which holds for all groups of finite order but fails for some groups of infinite order?

Well, of course, "the group is finite" is some property, but this is not what we want. Namely, we want our property to be stated in terms of group theory: we only want to use group elements, their multiplication structure, and first-order logic.

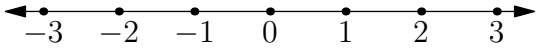
Non-Example 1.82. Here is something which doesn't work: $\forall g \in G, \exists n \in \mathbb{Z}, g^n = e$. This doesn't work because $n \in \mathbb{Z}$ is invalid.

However, this can be done. For example,

$$\forall g, h, \left[gg = hh = e \wedge g \neq e \neq h\right] \implies \left[(\exists g_0 : g_0 h g_0^{-1} = g \vee (\exists g_0 \neq e, gg_0 = g_0 g \wedge hg_0 = g_0 h)\right]. \qquad (*)$$

In other words, any two elements of order-2 are conjugate or have a nontrivial third element commuting with both of them. We see $(*)$ works for any finite group because g and h will generate a dihedral group (generated by the "reflection" g and the "rotation" gh), and we checked that this statement holds for dihedral groups.

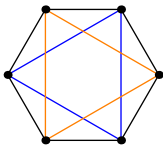
However, $(*)$ fails for the "infinite" dihedral group as symmetries of $\mathbb{Z}$, which is $\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$. To be explicit, we can imagine this as the group of symmetries of the number line.



We can check that $g : x \mapsto -x$ and $h : x \mapsto 1 - x$ are neither conjugate nor commuting with a third element[2] even though $g^2 = h^2 = \mathrm{id}$, so indeed, $(*)$ fails here.

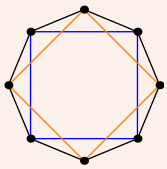
Something else funny about dihedral groups is that some of these split as products.

- $D_4 \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, generated by the vertical and horizontal flips.
- $D_{12} \cong D_6 \times \mathbb{Z}/2\mathbb{Z}$. Here, this is the symmetries of a hexagon, but inside the hexagon we can draw a triangle.

  We see that we can write D_{12} as the symmetries of the blue triangle, times perhaps a $180°$ rotation mapping the blue triangle to the orange triangle. This turns out to create a direct product, commuting because the $180°$ rotation is in the center.
- In general, $D_{8k+4} \cong D_{2k} \times \mathbb{Z}/2\mathbb{Z}$ by generalizing the above argument.

Remark 1.83. The above product decomposition does not work for (say) D_{16}. If we tried, we would get the following two squares.



The issue is that the $180°$ rotation that is supposed to send the blue square to the orange one already lives in the symmetries of the embedded square. Perhaps we could map the blue square to the red one in some other way, but this would lose being a direct product because $180°$ is the only element of the center.

Remark 1.84. It's not hard to see that any group generated by two elements of order 2 are either abelian or dihedral. However, for two elements of order three turns into a terrible mess: for example, we can achieve any finite simple group. So elements of order two are nice.

[2] They aren't conjugate because g has one fixed point while h has none. They don't commute with any third element by brute force: our group is generated by $x \mapsto x + 1$ and $x \mapsto -x$, so all elements take the form $g_0 : x \mapsto \pm x + n$ for some $n \in \mathbb{Z}$. We see $(gg_0)(1) = (g_0 hg)(1)$ implies $n = 0$, so the only nontrivial option for g_0 is g itself, but $gh \neq hg$.

1.4.3 Groups of Orders 11 and 12

Groups of order 11 are cyclic because 11 is prime. So let's just jump into 12. There are five of them; here are some obvious ones:

$$\mathbb{Z}/6\mathbb{Z} \times \mathbb{Z}/\mathbb{Z}, \qquad S_3 \times \mathbb{Z}/2\mathbb{Z}, \qquad \mathbb{Z}/12\mathbb{Z}, \qquad A_4.$$

We note that $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z} \cong \mathbb{Z}/12\mathbb{Z}$ and $D_{12} \cong S_3 \times \mathbb{Z}/2\mathbb{Z}$, so these are already in our list. So where's the fifth group? Well, let's find out.

1.4.4 Sylow Time

Let's return to our attempt at reversing Lagrange.

Question 1.85. We know that $H \subseteq G$ as groups implies $\#H \mid \#G$. But if we have $n \mid \#G$, then is there a subgroup of order n?

The answer turns out to be no: A_4 has no subgroups of order 6. (Again, we can check this by hand.)

However, we can salvage Question 1.85: it turns out to be true if n is a prime power, which is a special case of the Sylow theorems.

Remark 1.86. Nobody actually knows how to pronounce "Sylow." There's no point trying to pronounce it correctly because no matter how hard you try, a Norwegian will smile patronizingly at you and tell you you're wrong.

Here's the statement.

Theorem 1.87 (Sylow, I)**.** Fix $p^{\nu_p(n)}$ the largest prime power of p dividing $n := \#G < \infty$. Then there is a subgroup of order $p^{\nu_p(n)}$.

Definition 1.88 (Sylow subgroups)**.** The subgroups in Theorem 1.87 are called Sylow p-subgroups.

Proof. There are two possibilities.

- If there are some subgroups which have proper subgroups with index not divisible by p, then we can just induct on one of these subgroups because their orders will also be divisible by $p^{\nu_p(n)}$.
- Otherwise, all proper subgroups have index divisible by p. But we saw in Remark 1.49 that this implies that $Z(G)$ has order divisible by p. So Cauchy's theorem gives us an element $g \in Z(G)$ with order p. In particular, we have the short exact sequence

$$1 \to \langle g \rangle \to G \to G/\langle g \rangle \to 1.$$

  Indeed, because $\langle g \rangle$ is in the center, $\langle g \rangle$ is normal, so $G/\langle g \rangle$ is actually a group. To finish, we use induction to get a Sylow p-subgroup of $G/\langle g \rangle$, and we can pull this backwards along the modulo by g map to get a subgroup of G of the correct order. ∎

Remark 1.89. The end of this proof uses the fact that pre-images of subgroups are subgroups. To see why this is true, fix $\varphi : G \to H$ a group homomorphism and $B \subseteq H$ a subgroup. Then $A := \varphi^{-1}(B)$ contains $e \in \varphi^{-1}(\{e\})$, is closed under multiplication ($\varphi(a_1), \varphi(a_2) \in B \implies \varphi(a_1 a_2) \in B$), and is closed under inversion ($\varphi(a) \in B \implies \varphi(a^{-1}) = \varphi(a)^{-1} \in B$).

1.4.5 Back to Groups of Order 12

Let's return to groups of order 12; fix G with $\#G = 12$. From Theorem 1.87, we see that G has a subgroup of order 3 and a group of order 4. We would like a normal subgroup; do we have one? Well, let's do casework.

(a) If our subgroup H_3 of order 3 is normal, then $G = \mathbb{Z}/3\mathbb{Z} \rtimes H_4$, where H_4 is our Sylow 2-subgroup.

(b) If H_3 is not normal, then it has four conjugates, by another Sylow theorem we will prove later. This gives us $4 \cdot 2$ elements of order 3, so we have exactly four elements left over, which must be our H_4 of order 4, and we see that H_4 is normal because we can only have one of them.

So in this case, we see that $G = H_4 \rtimes \mathbb{Z}/3\mathbb{Z}$ by letting an $H_3 \cong \mathbb{Z}/3\mathbb{Z}$ act on our Sylow 2-subgroup.

We now work out our cases separately.

Case (b)

We start with (b). We have the following table.

	trivial $\mathbb{Z}/3\mathbb{Z}$-action	nontrivial $\mathbb{Z}/3\mathbb{Z}$-action
$(\mathbb{Z}/2\mathbb{Z})^2$		
$\mathbb{Z}/4\mathbb{Z}$		

When $\mathbb{Z}/3\mathbb{Z}$ acts trivially, our group is abelian, so the top left is $(\mathbb{Z}/2\mathbb{Z})^2 \times \mathbb{Z}/3\mathbb{Z}$, and the bottom left is $\mathbb{Z}/12\mathbb{Z}$.

	trivial $\mathbb{Z}/3\mathbb{Z}$-action	nontrivial $\mathbb{Z}/3\mathbb{Z}$-action
$(\mathbb{Z}/2\mathbb{Z})^2$	$(\mathbb{Z}/2\mathbb{Z})^2 \times \mathbb{Z}/3\mathbb{Z}$	
$\mathbb{Z}/4\mathbb{Z}$	$\mathbb{Z}/12\mathbb{Z}$	

We now work on the right column. We need to consider a nontrivial map $\mathbb{Z}/3\mathbb{Z} \to \operatorname{Aut}\left((\mathbb{Z}/2\mathbb{Z})^2\right)$. Writing our $(\mathbb{Z}/2\mathbb{Z})^2$ as $\{e, a_1, a_2, a_3\} \subseteq G$, we see that an element of $\operatorname{Aut}\left((\mathbb{Z}/2\mathbb{Z})^2\right)$ must fix e and hence essentially be a permutation in S_3 on $\{a_1, a_2, a_3\}$. It turns out that these are all actually automorphisms.[3]

Now, if $\mathbb{Z}/3\mathbb{Z} \to \operatorname{Aut}\left((\mathbb{Z}/2\mathbb{Z})^2\right) \cong S_3$ is to be nontrivial, then we need to send our order-3 element (which we name conjugation by b) to a three-cycle in S_3. By switching around our elements, it doesn't matter which one, so we have the restrictions

$$\left\langle b^3 = a_1^2 = a_2^2 = a_3^2 = a_1a_2a_3^{-1} = e, \quad ba_1b^{-1} = a_2, \quad ba_2b^{-1} = a_3, \quad ba_3b^{-1} = a_1 \right\rangle,$$

or after doing some reduction,

$$\left\langle b^3 = a^2 = (ab)^3 = e \right\rangle.$$

This turns out to be A_4 because we can take $a = (12)(34)$ and $b = (123)$. (We can check by hand that there are twelve elements in the above group presentation.) So our table looks like the following.

	trivial $\mathbb{Z}/3\mathbb{Z}$-action	nontrivial $\mathbb{Z}/3\mathbb{Z}$-action
$(\mathbb{Z}/2\mathbb{Z})^2$	$(\mathbb{Z}/2\mathbb{Z})^2 \times \mathbb{Z}/3\mathbb{Z}$	A_4
$\mathbb{Z}/4\mathbb{Z}$	$\mathbb{Z}/12\mathbb{Z}$	

Lastly we can have $\mathbb{Z}/3\mathbb{Z}$ act on $\mathbb{Z}/4\mathbb{Z}$. However, $\operatorname{Aut}(\mathbb{Z}/4\mathbb{Z}) \cong (\mathbb{Z}/4\mathbb{Z})^\times \cong (\mathbb{Z}/2\mathbb{Z})$ has no order-3 elements, so this is impossible. So in total, we have the following table.

	trivial $\mathbb{Z}/3\mathbb{Z}$-action	nontrivial $\mathbb{Z}/3\mathbb{Z}$-action
$(\mathbb{Z}/2\mathbb{Z})^2$	$(\mathbb{Z}/2\mathbb{Z})^2 \times \mathbb{Z}/3\mathbb{Z}$	A_4
$\mathbb{Z}/4\mathbb{Z}$	$\mathbb{Z}/12\mathbb{Z}$	impossible

[3] I think the most "pure thought" way to see this is to view $(\mathbb{Z}/2\mathbb{Z})^2$ as a $\mathbb{Z}$-module, so we see that it is actually a $\mathbb{F}_2$-vector space, so $\operatorname{Aut}\left((\mathbb{Z}/2\mathbb{Z})^2\right) = \operatorname{Hom}_{\mathbb{Z}}\left((\mathbb{Z}/2\mathbb{Z})^2, (\mathbb{Z}/2\mathbb{Z})^2\right) = \operatorname{GL}_2(\mathbb{F}_2)$. As described previously, we have an injection $\operatorname{Aut}\left((\mathbb{Z}/2\mathbb{Z})^2\right) \hookrightarrow S_3$, but $\#\operatorname{GL}_2(\mathbb{F}_2) = 6$, so $\operatorname{Aut}\left((\mathbb{Z}/2\mathbb{Z})^2\right) \cong S_3$.

Case (a)

In part (a), we have the following table.

	trivial action	nontrivial action
by $\mathbb{Z}/4\mathbb{Z}$		
by $(\mathbb{Z}/2\mathbb{Z})^2$		

Here everything is acting on our Sylow 3-subgroup $\mathbb{Z}/3\mathbb{Z}$. In the left column, our group is abelian, so we can fill these out.

	trivial action	nontrivial action
by $\mathbb{Z}/4\mathbb{Z}$	$\mathbb{Z}/12\mathbb{Z}$	
by $(\mathbb{Z}/2\mathbb{Z})^2$	$(\mathbb{Z}/2\mathbb{Z})^2 \times \mathbb{Z}/3\mathbb{Z}$	

In the bottom right, we have $(\mathbb{Z}/2\mathbb{Z})^2$ acting nontrivially on $\mathbb{Z}/3\mathbb{Z}$. Again, there is only nontrivial automorphism of $\mathbb{Z}/3\mathbb{Z}$, so we had better send some element of $(\mathbb{Z}/2\mathbb{Z})^2$ there. Without loss of generality, we send $(1,0)$ to the nontrivial automorphism; then exactly one of $(1,1)$ or $(0,1)$ will be nontrivial as well, so we'll say $(0,1)$ is nontrivial. In total, we have the presentation

$$\left\langle a_1^2 = a_2^2 = b^3 = e, \quad a_1 b a_1^{-1} = a_2 b a_2^{-1} = b^2 \right\rangle.$$

This turns out to be $\mathbb{Z}/2\mathbb{Z} \times S_3$. For example, we can take $a_1 \mapsto (1, \mathrm{id})$ and $a_2 \mapsto (1, (12))$ and $b \mapsto (0, (123))$. So our table so far is the following.

	trivial action	nontrivial action
by $\mathbb{Z}/4\mathbb{Z}$	$\mathbb{Z}/12\mathbb{Z}$	
by $(\mathbb{Z}/2\mathbb{Z})^2$	$(\mathbb{Z}/2\mathbb{Z})^2 \times \mathbb{Z}/3\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z} \times S_3$

In the top right, we have $\mathbb{Z}/4\mathbb{Z}$ acting nontrivially on $\mathbb{Z}/3\mathbb{Z}$. Well, $\mathrm{Aut}(\mathbb{Z}/3\mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z}$, letting $\mathbb{Z}/4\mathbb{Z} \cong \langle a \rangle \subseteq G$ and $\mathbb{Z}/3\mathbb{Z} \cong \langle b \rangle \subseteq G$, we are forced into $aba^{-1} = b^{-1}$. So we have the following presentation.

$$\left\langle a^4 = b^3 = e, \quad aba^{-1} = b^2 \right\rangle.$$

And here, we call it quits, having more or less classified all groups of order 12.

Remark 1.90. This last group is hard to visualize. It turns out to be a binary dihedral group. Namely, we recall that our binary dihedral groups were defined as pull backs from $\mathrm{SO}_3(\mathbb{R})$ in the short exact sequence

$$1 \to \{\pm 1\} \to S^3 \to \mathrm{SO}_3(\mathbb{R}) \to 1.$$

Namely, we take $D_6 \cong S_3 \subseteq \mathrm{SO}_3(\mathbb{R})$ as the group of symmetries of a triangle (which is in $\mathrm{SO}_3(\mathbb{R})$ by placing the triangle in 3-space) and pull it back into S^3 to get G_6.

1.4.6 Back to Sylow

Let's go back and prove that one fact that told us $H_3 \subseteq G$ has four subgroups of order 3 if not normal. We will go through the Sylow theorems one at a time, though we will not do this in order.

Theorem 1.91 (Sylow, III(a))**.** Fix G a finite group and p prime. Then the number n_p of Sylow p-subgroups is $1 \pmod p$.

Proof. We show that $n_p \equiv 1 \pmod p$. The following lemma is the meat of the argument.

Lemma 1.92. Fix G a finite group with Sylow p-subgroups named S_1 and S_2. Then S_1 does not normalize S_2.

Proof. Note that if S_1 normalizes S_2 for Sylow p-subgroups S_1 and S_2, then S_1S_2 is a subgroup of order which is larger than $p^{\nu_p(n)}$, which is a contradiction.

For completeness, we check that S_1S_2 a subgroup if S_1 normalizes S_2. Well, for $g_1, g_2 \in S_1$ and $h_1, g_2 \in S_2$, then

$$(g_1h_1)(g_2h_2) = (g_1g_2)(g_1^{-1}h_1g_1h_2) \in S_1S_2$$

because S_1 normalizes S_2. This gives us closure under multiplication. For closure under inversion, we see $g \in S_1$ and $h \in S_2$ has $(gh)^{-1} = h^{-1}g^{-1} = g^{-1}gh^{-1}g^{-1} \in S_1S_2$. ■

To finish, fix a Sylow p-subgroup S. We check the action of S on all Sylow p-subgroups by conjugation. One orbit is $\{S\}$ because S should certainly fix itself. Otherwise, all orbits have at least one element because S doesn't normalize any other Sylow p-subgroup.

However, each orbit size not 1 still must divide $\#S$ by Orbit-stabilizer, so the size of each orbit which isn't $\{S\}$ must be divisible by p. It follows that the sum of the sizes of all orbits is $1 \pmod p$, where the 1 comes from $\{S\}$. ■

Theorem 1.93 (Sylow, II)**.** Fix G a finite group and p prime. Then all Sylow p-subgroups are conjugate.

Proof. Fix S and T two Sylow p-subgroups so that we want to show they are conjugate. The trick is to let T act on the left cosets G/S of S. We note that we can use the end of the previous argument on this action to note that the number of fixed points by this action equals the number size-1 orbits, which is

$$[G : S] \pmod p$$

after adding in the sizes of all the other orbits (which divide T and hence are divisible by p). Because $[G : S] \not\equiv 0 \pmod p$, there is some fixed point; namely, for some $gP \in G/P$, we have $tgP = gP$ for any $t \in T$. It follows $T \subseteq gPg^{-1}$, so $T = gPg^{-1}$ for size reasons. This finishes. ■

Theorem 1.94 (Sylow, III(b))**.** Fix G a finite group and p prime. Then the number n_p of Sylow p-subgroups divides $\#G$.

Proof. Lastly, because all Sylow p-subgroups are conjugate, we see that their number is

$$\frac{\#G}{\#\{g \in G : g \text{ normalizes some } S\}},$$

which divides $\#G$. Namely, this is just the Orbit-stabilizer theorem because there is only orbit, so the size of this orbit is the index in G. ■

Remark 1.95. We used Theorem 1.91 in our classification of groups of order 12: the number of Sylow 3-subgroups needed to be $1 \pmod 3$ and divide into 12 and hence must have been 1 or 4.

As a consequence of the Sylow theorems, because all Sylow subgroups are conjugate, it follows that they are isomorphic. However, this is not true in general.

Example 1.96. The group $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ has non-conjugate subgroups of order 2. Namely, it has more than one subgroup of order 2, and $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ is abelian, so these are normal. (However, there is an outer automorphism connecting them.)

Example 1.97. The group D_8 has subgroups isomorphic to $\mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, which are not even isomorphic though they both have order 4.

So it is somewhat surprising that looking at the largest power of p forces the p-subgroups to be isomorphic.

1.4.7 Nilpotent Groups

Recall that if $\#G = p^n > 1$, then we know that G has nontrivial center. We showed this by using the class equation. This lets us fix $Z_1 \coloneqq Z(G)$ and we note that $G_1 \coloneqq G/Z_1$ again has order a power of p; if it is trivial, we declare we are done, and otherwise we can fix Z_2 to be the pre-image of $Z(G/Z_1)$ in G and look at $G_2 \coloneqq G/Z_2$. Then we can just continue this inductively.

Note that killing the center by modding might still have a center afterwards, so this process isn't trivial.

Example 1.98. With Q_8, we have center $\{\pm 1\}$, but $G/\{\pm 1\}$ has center $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ because $Q_8/\{\pm 1\}$ is abelian.

What's happening is that we get the sequence

$$\{e\} = Z_0 \subseteq Z_1 \subseteq Z_2 \subseteq \cdots \subseteq Z_{n-1} \subseteq Z_n = G,$$

where $G_0 = \{e\}$ and $Z_{k+1}/Z_k = Z(G/Z_k)$.

Definition 1.99 (Nilpotent)**.** Call a group G *nilpotent* if G has such a chain.

Non-Example 1.100. The group S_3 is not nilpotent. Indeed, S_3 has trivial center, so the chain of taking centers never descends.

It turns out that nilpotence and Sylow subgroups are connected.

Proposition 1.101. The following are equivalent for a finite group G.

(a) G is nilpotent.

(b) All proper subgroups H have normalizer strictly bigger than H.

(c) All Sylow p-subgroups are normal.

(d) G is a product of groups order a power of prime.

Proof. We take these one at a time, in varying amounts of detail.

- We show that (a) implies (b). We show the contrapositive: suppose that we have a proper subgroup $H \subsetneq G$ such that $N(H) = H$, and we show that G is not nilpotent.

  The idea is to use H to bound the subgroup chain. Indeed, we show that $Z_k \subseteq H \subsetneq G$ for each k, which keeps G from being nilpotent. Certainly this is true for $Z_0 = \{e\}$.

  For the inductive step, take $Z_k \subseteq H$. Now, $g \in Z_{k+1}$ implies that $gZ_k \in Z(G/Z_k)$ implies that

  $$(gh)Z_k = gZ_k \cdot hZ_k = hZ_k \cdot gZ_k = (hg)Z_k$$

  for any $h \in G$. In particular, we see that $ghg^{-1}h^{-1} \in Z_k \subseteq H$ for any $h \in G$.

  Now, taking $h \in H$, we see that $ghg^{-1} \in H$ for each $h \in H$, so $g \in N(H) = H$! So indeed, $g \in H$, from which it follows that $Z_{k+1} \subseteq H$, completing our induction.

- We show that (b) implies (c). Fix P a Sylow p-subgroup so that we want to show P is normal in G. Well, it suffices to show that $N(P) = G$.

  The main claim is that $N(N(P)) = N(P)$. From this it will follow that $N(P)$ is not a proper subgroup, forcing $N(P) = G$. Certainly $N(P) \subseteq N(N(P))$, so we spend our time with $N(N(P)) \subseteq N(P)$.

  Well, fixing any $g \in G$, we see that $g \in N(N(P))$ implies

  $$N\left(gPg^{-1}\right) = gN(P)g^{-1} = N(P).$$

In particular, $P \subseteq N(P) = N\left(gPg^{-1}\right)$.

But gPg^{-1} is normal in $N\left(gPg^{-1}\right)$ while both gPg^{-1} and P are both Sylow p-subgroups of $N\left(gPg^{-1}\right)$ (compare the powers of p). Because Sylow p-subgroups are all conjugate, it follows that

$$P = gPg^{-1}$$

because gPg^{-1} is normal in $N\left(gPg^{-1}\right)$. So indeed, $g \in N(P)$.

- We show that (c) implies (d). This follows from the homework; fix $H_1, \dots, H_n$ our Sylow p-subgroups. The main claim is that, for N_1 and N_2 normal subgroups with trivial intersection, we have

  $$N_1 N_2 \cong N_1 \times N_2.$$

  Our isomorphism is defined by $N_1 \times N_2 \to N_1 N_2$ with

  $$\varphi : (n_1, n_2) \mapsto n_1 n_2.$$

  We see φ is homomorphic because $n_1 n_2 = n_2 n_1$ for any $n_1 \in N_1, n_2 \in N_2$ because $n_1 n_2 n_2^{-1} n_1^{-1} \in N_1 \cap N_2 = \{e\}$. We see φ is surjective by definition of $N_1 N_2$. Lastly, we see that φ has trivial kernel because $n_1 n_2 = e$ implies that $n_1 = n_2^{-1} \in N_1 \cap N_2 = \{e\}$ implies $n_1 = n_2 = e$.

  Now, we can simply inductively say

  $$H_1 H_2 \cdots H_n \cong H_1 \times H_2 \cdots H_n \cong H_1 \times H_2 \times H_3 \cdots H_n \cong \cdots \cong H_1 \times H_2 \times \cdots \times H_n.$$

  This induction works because $H_{k+1} H_{k+2} \cdots H_n$ is a normal subgroup always[4] and has order coprime to H_k (and hence trivial intersection) because the prime-power orders separate.

- We show that (d) implies (a). The point that (d) means that G is the product of its Sylow p-subgroups, and we know that p-groups are nilpotent from the above discussion. It follows that G is also nilpotent by attaching the chains together; we will not be rigorous about this because I cannot be bothered. ■

The point is that nilpotent groups are the ones which are the product of groups of prime-power order, which seems very nice. However, it turns out that there are lots of groups of prime-power order.

1.4.8 Groups of Order 13, 14, and 15

We see that 13 is prime, so all groups are cyclic. As for 14, it's twice a prime, so it's either cyclic or D_{14}.

So let's look at groups of order 15.

Proposition 1.102. Suppose G is a group with $\#G = pq$ with $p < q$ primes. Then $G = \mathbb{Z}/q\mathbb{Z} \rtimes \mathbb{Z}/p\mathbb{Z}$. If $q \not\equiv 1 \pmod{p}$, then G is cyclic.

Proof. The number of Sylow q-subgroups is $1 \pmod{q}$ and divides pq, so it must be 1. So our Sylow q-subgroup is normal, which forces $G \cong \mathbb{Z}/q\mathbb{Z} \rtimes \mathbb{Z}/p\mathbb{Z}$.

In particular, if $q \not\equiv 1 \pmod{p}$, then the action of $\mathbb{Z}/p\mathbb{Z}$ on $\mathbb{Z}/q\mathbb{Z}$ must be trivial because $\mathrm{Aut}(\mathbb{Z}/q\mathbb{Z}) \cong \mathbb{Z}/(q-1)\mathbb{Z}$ has no order-p elements. ■

Example 1.103. With $q = 5$ and $p = 3$, we see that $5 \not\equiv 1 \pmod{3}$, so any group of order 15 is cyclic.

Example 1.104. With $q = 7$ and $p = 3$, we do have a nonabelian group of order $\#G = 21$, though it is still $\mathbb{Z}/7\mathbb{Z} \rtimes \mathbb{Z}/3\mathbb{Z}$. We can represent this group by

$$\left\{ \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} : a, b \in \mathbb{F}_7 \text{ and } a \in \{1, 2, 4\} \right\}.$$

This is closed under multiplication because $a \in \{1, 2, 4\}$ is the same thing as $a \in (\mathbb{F}_7^\times)^{\times 2}$.

[4] Note $gH_{k+1}H_{k+2}\cdots H_n g^{-1} = gH_{k+1}g^{-1} \cdot gH_{k+2}g^{-1} \cdots gH_n g^{-1} = H_1 H_2 \cdots H_n$.

1.4.9 Groups of Order 16

We're not going to classify all groups of order 16 because it is a mess. However, we can list them. We have the following cases.

- Abelian: we have $\mathbb{Z}/16\mathbb{Z}$, $\mathbb{Z}/8\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, $\mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$, $\mathbb{Z}/4\mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})^2$, and $(\mathbb{Z}/2\mathbb{Z})^4$.
- There are 4 cyclic subgroups of order 8: There is a generalized quaternion group, which is binary dihedral. Otherwise, there is an element a of order eight and an element b of order 2. Then we have the cases $bab^{-1} \in \{a, a^3, a^5, a^7\}$. Not all of these even have names.
- There are products: $Q_8 \times \mathbb{Z}/2\mathbb{Z}$ and $D_8 \times \mathbb{Z}/2\mathbb{Z}$.
- There are semidirect products: $\mathbb{Z}/4\mathbb{Z} \rtimes \mathbb{Z}/4\mathbb{Z}$ and $(\mathbb{Z}/2\mathbb{Z})^2 \rtimes \mathbb{Z}/4\mathbb{Z}$. Also there is $(\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}) \rtimes \mathbb{Z}/2\mathbb{Z}$. This is sometimes called the Pauli group because it is the group generated by the Pauli matrices.

So yes, this list is rather a mess. It turns out that as we add more powers of 2, it just gets worse. It's just that 2-groups and p-groups in general have terrible structure.

1.4.10 Classification of Finitely Generated Abelian Groups

So we gave up on classifying all groups of order 16, but we can classify the abelian ones.

Theorem 1.105 (Classification of finitely generated abelian groups)**.** Any finitely generated abelian group is a product of cyclic groups.

Remark 1.106. This is not unique because, for example, $\mathbb{Z}/6\mathbb{Z} \cong \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. However, we can make this unique by forcing

$$G \cong \bigoplus_{k=1}^{N} \mathbb{Z}/n_k\mathbb{Z}$$

with $n_1 \mid n_2 \mid \cdots \mid n_N$ or by forcing the $n_\bullet$ to be prime-powers. Either of these gives us uniqueness, though using prime powers is only unique up to ordering the prime powers.

Example 1.107. We can classify all groups of order p^5. This comes down to writing down all the permutations of 5, which are

$$\begin{gathered} 5, \qquad 4+1, \qquad 3+2, \qquad 3+1+1, \\ 2+2+1, \qquad 2+1+1+1+1, \qquad 1+1+1+1+1+1. \end{gathered}$$

Each partition gives us a group as follows.

$$\begin{gathered} \mathbb{Z}/p^5\mathbb{Z}, \qquad \mathbb{Z}/p^4\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}, \qquad \mathbb{Z}/p^3\mathbb{Z} \times \mathbb{Z}/p^2\mathbb{Z}, \qquad \mathbb{Z}/3\mathbb{Z} \times (\mathbb{Z}/p\mathbb{Z})^2, \\ \left(\mathbb{Z}/p^2\mathbb{Z}\right)^2 \times \mathbb{Z}/p\mathbb{Z}, \qquad \mathbb{Z}/p^2\mathbb{Z} \times (\mathbb{Z}/p\mathbb{Z})^3, \qquad (\mathbb{Z}/p\mathbb{Z})^5. \end{gathered}$$

Next lecture we will prove Theorem 1.105.

1.5 September 9

You feel like you're going to have a bad time.

1.5.1 Groups of Order 2^n

Last lecture we noticed that groups of order 16 were rather a mess. In general, it turns out that groups of order higher powers of 2 are even worse.

n	number of groups of order 2^n
4	14
6	267
10	49487365422

It turns out that the number of groups of order p^n is a roughly $p^{(2/27)n^3}$, which is frankly huge; not even the $n = 10$ case fully captures the enormity of having a cube in an exponential. There's an entire book for groups of order 2^n for $n \le 6$.

Remark 1.108. It turns out that the vast majority of groups of order less than some bound are going to be 2-groups; see this MathExchange thread. The next most common are groups of order $3 \cdot 2^n$, then $5 \cdot 2^n$. In general, classifying these is quite boring.

1.5.2 Classification of Finitely Generated Abelian Groups

Today we'll prove Theorem 1.105. Recall the statement.

Theorem 1.109 (Classification of finitely generated abelian groups)**.** Any finitely generated abelian group is a product of cyclic groups.

Proof. Fix our group G, and fix generators $\{g_1, \dots, g_m\}$. We will write the group operation of G additively. There might be a list of relations among these elements; we list all relations, which gives us a large system of equations

$$\begin{cases} a_{11}g_1 + \cdots + a_{1m}g_m = 0 \\ a_{21}g_1 + \cdots + a_{2m}g_m = 0 \\ \vdots \end{cases}$$

We will abbreviate this system to the (unaugmented) matrix

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ a_{21} & \cdots & a_{2m} \\ \vdots & \ddots & \vdots \end{bmatrix}$$

We would like to simplify this to be diagonal; more precisely, because the above matrix need not be square (in fact, it might have countably infinite height), we want nonzero elements off the diagonal.

So, roughly speaking, we want to row-reduce. Here are our row operations; these correspond to moving around our relations.

- We can swap rows. Effectively, swapping row k with row ℓ turns the system

$$\begin{cases} a_{11}g_1 + \cdots + a_{1m}g_m = 0 \\ a_{21}g_1 + \cdots + a_{2m}g_m = 0 \\ \vdots \\ a_{k1}g_1 + \cdots + a_{km}g_m = 0 \\ \vdots \\ a_{\ell 1}g_1 + \cdots + a_{\ell m}g_m = 0 \\ \vdots \end{cases}$$

into

$$\begin{cases} a_{11}g_1 + \cdots + a_{1m}g_m = 0 \\ a_{21}g_1 + \cdots + a_{2m}g_m = 0 \\ \vdots \\ a_{\ell 1}g_1 + \cdots + a_{\ell m}g_m = 0 \\ \vdots \\ a_{k1}g_1 + \cdots + a_{km}g_m = 0 \\ \vdots \end{cases}$$

Merely rearranging the relations does not change the structure of the group.

- We can negate a row. Because negation of a row is an involution, this doesn't change the underlying structure.
- We can add two rows. Adding row k to row ℓ, we see that we are essentially saying that the system

$$\begin{cases} a_{11}g_1 + \cdots + a_{1m}g_m = 0 \\ a_{21}g_1 + \cdots + a_{2m}g_m = 0 \\ \vdots \\ a_{k1}g_1 + \cdots + a_{km}g_m = 0 \\ \vdots \\ a_{\ell 1}g_1 + \cdots + a_{\ell m}g_m = 0 \\ \vdots \end{cases}$$

implies

$$\begin{cases} a_{11}g_1 + \cdots + a_{1m}g_m = 0 \\ a_{21}g_1 + \cdots + a_{2m}g_m = 0 \\ \vdots \\ a_{k1}g_1 + \cdots + a_{km}g_m = 0 \\ \vdots \\ (a_{\ell 1} + a_{k1})g_1 + \cdots + (a_{km} + a_{\ell m})g_m = 0 \\ \vdots \end{cases}$$

which is true. Also, the converse (the second system implies the first) holds by subtraction, so these do yield the same group.

- By induction, we can actually add any integer multiple of a row to another row.

Here are our column operations; these correspond to moving around our generators.

- We can swap columns. Effectively, swapping column k with column ℓ turns the system

$$\begin{cases} a_{11}g_1 + \cdots + a_{1k}g_k + \cdots + a_{1\ell}g_\ell + \cdots + a_{1m}g_m = 0 \\ a_{21}g_1 + \cdots + a_{2k}g_k + \cdots + a_{2\ell}g_\ell + \cdots + a_{2m}g_m = 0 \\ \vdots \end{cases}$$

into the system

$$\begin{cases} a_{11}g_1 + \cdots + a_{1\ell}g_k + \cdots + a_{1k}g_\ell + \cdots + a_{1m}g_m = 0 \\ a_{21}g_1 + \cdots + a_{2\ell}g_k + \cdots + a_{2k}g_\ell + \cdots + a_{2m}g_m = 0 \\ \vdots \end{cases}$$

We note that this is the same as taking $(g_k, g_\ell) \mapsto (g_\ell, g_k)$, and rearranging the generators does not alter the structure of the group.

- We can negate a column, say k. Effectively, this turns the system

$$\begin{cases} a_{11}g_1 + \cdots + a_{1k}g_k + \cdots + a_{1m}g_m = 0 \\ a_{21}g_1 + \cdots + a_{2k}g_k + \cdots + a_{2m}g_m = 0 \\ \qquad\qquad\qquad\qquad \vdots \end{cases}$$

into

$$\begin{cases} a_{11}g_1 + \cdots + a_{1k}(-g_k) + \cdots + a_{1m}g_m = 0 \\ a_{21}g_1 + \cdots + a_{2k}(-g_k) + \cdots + a_{2m}g_m = 0 \\ \qquad\qquad\qquad\qquad \vdots \end{cases}$$

Because inversion is an involution, we see that that the exchange of generators $g_k \mapsto -g_k$ does not change the group structure.

- We can add two columns, say k to ℓ. Effectively, this turns the system

$$\begin{cases} a_{11}g_1 + \cdots + a_{1k}g_k + \cdots + a_{1\ell}g_\ell + \cdots + a_{1m}g_m = 0 \\ a_{21}g_1 + \cdots + a_{2k}g_k + \cdots + a_{2\ell}g_\ell + \cdots + a_{2m}g_m = 0 \\ \qquad\qquad\qquad\qquad \vdots \end{cases}$$

to

$$\begin{cases} a_{11}g_1 + \cdots + a_{1k}(g_k - g_\ell) + \cdots + (a_{1\ell} + a_{1k})g_\ell + \cdots + a_{1m}g_m = 0 \\ a_{21}g_1 + \cdots + a_{2k}(g_k - g_\ell) + \cdots + (a_{2\ell} + a_{2k})g_\ell + \cdots + a_{2m}g_m = 0 \\ \qquad\qquad\qquad\qquad \vdots \end{cases}$$

So we have taken the generator g_k to $g_k - g_\ell$, which is a reversible process and hence does not actually change the group structure. (We could construct an isomorphism if we wanted.

- By induction, we can actually add any integer multiple of a row to another row.

To "row-reduce," we do row and column operations. Here are the steps.

1. Consider the smallest we can make a_{11} by applying row and column operations while keeping a_{11} nonnegative. We have two cases.
   - If $a_{11} = 0$, then we can apply operations to make the entire matrix vanish, so $G \cong \mathbb{Z}^n$. Indeed, if there is a nonnegative entry anywhere, then we can swap that entry to a_{11}.
   - Otherwise $a_{11} > 0$. Currently our matrix looks like the following.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix}$$

   We claim that $a_{11} \mid a_{k1}$ for each k. Indeed, if $a_{11} \nmid a_{k1}$, then we can write $a_{k1} = qa_{11} + r$ for some $r < a_{11}$, so subtracting q times the kth row from the first makes a_{11} smaller. So subtracting a_{k1}/a_{11} times the first row from the kth row gives the matrix

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ 0 & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix},$$

   where the first column is all 0. Similarly,
   The same division algorithm argument shows that $a_{11} \mid a_{1k}$ for each k. So subtracting a_{1k}/a_{11} times the first column from the kth column gives the matrix

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix}.$$

2. Then we simply repeat the process to the smaller matrix

$$\begin{bmatrix} a_{22} & \cdots & a_{2m} \\ \vdots & \ddots & \vdots \end{bmatrix}$$

inside of our larger matrices. Note that applying row and column operations (which are swaps, negations, or additions) will not affect the 0s surrounding this sub-matrix.

So again, making a_{22} as small as possible and repeat the previous step lets us assert a matrix of the form

$$\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & a_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{bmatrix}$$

and inductively continue

Once we are done with this process, we get the matrix of relations where all terms off the diagonal are 0. This looks like the system

$$\begin{cases} a_{11}g_1 + 0g_2 + \cdots + 0g_k + \cdots + 0g_m = 0 \\ 0g_1 + a_{12}g_2 + \cdots + 0g_k + \cdots + 0g_m = 0 \\ \qquad\qquad\qquad\qquad \vdots \\ 0g_1 + 0g_2 + \cdots + a_{kk}g_k + \cdots + 0g_m = 0 \\ \qquad\qquad\qquad\qquad \vdots \\ 0g_1 + 0g_2 + \cdots + 0g_k + \cdots + a_{mm}g_m = 0 \end{cases}$$

It follows that each generator g_k has the sole relation $a_k g_k = 0$ (with possibly $a_k = 0$), so $g_k \mapsto 1$ yields an isomorphism $\langle g_k \rangle \to \mathbb{Z}/a_k\mathbb{Z}$

$$G \cong (\mathbb{Z}/a_1\mathbb{Z}) \times (\mathbb{Z}/a_2\mathbb{Z}) \times \cdots .$$

■

Remark 1.110. We can actually guarantee that $a_{11} \mid a_{22} \mid \cdots$. Indeed, otherwise we could use our row reduction to apply the division algorithm dividing a_{22} by a_{11}, thus making a_{11} smaller.

1.5.3 Groups of Order 17 and 18

Groups of order 17 are cyclic because 17 is prime.

Let's talk about groups G of order 18. We see G they have a subgroup H_9 of order 9 by Sylow, which must be normal, so $G \cong H_9 \rtimes \mathbb{Z}/2\mathbb{Z}$, where the $\mathbb{Z}/2\mathbb{Z}$ appears because it is our Sylow 2-subgroup. We have the following cases.

- If $H_9 = \mathbb{Z}/9\mathbb{Z}$, we see that $\mathbb{Z}/2\mathbb{Z}$ only has the trivial action or the inversion action on $\mathbb{Z}/9\mathbb{Z}$.
- For $H_9 = (\mathbb{Z}/3\mathbb{Z})^2$, the trick is to view $(\mathbb{Z}/3\mathbb{Z})^2$ as a vector space over $\mathbb{F}_3$ of dimension 3. In particular, we are looking for maps $\mathbb{Z}/2\mathbb{Z} \to \mathrm{GL}_2(\mathbb{F}_3)$, which aside from the trivial map correspond to order-2 elements of $\mathrm{GL}_2(\mathbb{F}_3)$. We can now do this by hand.

One of these groups turns out to be more interesting. Namely, there is a $\mathbb{Z}/2\mathbb{Z}$-action on $(\mathbb{Z}/3\mathbb{Z})^2$ by switching the two copies of $\mathbb{Z}/3\mathbb{Z}$; this corresponds to the order-2 matrix

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \in \mathrm{GL}_2(\mathbb{F}_3).$$

In some sense, what is happening is that we have a nontrivial $\mathbb{Z}/2\mathbb{Z}$-action on $\mathbb{Z}/2\mathbb{Z}$-indexed sequences of $\mathbb{Z}/3\mathbb{Z}$. We can generalize this construction.

Definition 1.111 (Wreath products)**.** Pick up two groups G and H and a set Ω with an H-action. (By default, we take $\Omega = H$.) Then we define

$$\operatorname{Mor}(\Omega, G) = \{f : \Omega \to G\},$$

which is a group by with (say) pointwise operation: $(fg)(x) = f(x)g(x)$. This has an H-action defined by

$$h \cdot f(x) = f(hx)$$

for $h \in H$, $f \in \operatorname{Mor}(\Omega, G)$, and $x \in \Omega$. So we define the *wreath product* $G \wr_\Omega H \coloneqq \operatorname{Mor}(\Omega, G) \rtimes H$.

At a high level, what is happening is that we have a list of symmetries of G (indexed by Ω), but this list itself has symmetries we want to keep track of (which is the H-action on Ω). A perhaps more concrete way to look at $\operatorname{Mor}(\Omega, G)$ is as sequences $\{g_\omega\}_{\omega\in\Omega}$ in G indexed by Ω. Here, the group operation is component-wise, and the H-action on Ω essentially induces a rearranging of the sequence.

Another quick fact that can we see straight from the definition is that

$$\#(G \wr_\Omega H) = \#\big(\operatorname{Mor}(\Omega, G) \rtimes H\big) = \#\operatorname{Mor}(\Omega, G) \cdot \#H = \#G^{\#\Omega} \cdot \#H.$$

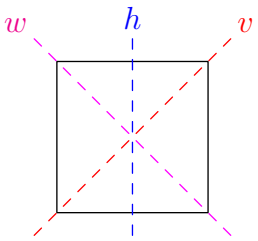
Anyways, let's do some examples.

Proposition 1.112. We have $\mathbb{Z}/2\mathbb{Z} \wr \mathbb{Z}/2\mathbb{Z} \cong D_8$.

Proof. For concreteness, we fix $G = H = \mathbb{Z}/2\mathbb{Z}$ so that we are computing $G \wr H$. To say that D_8 is the semidirect product of $\operatorname{Mor}(H, G)$ and H is to say that D_8 can fit in the short exact sequence

$$1 \to \operatorname{Mor}(H, G) \to D_8 \to H \to 1,$$

where $\operatorname{Mor}(H, G)$ is normal in D_8, $D_8 \to H$ has a pull-back into D_8, and we also have a prescribed conjugation action of H on $\operatorname{Mor}(H, G)$. We construct these manually from D_8 as follows.



Each of v, w, h refer to the reflection over the prescribed line.

(a) We claim that we want $\operatorname{Mor}(H, G) \cong \langle v, w\rangle$.

Note that $\operatorname{Mor}(H, G)$ consists of $\mathbb{Z}/2\mathbb{Z}$-indexed sequences of $\mathbb{Z}/2\mathbb{Z}$, so these are effectively ordered pairs $(\mathbb{Z}/2\mathbb{Z})^2$ where the group law is component-wise. So to check that $\operatorname{Mor}(H, G) \cong \langle v, w\rangle$, it suffices to say that v and w both have order 2, as does vw (which is the $180°$ rotation), so indeed, $\operatorname{Mor}(H, G)$ is an $\mathbb{F}_2$-vector space with 4 elements.

We also note that $\langle v, w\rangle$ is normal in D_8 because it is index $8/4 = 2$.

(b) We claim that we want $H \cong \langle h\rangle$. These are isomorphic because h has order 2. We also see that $h \in D_8 \setminus \langle v, w\rangle$, so $h\langle v, w\rangle \neq \langle v, w\rangle$, meaning that we do indeed have the short exact sequence

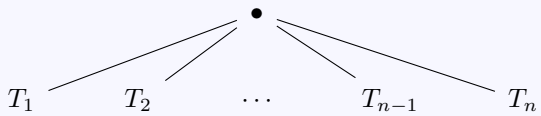
$$1 \to \langle v, w\rangle \to D_8 \to \langle h\rangle \to 1.$$

(c) Lastly, we need to check that the H-action on $\operatorname{Mor}(H, G)$ matches what it should be. Applying force, there are only two cases to check.

- We note that, given $\{a_0, b_1\} \in \operatorname{Mor}(H, G)$, $0 \cdot \{a_0, b_1\} = \{a_{0+0}, b_{1+0}\} = \{a_0, b_1\}$, so the action by $0 \in H$ is trivial; indeed, the action of e on $\langle v, w\rangle$ by conjugation is trivial.
- We note that, given $\{a_0, b_1\} \in \operatorname{Mor}(H, G)$, $1 \cdot \{a_0, b_1\} = \{a_{0+1}, b_{1+1}\} = \{b_0, a_1\}$, so the action by $0 \in H$ swaps; indeed, the action of h on $\langle v, w\rangle$ by conjugation swaps v and w, for $hvh^{-1} = w$ and $hwh^{-1} = v$. ■

Wreath products also show up naturally as symmetry groups of rooted trees. Here is the key lemma.

Proposition 1.113. Fix T_0 a rooted tree with symmetry group $\operatorname{Sym} T_0$. Then the symmetry group of the tree



made of a root and n copies of $T_0 = T_1 = T_2 = \cdots = T_n$ is $\operatorname{Sym} T_0 \wr_{\{1,\ldots,n\}} S_n$

Note that the wreath product is now actually taking in a named $\Omega = \{1, \ldots, n\}$ parameter. The action of S_n on Ω is by permuting, of course.

Proof. Name the big tree T and let $[n] \coloneqq \{1, \ldots, n\}$ for brevity. The main idea is that there are two steps to choose a symmetry T.

1. Pick a symmetry of each of the n copies of T_0. This more or less corresponds to an ordered sequence $\{\sigma_k\}_{k=1}^n$ of elements in $\operatorname{Sym}(T_0)$, which is the same thing as picking up an element of $\operatorname{Mor}([n], \operatorname{Sym} T_0)$.
2. Pick a way to rearrange the copies of T_0 itself. This corresponds to picking a permutation $\sigma \in S_n$.

These steps combine into something which is believably $\operatorname{Mor}([n], \operatorname{Sym} T_0) \rtimes S_n = \operatorname{Sym} T_0 \wr_{[n]} S_n$. We now rigorize this but not by too much because I don't hate myself. We want to build a split short exact sequence

$$1 \to \operatorname{Mor}([n], \operatorname{Sym} T_0) \to \operatorname{Sym} T \to S_n \to 1$$

with prescribed S_n-action on $\operatorname{Mor}([n], \operatorname{Sym} T_0)$.

(a) We claim that $N \coloneqq \{\sigma \in \operatorname{Sym} T : \sigma T_k \subseteq T_k \text{ for each } k\}$ is normal in $\operatorname{Sym} T$ and isomorphic to $\operatorname{Mor}([n], \operatorname{Sym} T_0)$.

- We show that N is normal in $\operatorname{Sym} T$. Indeed, for any $\tau \in \operatorname{Sym} T$ and $\sigma \in N$, then we have to check that
$$\left(\tau^{-1}\sigma\tau\right)(T_k),$$
for any of the subtrees T_k. Well, because the T_k are rooted trees, we see that τ must move the entire subtree T_k to some other tree T_ℓ wholesale. So any vertex $t \in T_k$ has $\tau t \in T_\ell$, so $\sigma\tau t \in T_\ell$, and $\tau^{-1}\sigma\tau t \in T_k$, finishing.
- We show that $N \cong \operatorname{Mor}([n], \operatorname{Sym} T_0)$. We construct our map $\varphi : N \to \operatorname{Mor}([n] \to \operatorname{Sym} T_0)$ by restriction:
$$\varphi(\sigma) \mapsto \{\sigma|_{T_k}\}_{k=1}^n.$$
This is homomorphic because look at it. It has an inverse map by taking $\{\sigma_k\}_{k=1}^n$ to the symmetry of $\operatorname{Sym} T$ which applies σ_k to T_k. It follows that φ is an isomorphism.

(b) We note that there is an embedding S_n into $\operatorname{Sym} T$ by sending $\sigma \in S_n$ to the permutation which merely permutes the $\{T_k\}_{k=1}^n$. We claim that we can set H to be the image of this permutation. (Technically, we have to fix a standard equality of the T_k to T_0 and then state that our elements of H do not alter this, but we will not bother.) We get $H \cong S_n$ for free.

We also see that there is a map
$$\operatorname{Sym} T \to S_n$$

by viewing $\sigma \in \operatorname{Sym} T$ as a permutation of the $\{T_k\}_{k=1}^n$. Here we again use the fact that a symmetry of T must send a subtree T_k to a T_ℓ wholesale.

We see that the image of H fully covers S_n because H describes all the ways we can rearrange the $\{T_k\}_{k=1}^n$. Further, we see that the kernel of this map consists of the maps which fix each tree in place, which is exactly N. So we indeed have the split short exact sequence

$$1 \to N \to \operatorname{Sym} T \to H \to 1.$$

In particular, we do have $\operatorname{Sym} T = \operatorname{Mor}([n], \operatorname{Sym} T_0) \rtimes S_n$.

(c) It remains to check that the H-action on N (by conjugation: $h \cdot x = h^{-1}xh$) matches the S_n-action on $\operatorname{Mor}([n], \operatorname{Sym} T_0)$. Fix $h \in H$ corresponding to $\sigma \in S_n$ and $g \in N$ corresponding to $\{g|_{T_k}\}_{k=1}^n \in \operatorname{Mor}([n], \operatorname{Sym} T_0)$. On one hand,

$$\sigma \cdot \{g|_{T_k}\}_{k=1}^n = \{g|_{T_{\sigma k}}\}_{k=1}^n. \tag{$*$}$$

On the other hand, for any $t \in T_k$ in any subtree T_k, we see $ht \in T_{\sigma k}$, so g will behave like $g|_{T_{\sigma k}}$ on ht, which then gets sent back to $g|_{T_{\sigma k}} t$ after another h^{-1}.

So indeed, $h^{-1}gh$ restricts to $g|_{T_{\sigma k}}$ on each T_k, which matches $(*)$. ■

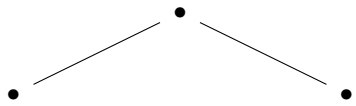
That was a lot of work, so here is a nice corollary.

Corollary 1.114. Fix T a complete binary rooted tree with $n+1$ levels for $n > 0$. Then

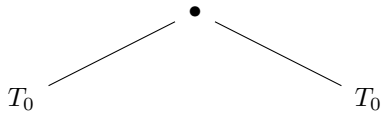
$$\operatorname{Sym} T \cong \underbrace{S_2 \wr S_2 \wr \cdots \wr S_2}_{n},$$

where $\wr$ is left-associative. It follows (by induction) that there are $2^{2^n - 1}$ total symmetries.

Proof. We induct. For $n = 1$, we have a complete binary tree with two levels, which looks like the following.



This has symmetry group S_2, which is our base case. For the inductive step, we fix T_0 the completed binary rooted tree with $n+1$ levels and construct the completed binary rooted tree with $n+2$ levels as follows.



By Proposition 1.113, we see that the symmetry group of the big tree is $\operatorname{Sym} T_0 \wr S_2$, which is what we wanted. ■

Remark 1.115. Technically we may permit the $n = 0$ as the base of our induction, which is the tree with only a root.

Here are some more miscellaneous examples of the wreath product.

Example 1.116. In electrodynamics, it turns out that the symmetry group is also a wreath product. Space-time acts as $\mathbb{R}^4$, and the Gauge group is S^1. The symmetry group consists of (smooth) functions $\mathbb{R}^4 \to S^1$, on which the Poincare group acts.

Example 1.117. We can also have the group of symmetries of an n-dimensional cube. Fix its vertices are $(\pm 1, \pm 1, \cdots)$. The symmetries are various inversions and permutations of coordinates, so our group of symmetries is $(\mathbb{Z}/2\mathbb{Z})^n \wr_{[n]} S_n$ using similar logic as in Proposition 1.113.

1.5.4 Groups of Order 24

We're going to skip over groups of order 19, 20, 21, 22, and 23, and we're not even going to fully classify groups of order 24. But let's sketch this; fix G of order 24.

1. If there is a normal Sylow 3-subgroup, then this is a semidirect product. There are many possibilities here for what is acting on our Sylow 3-subgroup.
2. Otherwise, the number of Sylow 3-subgroup is $1 \pmod 3$ and divides 24 and hence must divide 8 and hence must be 4. The trick is for G to act by conjugation of G on its Sylow 4-subgroups, which gives a homomorphism
$$G \to S_4.$$
What is the kernel? Well, it has order dividing 24, and in fact it has order $1, 2, 3$, or 6 because G acts transitively on the Sylow 3-subgroups.
   (a) If the order is 1, we get $G \cong S_4$.
   (b) If the order is 2, we get the binary tetrahedral group. Namely, we can realize A_4 as the group of rotations of a tetrahedron, which we can pull back along the short exact sequence
$$1 \to \{\pm 1\} \to S^3 \to \mathrm{SO}_3(\mathbb{R}) \to 1.$$
We can work out the other cases if we want, but we won't here.

1.5.5 Symmetric Groups

While we're here, let's use this as a discussion to talk about symmetric groups. Recall the following definition.

Definition 1.118 (Symmetric group)**.** The symmetric group S_n consists of the permutations of $\{1, \dots, n\}$.

To talk about the conjugacy classes, we note that we can write any permutation as a product of cycles by tracking the orbits of single elements. It turns out that the structure we need is the notion of the "cycle shape."

Definition 1.119 (Cycle shape)**.** Fix $\sigma \in S_n$. We say that σ has "cycle shape
$$1^{n_1} 2^{n_2} 3^{n_3} \cdots$$
if and only if the cycle decomposition of σ has exactly n_k k-cycles. For example, note that $n_k = 0$ for $k > n$ and that $\sum_{k=1}^{n} n_k k = n$. We will not show that cycle shape is well-defined.

We have the following proposition.

Proposition 1.120. Any two permutations σ and τ of S_n are conjugate if and only if they have the same cycle shape.

Proof. In one direction, suppose that σ and τ have the same cycle shape. We'll give the proof idea: conjugation "renames" the elements that σ and τ are acting on. This is clearer with an example.

Example 1.121. Take $\sigma = (123)(45)(6)$ and $\tau = (425)(16)(3)$. The point is that, for any $x \in S_n$,

$$x(123)(45)(6)x^{-1} = x(123)x^{-1} \cdot x(45)x^{-1} \cdot x(6)x^{-1} = (x1, x2, x3)(x4, x5)(x6).$$

So we can make this equal to τ by setting $x1 = 4, x2 = 2, x3 = 4, x4 = 1, x5 = 6$, and $x3 = 6$. Visually, we see that x is the vertical map in the following diagram.

$$\begin{array}{cccccc} (1 & 2 & 3) & (4 & 5) & (6) \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ (4 & 2 & 5) & (1 & 6) & (3) \end{array}$$

This idea generalizes into the proof. Indeed, the main trick is the following lemma.

Lemma 1.122. Fix a cycle $(a_1, a_2, \ldots, a_k) \in S_n$ and some $\sigma \in S_n$. Then

$$\sigma(a_1, a_2, \ldots, a_k)\sigma^{-1} = (\sigma a_1, \sigma a_2, \ldots, \sigma a_k).$$

Proof. This is by brute force. The main thing to check is that $\left(\sigma(a_1, a_2, \ldots, a_k)\sigma^{-1}\right)(\sigma a_\ell) = \sigma a_{\ell+1}$. ∎

Now dissolve our given permutations σ and τ into a cycle decompositions

$$\sigma = \prod_{k=1}^{n}\prod_{\ell=1}^{n_k}(a_{k,\ell 1}, a_{k,\ell 2}, \ldots, a_{k,\ell k}) \qquad \text{and} \qquad \prod_{k=1}^{n}\prod_{\ell=1}^{n_k}(k, b_{\ell 1}, b_{k,\ell 2}, \ldots, b_{k,\ell k}).$$

(Here we have organized the cycle decomposition by cycle length.) Our conjugating element x takes $a_{k,ij}$ to $b_{k,ij}$; because all elements of $\{1, \ldots, n\}$ appear in the $a_{k,ij}$ and $b_{k,ij}$, this x is surjective map $\{1, \ldots, n\} \to \{1, \ldots, n\}$, so $x \in S_n$. So we merely check

$$x\sigma x^{-1} = \prod_{k=1}^{n}\prod_{\ell=1}^{n_k} x(a_{k,\ell 1}, a_{k,\ell 2}, \ldots, a_{k,\ell k})x^{-1} = \prod_{k=1}^{n}\prod_{\ell=1}^{n_k}(b_{k,\ell 1}, b_{k,\ell 2}, \ldots, b_{k,\ell k}) = \tau,$$

which finishes this direction of the proof.

In the other direction, we show that all conjugates of σ have the same cycle shape. Well, fix the cycle shape of σ by

$$\sigma = \prod_{k=1}^{n}\prod_{\ell=1}^{n_k}(a_{k,\ell 1}, a_{k,\ell 2}, \ldots, a_{k,\ell k}).$$

Then, for any permutation $x \in S_n$, we can compute the conjugate

$$x\sigma x^{-1} = \prod_{k=1}^{n}\prod_{\ell=1}^{n_k} x(a_{k,\ell 1}, a_{k,\ell 2}, \ldots, a_{k,\ell k})x^{-1} = \prod_{k=1}^{n}\prod_{\ell=1}^{n_k}(xa_{k,\ell 1}, xa_{k,\ell 2}, \ldots, xa_{k,\ell k}).$$

Because x is injective, and the $a_{k,ij}$ appear exactly once in the original cycle decomposition, we see that the $xa_{k,ij}$ still make a valid cycle decomposition of $x\sigma x^{-1}$. However, then the given cycle decomposition forces the cycle shape of $x\sigma x^{-1}$ to match σ, finishing. ∎

We can also ask how many possible values of x there are which conjugate σ into τ. Essentially, we are having S_n act on the conjugacy class of σ by conjugation and asking how many elements take σ to τ; by Orbit-stabilizer logic, it suffices to count $\#\operatorname{Stab}\sigma$. The image is that we are roughly asking how many ways we can rewrite

$$\sigma = \underbrace{(*)\cdots(*)}_{n_1 \text{ 1-cycles}}\underbrace{(**)\cdots(**)}_{n_2 \text{ 2-cycles}}\underbrace{(***)\cdots}_{\cdots}.$$

Formally, we have the terrible cycle decomposition

$$\sigma = \prod_{k=1}^{n} \prod_{\ell=1}^{n_k} (a_{k,\ell 1}, a_{k,\ell 2}, \dots, a_{k,\ell k}).$$

Well, for each $(a_{1,\ell 1})$ of these 1-cycles, there are $n_1!$ ways we can rearrange them. Then there are $n_2!$ ways to rearrange the 2-cycles, but then each individually 2-cycle has two ways to rearrange it internally, so we get $2^{n_2} n_2!$. Continuing, we get

$$\prod_{k=1}^{n} k^{n_k} n_k!$$

total permutations stabilizing σ. With this in mind, we can compute the number of elements of the conjugacy class of σ is

$$\frac{\#S_n}{\#\operatorname{Stab}(\sigma)} = \frac{n!}{\prod_{k=1}^{n} k^{n_k} n_k!}$$

by the Orbit-stabilizer theorem.

Example 1.123. Let's work this out for S_4. We get the following table.

cycle shape	centralizer of element	size of class
4	4	6
31	3	8
2^2	$2^2 \cdot 2! = 8$	3
$1^2 2$	$2! \cdot 2 = 4$	6
1^4	$4! = 24$	1

We can check the union of our conjugacy classes has 24 elements.

1.5.6 Solvability

We can also asking about the normal subgroups of S_n. Of course, there is $\{\mathrm{id}\}$ and S_n, but there is also the alternating group.

Definition 1.124 (Alternating)**.** Fix the determinant

$$\Delta = \prod_{1 \le k < \ell \le n} (x_\ell - x_k).$$

Then S_n acts on Δ by permuting the coordinates, and the orbit of Δ is $\{\pm\Delta\}$ because it can only add signs to each factor (and indeed, we can check by hand that transpositions do add signs). We define $A_n = \operatorname{Stab}(\Delta)$. Note that because the orbit is fully $\{\pm\Delta\}$, we have $\#A_n = \#S_n / \#\{\pm\Delta\} = n!/2$ by the Orbit-stabilizer theorem.

These are almost all the normal subgroups of S_n. We have the following.

Proposition 1.125. The normal subgroups of S_n are the following.

- 1.
- S_n.
- A_n.
- For $n = 4$, we also have the normal subgroup $\{\mathrm{id}, (12)(34), (13)(24), (14)(23)\}$. Namely, it just so happens that the set $\{(12)(34), (13)(24), (14)(23)\}$ is a conjugacy class in S_4 (those are all the elements with cycle type 2^2) so this subgroup is a union of conjugacy classes, by magic.

Proof. We omit this proof because it is actually nontrivially annoying to classify the normal subgroups of S_n. The key trick is that any normal nontrivial subgroup has an element $\sigma \in S_n \setminus \{e\}$; then a commutator $\tau\sigma\tau\sigma^{-1}$ for τ some transposition will force the normal subgroup to have a three-cycle, which forces the normal subgroup to contain A_n. ■

This shows that S_4 is solvable.

Definition 1.126 (Solvable)**.** A group G is *solvable* if we can construct an ascending sequence of normal subgroups

$$\{\mathrm{id}\} = G_0 \subseteq G_1 \subseteq G_2 \subseteq \cdots \subseteq G_n = G$$

such that each quotient G_{k+1}/G_k is cyclic.

Remark 1.127. We can weaken the quotient condition to abelian, but it doesn't matter that much.

Example 1.128. For S_4, we have

$$\{\mathrm{id}\} \subseteq \{(12)(34)\} \subseteq \{\mathrm{id}, (12)(34), (13)(24), (14)(23)\} \subseteq A_4 \subseteq S_4.$$

These quotients have order $2, 2, 3, 2$ respectively, so they are cyclic because their order is prime.

This notion of solvable will come up again in Galois theory; it "turns out" that S_5 is not solvable, and this has to do with non-solvability of the quintic by radicals. Perhaps stranger, the weird exception for $n = 4$ in Proposition 1.125 is why quartics are solvable by radicals.

Definition 1.129 (Simple)**.** A group is called *simple* if it has no proper, nontrivial, normal subgroups.

Example 1.130. Consider A_5, the set of rotations of the icosahedron, of which there are $3 \cdot 20 = 60$ elements by the Orbit-stabilizer theorem. (Each of the 20 vertex has 3 rotations fixing it.) Let's write out its conjugacy classes.

- We have id, which is a conjugacy class of size 1.
- Face symmetries: we can rotate a face by $2\pi/3$, of which there are 20 elements, of order 3.

  Note that rotating a face by $4\pi/3$ is the same as rotating its opposite face by $2\pi/3$, so we don't count this symmetry.
- Edge symmetries: we can rotate an edge by $\pi/2$, of which there are 30 elements, of order 2. However, flipping over an edge is the same as flipping over the opposite edge, so there are only 15 here.
- Vertex symmetries: we can rotate a vertex by $2\pi/5$, of which there are 12 elements, of order 5. We can also rotate a vertex by $4\pi/5$, of which there are 12 elements, of order 5.

  Note that rotating by $6\pi/5$ is rotating the opposite vertex by $4\pi/5$, so we don't count it; similarly, we don't count rotation by $8\pi/5$.

We can check that $1 + 20 + 15 + 12 + 12 = 60$, so these are all our conjugacy classes. Now suppose we have a normal subgroup. It must be a sum of the above conjugacy classes and have size dividing 60, but it turns out that the only ways to do this are to have size 1 or 60.

It follows that A_5 have no proper, nontrivial, normal subgroups, so A_5 is simple and not solvable.

Example 1.131. We also have that $\mathbb{Z}/p\mathbb{Z}$ is simple for p prime.

It turns out that all simple groups of order less than 60 are of the form $\mathbb{Z}/p\mathbb{Z}$ for p prime. The hard cases here are 48 or 56. The point is that A_5 is the first group we have some trouble understanding.

This gives us the following way to study groups. For any group G, we can find some maximal chain of normal subgroups

$$\{e\} \subseteq G_0 \subseteq G_1 \subseteq \cdots \subseteq G_n = G$$

such that G_k is normal in G_{k+1} and G_{k+1}/G_k is simple. So we have two problems.

1. Find all simple groups.
2. Find all ways to take the above chain and combine the simple groups into larger groups.

The second question is hopeless: for example, if we just have ten copies of $\mathbb{Z}/2\mathbb{Z}$, then we are back to trying to classify groups of order 2^{10} again, which is very sad. The first question does have an answer: there are 18 infinite families of simple groups and 26 some exceptions.

Remark 1.132. Nobody actually knows how long the proof of the classification of simple groups is. It's probably somewhere between ten or twenty thousand pages. It has not been computer verified because it's too long and hard.

1.5.7 Miscellaneous Group Theory

There are some interesting groups of order 120. There are three groups built from $\mathbb{Z}/2\mathbb{Z}$ and A_5.

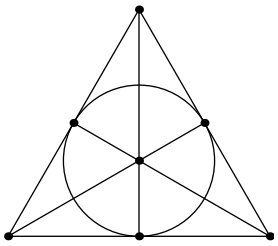
- We can take $\mathbb{Z}/2\mathbb{Z} \times A_5$.
- We can also take S_5 which has a normal subgroup A_5 and quotient $\mathbb{Z}/2\mathbb{Z}$.
- We also have the binary icosahedron group, created by pulling back A_5 as the symmetries of the icosahedron as a subgroup of $\mathrm{SO}_3(\mathbb{R})$ along

$$1 \to \{\pm 1\} \to S^3 \to \mathrm{SO}_3(\mathbb{R}) \to 1.$$

These groups are different by counting the number of elements which square to e; for example, the binary icosahedron group has exactly one element of order 2 from $\{\pm 1\}$, yielding 2 elements. In contrast, $\mathbb{Z}/2\mathbb{Z} \times A_5$ has $2 \cdot \binom{5}{4} \cdot 3 = 30$ such elements, and S_5 has more.

Remark 1.133. It turns out that binary icosahedron group G shows up in algebraic topology. It turns out that S^3/G is a three-dimensional manifold M with $\pi_1(M) = G$ and first homology group the maximal abelian group of G, which is trivial. This motivated the Poincare conjecture: if π_1 vanishes for a 3-manifold, then it must be a 3-sphere. (This is not true if the first homology group vanishes, as shown.)

After the cyclic groups and the alternating groups, the next simple group comes up as the symmetry group of the Fano plane, which is the following finite geometry. (Here, the unit circle is a "line.")



This group also turns out to be $\mathrm{GL}_3(\mathbb{F}_2)$ because the Fano plane is the the projective plane over $\mathbb{F}_2$. Alternatively, this group is $\mathrm{PSL}_2(\mathbb{F}_7) \cong \mathrm{SL}_2(\mathbb{F}_7)/\{\pm I\}$. There is no good reason why we should expect these groups to be isomorphic, but they are.

1.6 September 14

If you can't stand the heat, turn the A/C on.

1.6.1 Free Abelian Groups

This is going to be our last lecture on group theory, and it will be on free groups. A free groups on generators a, b, c is the "largest possible" group generated by those elements.

As an exercise, let's talk about free abelian groups. The free abelian group on generators $\{a_k\}_{k=1}^n$ can more or less be tracked by the sums

$$\sum_{k=1}^n m_k a_k$$

for integers $\{m_k\}_{k=1}^n$. Indeed, a group containing $\{a_k\}_{k=1}^n$ must have the above elements, and being abelian, we can always coerce a word in the above form. All of these elements are different, so we get the direct sum

$$\bigoplus_{k=1}^n \mathbb{Z} a_k \cong \mathbb{Z}^n,$$

where the isomorphism consists of coordinate-extraction. With this in mind, we take this as our definition of the free abelian group.

Definition 1.134 (Free abelian group)**.** Given letters $\{a_k\}_{k=1}^n$, we define the free abelian group F on the letters $\{a_k\}_{k=1}^n$ as the group

$$\bigoplus_{k=1}^n \mathbb{Z} a_k.$$

Let's prove some things; it turns out that free abelian groups are quite nice. We start with the universal property because it's nicer than the actual definition we gave.

Proposition 1.135 (Universal property of free abelian groups)**.** Fix F the free abelian group generated by $\{a_k\}_{k=1}^n$. Then, given an abelian group G with elements $\{g_k\}_{k=1}^n$, there is a unique group homomorphism $\varphi : F \to G$ such that $\varphi : a_k \mapsto g_k$ for each k.

Proof. On one hand, certainly if φ exists, then, for any $\sum_{k=1}^n m_k a_k \in F$, then we have

$$\varphi\left(\sum_{k=1}^n m_k a_k\right) = \sum_{k=1}^n m_k \varphi(a_k) = \sum_{k=1}^n m_k g_k,$$

so φ has only one option for where it can send all the elements.

In the other direction, we claim that

$$\varphi\left(\sum_{k=1}^n m_k a_k\right) \coloneqq \sum_{k=1}^n m_k g_k$$

actually defines a group homomorphism. This is well-defined because every element of F has a unique

representation as $\sum_{k=1}^{n} m_k a_k$ (by definition of the direct sum). This is homomorphic because we can sum

$$\begin{aligned}\varphi\left(\sum_{k=1}^{n} m_k a_k + \sum_{k=1}^{n} m_k' a_k\right) &= \varphi\left(\sum_{k=1}^{n} (m_k + m_k') a_k\right) \\ &= \sum_{k=1}^{n} (m_k + m_k') g_k \\ &= \sum_{k=1}^{n} m_k g_k + \sum_{k=1}^{n} m_k' g_k,\end{aligned}$$

where the last equality holds because G is abelian. (Note this is the only place where we used that G is abelian.) ■

This gives the following properties with ease.

Proposition 1.136. The following are true.

(a) The rank of a free abelian group is well-defined.

(b) Any subgroup of a free abelian group on n generators is free abelian on at most n generators.

Proof. We do these one at a time.

(a) Essentially, for $n_1 \neq n_2$, we want to show that $\mathbb{Z}^{n_1} \ncong \mathbb{Z}^{n_2}$. To show this, we look at the number of homomorphisms from $\mathbb{Z}^n \to \mathbb{Z}/2\mathbb{Z}$. Each homomorphism can be tracked by if it sends each generator to 1 or 0, so there are 2^n of these.

Explicitly, there is a unique homomorphism from the free abelian group on n letters

$$\varphi : \bigoplus_{k=1}^{n} \mathbb{Z}\alpha_k \to \mathbb{Z}/2\mathbb{Z}$$

for each function $f : \{\alpha_k\}_{k=1}^{n} \to \mathbb{Z}/2\mathbb{Z}$, by the universal property. Because each homomorphism also gives rise to a function f, we see that the number of homomorphisms is equal to the number of functions f, so there are 2^n total homomorphisms.

To finish, we see that the free abelian group on n_1 letters and the free abelian group on n_2 letters being isomorphic implies that the number of homomorphisms to $\mathbb{Z}/2\mathbb{Z}$ is equal, so $2^{n_1} = 2^{n_2}$, so $n_1 = n_2$.

(b) This proof is the same in spirit to the classification of finitely generated abelian groups.[5] It suffices to show that any subgroup $G \subseteq \mathbb{Z}^n$ is free abelian on at most n letters.

We first show that G is finitely generated by n vectors; this is by induction. The idea is to project onto the last coordinate, yielding the subgroup

$$G_n \coloneqq \{k_n : (k_1, k_2, \ldots, k_n) \in G\} \subseteq \mathbb{Z},$$

which must take the form $d_n\mathbb{Z}$ for some $d_n \in \mathbb{Z}$. Fixing $(d_1, \ldots, d_n) \in G_n$ its vector, then we can embed $G/(d_1, \ldots, d_n)\mathbb{Z}$ into $\mathbb{Z}^{n-1}$, so $G/(d_1, \ldots, d_n)$ can be generated by $n-1$ vectors by induction, so G can be generated by at most n vectors.

So suppose that G is generated by the row vectors of

$$\begin{bmatrix} d_{11} & \cdots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{nn} \end{bmatrix}.$$

[5] In fact, this property can be used to show the Classification of finitely generated abelian groups, but I am having trouble going the other direction.

We can check again that the row and column operations seen in the Classification of finitely generated abelian groups do not change the actual group structure (the check is identical, so we won't do it here), so we can reduce our matrix to one that looks like

$$\begin{bmatrix} m_1 & 0 & \cdots & 0 \\ 0 & m_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_n \end{bmatrix}$$

so that $G \cong \bigoplus_{k=1}^{n} m_k\mathbb{Z}$, finishing the proof. ■

1.6.2 Free Nonabelian Groups

So let's look at free nonabelian groups. We want the biggest possible group generated by the elements. Observe that it's not even obvious that such a thing exists!

Proposition 1.137. The free group F on $\{a_k\}_{k=1}^n$ exists and is a group.

Proof. Let S be the set of all possible words (empty allowed) whose letters are in $\{a_k\}_{k=1}^n$ or $\{a_k^{-1}\}_{k=1}^n$, and we simply mod this out by all relations which give the group axioms. For example, we should mod out by the relation that

$$(ab)(cc) = (a(bc))c$$

and all of its friends. To be explicit, we define the equivalence relation $\equiv$ on S defined as follows.

- Inverse: if we have $w = w_1\ell\ell^{-1}w_2$ for some words w_1, w_2 and letter $\ell \in \{a_k\}_{k=1}^N \cup \{a_k^{-1}\}_{k=1}^n$, then $w \equiv w_1w_2$.
- Well-defined concatenation: if $w_1 \equiv w_2$ and $v_1 \equiv v_2$, then $w_1v_1 \equiv w_2v_2$.

To be rigorous, we could do something like declare S a graph where the above two rules define edges; then $\equiv$ consists of equivalence classes of vertices, where two vertices are in the same equivalence if there is a finite path connecting them. We now check the group axioms by hand.

Remark 1.138. It is almost obvious, but it's not obvious that it's obvious.

- We make our group law concatenation. It is well-defined because our equivalence class forced it to be.
- Associativity: given w_1, w_2, w_3, then concatenating w_1w_2 with w_3 is the same as concatenating w_1 with w_2w_3.
- Identity: our identity is the empty string because concatenating the empty string does nothing.
- Inverse: given a string $w = \prod_{k=1}^{N} \ell_k$ for letters ℓ_k, the inverse law implies

$$\left(\prod_{k=1}^{N} \ell_k\right)\left(\prod_{k=1}^{N} \ell_{N+1-k}^{-1}\right)$$

is the empty string; formally we would do an induction here, but we won't bother. ■

Even though we have defined the free group as being equivalence classes of words, we will liberally call the elements of the free group "words" and refer to specific representatives.

In reality, the easiest way to handle the free group is by universal property.

Proposition 1.139 (Universal property of the free group)**.** Given a group G with elements $\{g_k\}_{k=1}^n$, the free group F on $\{a_k\}_{k=1}^n$ has a unique map $\varphi : F \to G$ such that $\varphi(a_k) = g_k$.

Proof. Again, the uniqueness of this map is the easier part: given a word $w = \prod_{k=1}^N a_k^{\varepsilon_k}$ for $\varepsilon_k \in \{\pm 1\}$, we see that φ being a homomorphism forces

$$\varphi(w) = \varphi\left(\prod_{k=1}^N a_k^{\varepsilon_k}\right) = \prod_{k=1}^N \varphi(a_k)^{\varepsilon_k} = \prod_{k=1}^N g_k^{\varepsilon_k},$$

so indeed, φ is forced. It remains to show that

$$\varphi\left(\prod_{k=1}^N a_k^{\varepsilon_k}\right) \coloneqq \prod_{k=1}^N g_k^{\varepsilon_k}$$

is actually a group homomorphism. Namely, we have to show that φ is well-defined and a homomorphism.

- We show that φ satisfies $\varphi(wv) = \varphi(w)\varphi(v)$ for words w and v, where now we are not treating w and v as equivalence classes but as actual words. Well, writing $w = \prod_{k=1}^N a_k^{\alpha_k}$ and $v = \prod_{k=N+1}^M a_k^{\alpha_k}$ (where we have continued our indexing implicitly), we see that

$$\varphi(wv) = \varphi\left(\prod_{k=1}^N a_k^{\alpha_k} \cdot \prod_{k=N+1}^M a_k^{\alpha_k}\right) = \varphi\left(\prod_{k=1}^M a_k^{\alpha_k}\right) = \prod_{k=1}^M g_k^{\alpha_k} = \prod_{k=1}^N g_k^{\alpha_k} \cdot \prod_{k=N+1}^M g_k^{\alpha_k} = \varphi(w)\varphi(v).$$

- We show that φ is well-defined. Because two elements are equal if and only if we can finitely apply the inverse and well-defined concatenation laws to make them term-wise equal, it suffices to show that φ is well-defined up to one application of each of these (and finish by induction).

  For the inverse law, we show that $\varphi(w_1 a_k^{\varepsilon} a_k^{-\varepsilon} w_2) = \varphi(w_1 w_2)$ for some letter a_k^{ε}. We know that φ satisfies the homomorphism property for words, so we may freely write

$$\varphi(w_1 a_k^{\varepsilon} a_k^{-\varepsilon} w_2) = \varphi(w_1) g_k^{\varepsilon} g_k^{-\varepsilon} \varphi(w_2) = \varphi(w_1)\varphi(w_2) = \varphi(w_1 w_2).$$

  For the well-defined concatenation law, we show that $w_1 \equiv w_2$ and $v_1 \equiv v_2$ implies that $\varphi(w_1 v_1) = \varphi(w_2 v_2)$. Well, by induction (say on the maximum word length among w_1, w_2, v_1, v_2), we may suppose that $\varphi(w_1) = \varphi(w_2)$ and $\varphi(v_1) = \varphi(v_2)$ so that

$$\varphi(w_1 v_1) = \varphi(w_1)\varphi(v_1) = \varphi(w_2)\varphi(v_2) = \varphi(w_2 v_2).$$

  This finishes. ∎

Note that the same proof as above works for other algebraic structures. So we can also define free rings, free algebras, and so on as the "universal object" by all possible words and modding out by all relations.

Warning 1.140. As a warning, there are no "free fields."

Here are two reasons why there are no "free fields."

- The core problem here is that the inverse function $a \mapsto a^{-1}$ is necessary but not defined everywhere, so fields aren't as nice as algebraic structures, in the sense that an algebraic structure should have operations defined everywhere with some relations. This makes "all possible words" in the above argument somewhat difficult.
- We can actually prove that there is no free functor for Fld. In short, all morphisms are injective, so if our free object on n elements is to have a map into $\mathbb{F}_2$, then our free object must inject into $\mathbb{F}_2$ and be $\mathbb{F}_2$. But then there is no map $\mathbb{F}_2$ into $\mathbb{F}_3$.

So free abelian groups are nicely behaved. Let's move on to the nonabelian case.

1.6.3 Reduced Words

Free groups are pretty huge; what do they look like? For example, is it nontrivial? This is not immediately obvious because our construction was complicated in the sense that there were a lot of equivalence relations. So let's try to bound the size of our group.

To upper-bound the size of our group, we note that every element is a word in the letters $a_\bullet$ and $a_\bullet^{-1}$, but this is somewhat inefficient because we can immediately cancel $a_1 a_1^{-1}$ and its friends. So we actually count with the following definition.

Definition 1.141 (Reduced words)**.** Let F be the free group on $\{a_k\}_{k=1}^n$. Then we define *reduced* words as words in F which do not contain $\ell\ell^{-1}$ for some letter $\ell \in \{a_k\}_{k=1}^n \cup \{a_k^{-1}\}_{k=1}^n$, which is our upper bound.

In particular, we see that every word in the free group has at least one representation as a reduced word simply by removing all $\ell\ell^{-1}$ substrings.

It feels like reduced words cannot collide, but it is nontrivial to prove this. Well, suppose that we have two reduced words w_1 and w_2 which are not equal term-wise so that we want to show $w_1 w_2^{-1} \neq e$. In other words, we have that $w_1 w_2^{-1}$ does not immediately reduce to the identity (w_1 and w_2 are not term-wise equal), and we want $w_1 w_2^{-1}$.

So it suffices to show the following lemma.

Lemma 1.142. Fix F the free group on $\{a_k\}_{k=1}^n$. Then all nontrivial reduced words w cannot collapse to e.

The idea here is to use the universal property. Let's give some examples of things that we can do, just to get the feeling for our power.

- Let's show that $a_k \neq e$. Well, we can map $a_k \mapsto 1$ in $\mathbb{Z}$ and $a_\ell \mapsto 0$ in $\mathbb{Z}$ so that $a_k \mapsto 1$, which is not the identity, so $a_k \neq e$.

- Let's show that $a_k \neq a_\ell$ for $k \neq \ell$. Well, we map $a_k \mapsto 1$ and $a_\ell \mapsto 0$ again, and the map any other generator $a_\bullet \mapsto 0$. Then we see that $a_k a_\ell \mapsto 1$ and is not the identity, so $a_k a_\ell^{-1} \neq e$.

In other words, we just showed that the map from our set of generators to the group is injective; I'm glad we got that squared away. We continue.

- We show that $a_1^2 a_2 a_1^{-1} a_2^{-1}$ is nontrivial. Well, send $a_1 \mapsto 1$ and $a_2 \mapsto 0$ in $\mathbb{Z}$ and our word gets sent to $1 \neq 0$.

- In general, if a word w has an unequal number of a_k and a_k^{-1} letters, then we can send $a_k \mapsto 1 \in \mathbb{Z}$ and all other generators to 0. Then w gets sent to some nontrivial integer.

In some sense, this last condition is the best we can do by mapping to abelian groups, for abelian groups will always send elements with an equal number of ℓ letters and ℓ^{-1} letters to the identity.

Now that we've gotten a feeling for the universal property, let's jump into the general case.

Proof of Lemma 1.142. Suppose that w is a word of length N; we map F into S_{N+1}. For concreteness, here is an example of our idea.

Example 1.143. Let's show that $aaba^{-1}b^{-1}a^{-1}$ is nontrivial in the free group on $\{a, b\}$. The idea is that we want to give permutations a and b which satisfy the following movement.

$$1 \xleftarrow[a]{} 2 \xleftarrow[b]{} 3 \xleftarrow[a]{} 4 \xrightarrow[b]{} 5 \xrightarrow[a]{} 6 \xrightarrow[a]{} 7$$

The point is that $aaba^{-1}b^{-1}a^{-1}$ will surely get sent to a nontrivial permutation now: $aaba^{-1}b^{-1}a^{-1}1 = 7$. Actually exhibiting a and b a matter of extending the constraints

$$1 \xleftarrow[a]{} 2 \qquad 3 \xleftarrow[a]{} 4 \qquad 5 \xrightarrow[a]{} 6 \xrightarrow[a]{} 7$$

to a permutation a and the constraints

$$1 \qquad 2 \xleftarrow[b]{} 3 \qquad 4 \xrightarrow[b]{} 5 \qquad 6 \qquad 7$$

to a permutation b. There are lots of ways to do this.

In general, fix our nontrivial reduced word $w = \prod_{k=1}^{N} a_k^{\varepsilon_k}$, where $\varepsilon_k \in \{\pm 1\}$. Then we would like to send a_k to a permutation so that we can have the following computation.

$$1 \xrightarrow{a_N^{\varepsilon_N}} 2 \xrightarrow{a_{N-1}^{\varepsilon_{N-1}}} 3 \longrightarrow \cdots \longrightarrow N-1 \xrightarrow{a_1^{\varepsilon_1}} N$$

(A forward arrow of a_k^{-1} is intended to mean a backward arrow for a_k.) Here, w will be sent to a nontrivial permutation, namely sending 1 to N. It remains to show that we can actually extend the above constraints to actual permutations.

There are some obstructions to extending our constraints. For example, if we every end up with the following constraints, we immediately violate injectivity and cannot be a permutation.

$$\bullet \longrightarrow \bullet \longleftarrow \bullet$$

Similarly, the following constraints cannot even make a function.

$$\bullet \longleftarrow \bullet \longrightarrow \bullet$$

However, these are the only obstructions to extending a permutation.[6] Further, neither of these obstructions in our constraints for a particular letter: having

$$\bullet \xrightarrow{a_k} \bullet \xleftarrow{a_k} \bullet$$

would mean that w contains $a_k a_k^{-1}$, violating that w is reduced. And having

$$\bullet \xleftarrow{a_k} \bullet \xrightarrow{a_k} \bullet$$

would mean that w contains $a_k^{-1} a_k$, again violating that w is reduced. So indeed, we can extend our constraints on $\{a_k\}_{k=1}^{n}$ to actual permutations on S_{n+1}, finishing. ■

So we see that all of our talk about reduced words has given us the following way to look at the free group.

Proposition 1.144. Elements of the free group are in bijection with reduced words.

[6] Showing this is annoying. Lacking the given obstructions, any constraint arrow $k \to k+1$ must either have no arrow into k or an arrow into k, but no arrow out of k; similar holds for $k+1$. Essentially this means that all of our constraints look like disjoint "chains" $x \to x+1 \cdots \to \cdots \to y-1 \to y$ (possibly backwards). We extend this to a permutation by fixing any element not in a chain and sending y to x.

Proof. This follows from the above discussion. ■

As a bonus, we get the following.

Definition 1.145 (Residually finite)**.** A group G is *residually finite* if, for each $g \in G \setminus \{e\}$, there exists a finite group H and a homomorphism $\varphi : G \to H$ such that $\varphi(g) \neq e$.

Proposition 1.146. If $g \in F$ the free group on $\{a_k\}_{k=1}^n$, then if $g \neq e$, then we can find some finite group H such that g does not go to $e \in H$.

Proof. Indeed, in the proof of Lemma 1.142, we showed that we can take $H = S_{n+1}$. ■

Non-Example 1.147. The rational numbers $\mathbb{Q}$ is residually finite: given any finite group H with a map $\varphi : \mathbb{Q} \to H$, we claim that φ is the trivial map. Indeed, for any element $\frac{n}{m}$, we claim that $\varphi(\frac{n}{m}) = e$, for

$$\varphi\left(\frac{n}{m}\right) = \varphi\left(\#H \cdot \frac{n}{m\#H}\right) = \varphi\left(\frac{n}{m\#H}\right)^{\#H} = e,$$

where the last equality is by Lagrange's theorem.

Remark 1.148. There is a terrible way to define the free group on reduced words by brute force defining the multiplication law on reduced words, simply cancelling out neighbors. This gets bad when trying to check associativity: for example, we have to keep track of how to associate

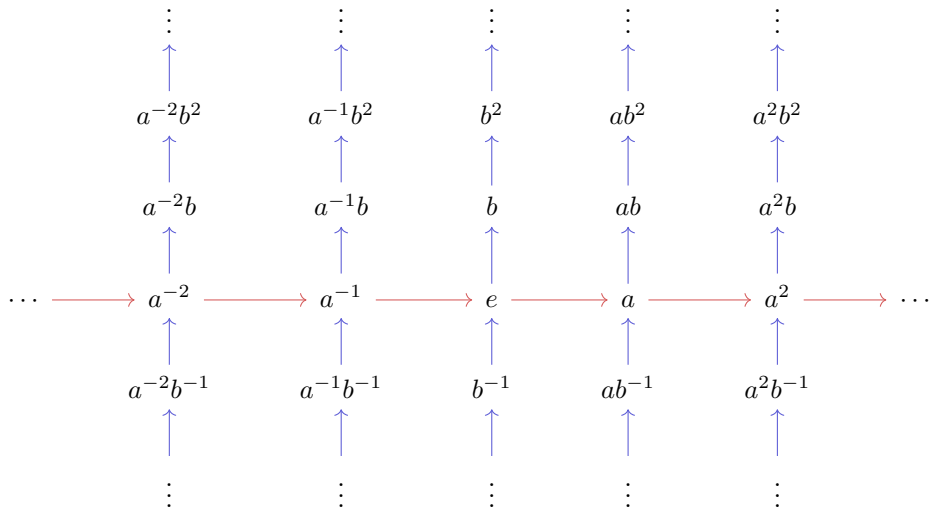
$$(ab)(b^{-1})(a^{-1}b).$$

While we're here, we present an alternate proof of Lemma 1.142. The idea is to find a set which G acts on and then show that every element acts nontrivially. We choose a Cayley graph; as a warning they are large graphs. For example, here is what our Cayley graph looks like for the free group F_2 generated by $\{a, b\}$.

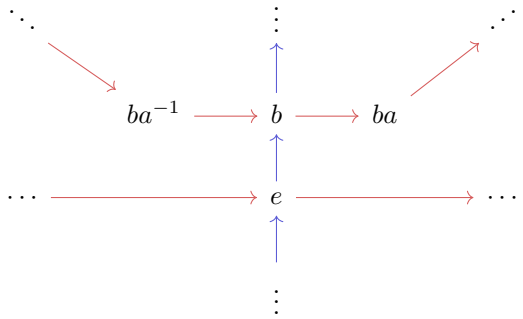
- We start by writing down vertices for each word consisting of only as, connecting $a^k \to a^{k+1}$ by a directed red edge, as follows.

$$\cdots \longrightarrow a^{-2} \longrightarrow a^{-1} \longrightarrow e \longrightarrow a \longrightarrow a^2 \longrightarrow \cdots$$

- Next, for each a^k, we add in vertices $a^k b^\ell$, where our edges are directed blue edges connecting $a^k b^\ell \to a^k b^{\ell+1}$, as follows.



- Then at each $a^k b^\ell$, we add in another red line using the same joining rules. This roughly looks like the following; to avoid crowding, we choose one small part.



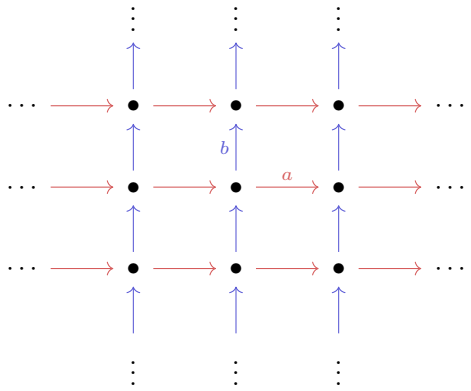
- Then we can continue adding layers to the above graph, building a giant monstrosity recursively.

With this example in mind, here is the general case.

Proof of Lemma 1.142. We build our graph X as with vertices which are reduced words and add a directed edge $w_1 \to w_2$ of color k if and only if $w_1 a_k = w_2$. In particular, an outgoing edge implies that the length of the word is strictly increasing; rigorously, we would build X recursively as in the example (to make sure X is a tree), but we will not bother here.

Then the action of $g \in G$ on X consists of sending vertices $v \in X$ to gv; it's not hard to check that this is in $\mathrm{Aut}(X)$, and in fact $g \neq e$ yields a nontrivial element of $\mathrm{Aut}(X)$ because the empty word is taken to g in X. ∎

This sort of process turns out to be easier for free abelian groups. For example, the Cayley graph for the free abelian group on $\{a, b\}$ looks like the following. (Add dimensions with more letters.)



So this free abelian group has Cayley graph which fits nicely in Euclidean space. On the other hand, the free group fits nicely on hyperbolic space, which itself does not fit nicely on Euclidean space.

Remark 1.149. In some sense, the free group is approximately the size of hyperbolic space in the same way that the free abelian group is approximately the size of Euclidean space.

1.6.4 Free Groups as Fundamental Groups

Let's talk about some properties of free groups, in the same way we talked about free abelian groups. For example, we still have the following.

Proposition 1.150. The rank of a free group is well-defined.

Proof. Again, given a free group on n letters named F, exactly the same argument as in Proposition 1.136 shows that there are 2^n homomorphisms $F \to \mathbb{Z}/2\mathbb{Z}$. It follows that two free groups are isomorphic if and only if they are generated by sets of the same cardinality. ■

However, rank does not behave the way we might want it to. For example, intuitively the free group on three elements ought to be larger than the free group on two elements, and indeed, for abelian groups, there is no injective homomorphism $\mathbb{Z}^3 \to \mathbb{Z}^2$.[7] But life is not so good with general free groups.

Proposition 1.151. There is an injective homomorphism from the free group on three letters to the free group on two letters.

Regardless, we will still be able to show the following, albeit with more effort.

Theorem 1.152. Any subgroup of a free group is free.

The main idea of this theorem is to show that G is free if and only if G is the fundamental group of some graph. So we begin by defining the fundamental group.

Definition 1.153 (Circuit)**.** Given a graph X, a *circuit* is an alternating sequence of vertices and adjacent edges of X, say $x_1e_1x_2e_2\dots e_{n-1}x_n$, such that $x_n = x_1$.

Warning 1.154. We are going to direct our edges but be sloppy about it: every edge $e : v \to w$ will have a designated inverse edge $e^{-1} : w \to v$. Because our circuits are also tracking the endpoints where we move, this does not matter most of the time, but it does matter for loops. The short version is that we need to keep track of "which" way we move along loops but no other edges, so I will not keep careful track of which way we move along non-loop edges (and so may write $e = e^{-1}$ when $v \neq w$).

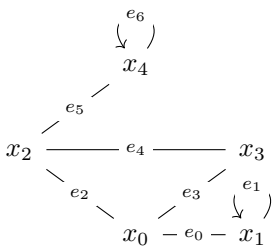
Definition 1.155 (Homotopy)**.** Fix a graph X and a basepoint $x_0 \in X$. We define the equivalence class $\equiv$ on circuits starting and ending at $x_0 \in X$ by asserting that, if $x_0x_1\cdots x_nx_0$ is a circuit, then

$$x_0e_0x_1e_1\cdots x_ne_nx_0 \equiv x_0e_0x_1e_1\cdots x_keye^{-1}x_ke_k\cdots x_n, \tag{$*$}$$

and closing this under the requirements to be an equivalence relation. (Any two paths with that can be translated into each other using a finite number of these moves are equivalent.) Here e^{-1} explicitly means moving backwards along the edge e.

If two circuits are equivalent under $\equiv$, then we say that they are *homotopic*. Also, for brevity, we will call a move of the form $(*)$ a "back-and-forth move."

Essentially, two paths are homotopic if there is some "back and forth" steps we can optimize out to make the paths equivalent to each other. As an example, consider the following graph X.



[7] The basis "vectors" of $\mathbb{Z}^3$ go to some three "vectors" in $\mathbb{Z}^2$, but any three vectors in $\mathbb{Z}^2 \subseteq \mathbb{Q}^2$ have a nontrivial $\mathbb{Q}$-linear relation for dimension reasons, which can be lifted to a nontrivial $\mathbb{Z}$-linear relation by clearing denominators.

In this graph, the circuits $x_0e_2x_2e_4x_3e_3x_0$ is homotopic with $x_0e_2x_2e_5x_4e_5x_2e_4x_3e_3x_0$ but is not homotopic with $x_0e_2x_2e_5x_4e_6x_4e_5x_2e_4x_3e_3x_0$.

Now, our fundamental group is more or less the same thing as seen in algebraic topology, but we will define this algebraically.

Definition 1.156 (Fundamental group)**.** Suppose that X is a connected graph, with loops permitted but not multiple edges, with a particular basepoint $x_0 \in X$. Then the *fundamental group* of (X, x_0), notated $\pi_1(X, x_0)$, consists of the circuits around x_0 up to homotopy.

Given a circuit C starting and ending at x_0, we will denote its homotopy equivalence class by $[C]$.

Note that we have not actually made our fundamental group into a group yet. Our composition law will be composition of circuits (follow the first circuit; then follow the second), but it might feel like we don't need homotopy for this. It turns out that homotopy is what makes this group law a group, giving us inverses.

Lemma 1.157. Fix X a connected graph and $x_0 \in X$ a basepoint. Then $\pi_1(X, x_0)$ is a group.

Proof. As promised, our group law is composition of circuits: write down the first circuit, subtract its last x_0 to avoid duplicates, and then write down the second circuit. We check the group conditions by hand.

- We show that the group law is well-defined: given circuits $C_1 \equiv C_2$ and $D_1 \equiv D_2$, we have to show that $C_1D_1 = C_2D_2$. Well, it requires finitely many back-and-forth moves to turn C_1 into C_2 and finitely many moves to turn D_1 into D_2, so it still requires finitely many moves to turn C_1D_1 into C_2D_2.
- Associative: given circuits C_1, C_2, C_3, we see that concatenating C_1 with C_2 first and then with C_3 is the same total string as concatenating C_2 with C_3 and then concatenating C_1 at the front.
- Identity: our identity is the do-nothing circuit "x_0." Concatenating with it does nothing.
- Inverse: reversing a circuit $C \coloneqq x_0e_0x_1e_1\cdots x_ne_nx_0$ to $C^{-1} \coloneqq x_0e_n^{-1}x_n\cdots e_1^{-1}x_1e_1^{-1}x_0$ gives its inverse. Indeed, concatenating gives

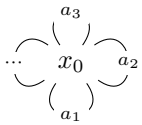
$$\begin{aligned} C \cdot C^{-1} &\equiv x_0e_0x_1e_1\cdots x_{n-1}e_{n-1}x_ne_nx_0e_n^{-1}x_ne_{n-1}^{-1}x_{n-1}\cdots e_1^{-1}x_1e_0^{-1}x_0 \\ &\equiv x_0e_0x_1e_1\cdots x_{n-1}e_{n-1}x_ne_{n-1}^{-1}x_{n-1}\cdots e_1^{-1}x_1e_0^{-1}x_0 \\ &\equiv \cdots \\ &\equiv x_0e_0x_1e_0^{-1}x_0 \\ &\equiv x_0, \end{aligned}$$

  which is our identity. ∎

We now begin moving towards our description of fundamental groups.

Lemma 1.158. The free group F on $\{a_k\}_{k=1}^n$ is a fundamental group.

Proof. The idea is to assign a_k to a loop around the basepoint so that reduced words roughly correspond to unique homotopy classes. Explicitly, we choose a graph X with one vertex x_0 and n different loops named $a_1, \ldots, a_k$. This looks like the following.



We would like to show that $\pi(X, x_0) \cong F$, where F is the free group on $\{a_k\}_{k=1}^n$. Because any circuit on X must have x_0 as its only vertex, so all of our circuits look like

$$x_0\ell_0x_0\ell_1\cdots x_n\ell_nx_0$$

for some letters $\ell_\bullet \in \{a_k\}_{k=1}^n \cup \{a_k^{-1}\}_{k=1}^n$. Now, the point is that we have a function φ from circuits to F by

$$\varphi : x_0\ell_0x_0\ell_1\cdots x_n\ell_nx_0 \longmapsto \ell_0\ell_1\cdots\ell_n.$$

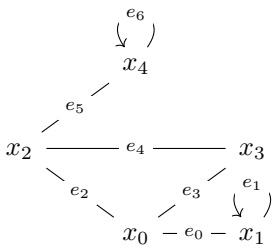
We can show that this is a group isomorphism $\pi_1(X, x_0) \to F$, which we do by hand.

- We can show that φ satisfies $\varphi(C_1C_2) = \varphi(C_1)\varphi(C_2)$ on circuits C_1, C_2. This follows directly from the fact C_1C_2 corresponds to concatenation, as does $\varphi(C_1)\varphi(C_2)$.
- The main obstruction is showing that φ is well-defined. By induction, it suffices to show that if two circuits differ by a single back-and-forth move give the same element of F. Well, by the previous part, we can take
$$\varphi(x_0\ell_0x_0\ell_1\cdots x_k\ell x_0\ell^{-1}x_k\ell_k\cdots x_0)$$
to
$$\varphi(x_0\ell_0x_0\ell_1\cdots x_k)\varphi(x_k\ell x_0\ell^{-1}x_k)\varphi(x_k\ell_k\cdots x_0),$$
which is indeed $\varphi(x_0\ell_0x_0\ell_1\cdots x_k)\varphi(x_k\ell_k\cdots x_0) = \varphi(x_1\ell_0\cdots x_n)$.
- Surjectivity is relatively apparent: pull a word of $w \in F$ directly backwards by punctuating it with x_0.
- Injectivity is also difficult. In short, we can reduce a circuit by removing any stray $\ell x_0\ell^{-1}$ terms by a back-and-forth move. Then nontrivial homotopy classes will map to nontrivial reduced words, so nontrivial homotopy classes are not in the kernel. So the kernel is trivial. ■

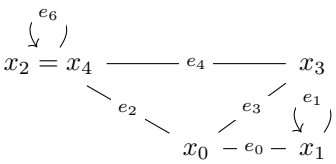
And here is the other direction.

Lemma 1.159. Any fundamental group of a finite, connected graph is a free group. In fact, given a finite, connected graph and basepoint (X, x_0) and a spanning tree T, then $\pi_1(X, x_0)$ is isomorphic to the free group on the edges of $X \setminus T$.

Proof. The idea is to contract edges by homotopy. For example, we can take



to



by contracting along the edge e_5. The main thing we need to show is that the fundamental group does not change after contracting along a non-loop edge.

Indeed, given a graph X with basepoint $x_0 \in X$, and non-loop edge e connecting v_1 and v_2, we construct the contracted graph X/e by deleting e and declaring $v_1 = v_2$. It remains to show $\pi_1(X, x_0) \cong \pi_1(X/e, x_0)$. We construct $\varphi : \pi_1(X, x_0) \to \pi_1(X/e, x_0)$ by taking a circuit

$$x_0e_0x_1e_1\cdots x_n$$

and deleting any occurrence of e while replacing each v_2 with a v_1. We need to show that φ is a homomorphism; this is generally annoying but visually makes sense, so we outline.

- Homomorphic on paths: we won't be rigorous about this. Essentially, concatenating two paths and then contracting the paths about e to some reduced word is the same as contracting the paths about e first and then concatenating.
- Well-defined: we can use the fact we are homomorphic on paths. Any back-and-forth move can be isolated from the rest of the circuit in the same way as in the proof of Lemma 1.158 so that φ is well-defined up to back-and-forth moves. Thus, φ is well-defined up to homotopy.
- Surjective: any path on X/e can be lifted to a path of X by, roughly speaking, just following the path in X/e in X. Any time we hit v_1 or v_2 in X/e, the next edge in X in the path might not adjacent to our current v_1 or v_2, but it will be adjacent to one of $\{v_1, v_2\}$, so we can use e to cross between with no repercussions from φ.
- Injective: essentially, it suffices to show that applying a back-and-forth move to a circuit in X/e does not affect its lift described to X described in the surjectivity. Because our lift essentially just follows the circuit in X/e with minor adjustments around e, a back-and-forth move will lift directly to a back-and-forth move, so the lift is well-defined up to homotopy.

To finish, we fix T a spanning tree of X. Recursively applying contraction along the edges of T will eventually[8] leave us with a single vertex and $\#E(X) - \#E(T)$ loops around our basepoint. So we see $\pi_1(X, x_0)$ is isomorphic to the free group on $\#E(X) - \#E(T)$ letters from Lemma 1.158. ■

Remark 1.160. Technically, we can extend the above argument to work for all connected graphs, but this requires more technical effort. Essentially, given a connected graph X and spanning tree T, we can mod X by the entire tree T in one blow. I am under the impression that the same arguments that work for a single edge generalize.

We now show our theorem.

Theorem 1.152. Any subgroup of a free group is free.

Proof of Theorem 1.152. Suppose that $G \subseteq F$ is a subgroup of a free group. Then we define the graph X whose points are the cosets in F/G and the edges are the actions of the generators of F. Note that F acts transitively on F/G, so X is connected; we choose eG as our basepoint.

We now claim that G is $\pi_1(X, eG)$, which will be sufficient because fundamental groups are free. Essentially, the idea is that we can map words $w = \prod_{k=1}^{N} \ell_k \in G$ to the circuit of X starting at eG and following ℓ_k as edges:

$$eG \xrightarrow{\ell_N} \ell_N eG \xrightarrow{\ell_{N-1}} \ell_{N-1}\ell_N eG \xrightarrow{\ell_{N-2}} \ell_{N-2}\ell_{N-1}\ell_N G \to \cdots.$$

Then we see that the last coset we hit in the circuit is wG, so the path we make is a circuit if and only if $w \in G$. We briefly talk through the checks to show we have an isomorphism. Call this map from words to paths φ.

- Well-defined: we know that every word can be reduced to a unique reduced representative by simply recursively removing $\ell\ell^{-1}$ subword, so it suffices to show that introducing an $\ell\ell^{-1}$ does not change the homotopy class of the output. But introducing $\ell\ell^{-1}$ means inserting

$$\cdots \to gG \xrightarrow{\ell^{-1}} \ell^{-1}gG \xrightarrow{\ell} \ell\ell^{-1}gG \to \cdots,$$

  where we see that this is just a back-and-forth move and therefore does nothing.
- Homomorphic: both group laws are concatenation, and we concatenate before or after.

[8] Technically, we have to show that the edges of the spanning tree never become loops when contracted. Well, contraction really just declares vertices equal, so the only way to have a loop would be to have a circuit in our spanning tree.

- Surjective: all of our circuits take the form

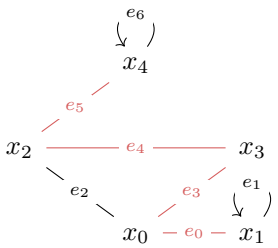
$$eG \xrightarrow{\ell_N} \ell_N eG \xrightarrow{\ell_{N-1}} \ell_{N-1}\ell_N eG \xrightarrow{\ell_{N-2}} \ell_{N-2}\ell_{N-1}\ell_N G \to \cdots \xrightarrow{\ell_1} \underbrace{\left(\prod_{k=1}^{N} \ell_k\right)}_{=:w} G$$

  for some word w, where $w \in G$ so that $wG = G$. It follows that $w \in G$ maps to this circuit.

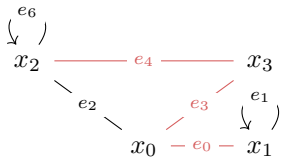
- Injective: in the well-defined point, we showed that back-and-forth moves correspond to removing $\ell\ell^{-1}$ substrings of words, so it follows that the inverse map introduced in the surjective point is also well-defined. ■

1.6.5 Applications of Fundamental Groups

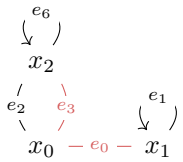
We are going to actually use this spanning tree contraction algorithm, so we give an example of this algorithm. We start with X as above with the designated red spanning tree.



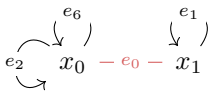
Contracting along e_5 gives X/e_5.



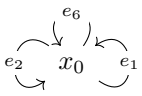
Contracting along e_4 gives $X/\{e_4, e_5\}$.



Contracting along e_3 gives $X/\{e_3, e_4, e_5\}$.



Lastly, contracting along e_0 gives $X/\{e_0, e_3, e_4, e_5\}$.



So we see that $\pi_1(X, x_0)$ is isomorphic to the free group on $\{x_0e_1x_0, x_0e_6x_0, x_0e_2x_0\}$, which were exactly the edges in X minus the spanning tree. We can even track this backwards to find generators of $\pi_1(X, x_0)$. We will omit the edges to prevent overcrowding.

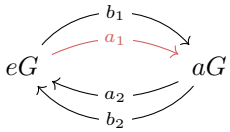
- We have that $\pi_1(X/\{e_0, e_3, e_4, e_5\}, x_0)$ is generated by $\{e_1, e_6, e_2\}$.
- We have that $\pi_1(X/\{e_3, e_4, e_5\}, x_0)$ is generated by $\{e_0 e_1 e_0^{-1}, e_6, e_2\}$ by lifting along e_0.
- We have that $\pi_1(X/\{e_4, e_5\}, x_0)$ is generated by $\{e_0 e_1^{-1} e_0^{-1}, e_3 e_2, e_3 e_6 e_3^{-1}\}$ by lifting along e_3.
- We have that $\pi_1(X/e_5, x_0)$ is generated by $\{e_0 e_1^{-1} e_0^{-1}, e_3 e_4 e_2, e_3 e_4 e_6 e_4^{-1} e_3^{-1}\}$ by lifting along e_4.
- We have that $\pi_1(X, x_0)$ is generated by $\{e_0 e_1^{-1} e_0^{-1}, e_3 e_4 e_2, e_3 e_4 e_5 e_6 e_5^{-1} e_4^{-1} e_3^{-1}\}$ by lifting along e_4.

Let's do a more sophisticated example of this algorithm. We show the following.

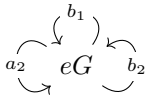
Proposition 1.151. There is an injective homomorphism from the free group on three letters to the free group on two letters.

Proof. We use the construction of graphs from the proof of Theorem 1.152. Fix F the free group generated by the two letters a and b, and we study subgroups G of index 2. Using the construction from Theorem 1.152, we make a graph with two vertices and edges dictated by the action of $\{a, b\}$ on these vertices.

The case we care about is when a and b swaps both of the vertices, as follows. Call this graph X; we label the a and b by subscripts for clarity.



We need to contract along something to make this a flower graph, so we contract along the red a_1, which gives the following graph X/a_1.



So we see that X/a_1 is freely generated by $\{a_2, b_1, b_2\}$, so pulling back along a_1 we have that X is freely generated by $\{a_1 a_2, b_1 a_1^{-1}, a_1 b_2\}$. In other words, we can check that the subgroup generated by $\{a^2, ba^{-1}, ab\}$ is free on those generators, finishing. ∎

Note that the above discussion can extend to classify all subgroups of index two of F. We simply have to do casework on what the possible (connected) graphs on two vertices are. We will not do this computation here.

1.7 September 16

You feel your sins crawling on your back.

1.7.1 Categories

We interrupt our regularly scheduled programming to talk about category theory. We're skipping most of the proofs.

Remark 1.161. All proofs in category theory are trivial and not very interesting.

We start some examples of categories.

Example 1.162. The basic examples of categories are as follows.

- The category of sets is Set. There are objects which are sets and morphisms which are functions.
- The category of groups is Grp. There are objects which are groups and morphisms which are homomorphisms.
- The category of topological spaces is Top. There are objects which are spaces and morphisms which are continuous maps.

So we have the following definition, abstracting the common features.

Definition 1.163 (Category)**.** A *category* $\mathcal{C}$ consists of the data of a set (or class) of objects with some morphisms. Namely, given objects A and B in the category $\mathcal{C}$, there is a set (or class) $\mathrm{Mor}(A, B)$ of morphisms from A to B satisfying the following.

(a) We can compose morphisms: given $f : \mathrm{Mor}(A, B)$ and $g : \mathrm{Mor}(B, C)$, there is a morphism $g \circ f : \mathrm{Nor}(A, C)$.

(b) There are identity morphisms: given any object A, there is a morphism $\mathrm{id}_A \in \mathrm{Mor}(A, A)$.

(c) Composition should associate: given objects A, B, C, D with morphisms $h \in \mathrm{Mor}(A, B)$ and $g \in \mathrm{Mor}(B, C)$ and $f \in \mathrm{Mor}(C, D)$, then $(f \circ g) \circ h = f \circ (g \circ h)$.

(d) Identity should be an identity: given a morphism $f \in \mathrm{Mor}(A, B)$, we have $f \circ \mathrm{id}_A = \mathrm{id}_B \circ f$.

Warning 1.164. I may occasionally abbreviate $f \in \mathrm{Mor}(A, B)$ to $f : A \to B$. For example, I might write composition as $\circ : \mathrm{Mor}(B, C) \times \mathrm{Mor}(A, B) \to \mathrm{Mor}(A, C)$ as $\circ : (B \to C) \times (A \to B) \to (A \to C)$.

Remark 1.165. We don't want to force all objects to live in a set because we would like Set to be a category, and there is no set containing all sets.

Let's have more examples.

Example 1.166. Groups are categories, where the object is a single point, where the morphisms are the elements of group, and we define the composition of two morphisms g, h to be $g \circ h \coloneqq gh$.

Example 1.167. Preordered sets X are categories. A preordered set is an ordered set where the order has reflexivity and transitivity. Our objects are elements of X, and we place exactly one morphism $A \to B$ if and only if $A \leq B$ for $A, B \in X$. Namely, the identity morphism comes from $A \leq A$, and associativity of composition comes from transitivity.

Of course, there are lots more things which are categories.

1.7.2 Fun with Morphisms

Here is the fundamental idea of category theory.

Idea 1.168. Ignore any internal structure of an object and instead look at its morphisms.

To be more explicit, given a category $\mathcal{C}$ with an object $A \in \mathcal{C}$, we study the morphisms of A instead of A directly.

For example, let's redefine injective. For sets, we usually say that $f : A \to B$ is injective if and only if $f(a_1) = f(a_2)$ for $a_1, a_2 \in A$, then $a_1 = a_2$. However, we would like to avoid talking about elements of A. Here is our new definition.

Definition 1.169 (Monic)**.** Fix a category $\mathcal{C}$ with a morphism $f \in \mathrm{Mor}(A, B)$. Then we say that f is *monic* (or "is a *monomorphism*") if and only if, for each other object X and morphisms $g_1, g_2 : \mathrm{Mor}(X, A)$, then $f \circ g_1 = f \circ g_2$ implies $g_1 = g_2$. Here is the diagram.

$$X \underset{g_2}{\overset{g_1}{\rightrightarrows}} A \xrightarrow{f} B$$

Think "monic means left-cancellative," which makes more sense after the example in Set.

Example 1.170. We check that monic is equivalent to injective in Set.

- Suppose that $f : A \hookrightarrow B$ is injective. Then, suppose we have an object X with maps $g_1, g_2 : X \to A$ such that $f \circ g_1 = g \circ g_2$. Then, for any $x \in X$, we have $f(g_1(x)) = f(g_2(x))$, from which $g_1(x) = g_2(x)$ follows by injectivity of f. So indeed, $g_1 = g_2$.
- Conversely, suppose that $f : A \hookrightarrow B$ is monic. Fix elements $a_1, a_2 \in A$ such that we want to show $f(a_1) = f(a_2)$ implies $a_1 = a_2$. Then consider the object $X := \{*\}$ for any a. There is a map $g_1 : X \to A$ by $* \mapsto a_1$ and a map $g_2 : X_2 \to A$ by $g_2 : * \mapsto a_2$. Then we have that $f \circ g_1 = f \circ g_2$ because
$$f(g_1(*)) = f(a_1) = f(a_2) = f(g_2(*)).$$
So because f is monic, it follows $g_1 = g_2$, so $a_1 = g_1(*) = g_2(*) = a_2$.

So in Set, monic means injective, but now monic works for all categories.

What about surjectivity? Well, in Set, we are saying that $f : A \to B$ has, for each $b \in B$, some $a \in A$ with $f(a) \in b$. How do we do this without talking about elements? The answer turns out to be dualize the definition of monic.

Definition 1.171 (Epic)**.** Fix a category $\mathcal{C}$ with a morphism $f \in \mathrm{Mor}(A, B)$. Then we say that f is *epic* (or "is an *epimorphism*") if and only if, for each other object X and morphisms $g_1, g_2 : \mathrm{Mor}(B, X)$, then $g_1 \circ f = g_2 \circ f$ implies $g_1 = g_2$. Here is the diagram.

$$A \xrightarrow{f} B \underset{g_2}{\overset{g_1}{\rightrightarrows}} X$$

Think "epic means right-cancellative."

Example 1.172. We check that epic is equivalent to surjective in Set.

- Suppose that $f : A \twoheadrightarrow B$ is surjective. Then, suppose that we have an object X with maps $g_1, g_2 : B \to X$ such that $g_1 \circ f = g_2 \circ f$ so that we want $g_1 = g_2$. Well, for any $b \in B$, there exists an $a \in A$ such that $f(a) = b$ because f is surjective, so
$$g_1(b) = g_1(f(a)) = (g_1 \circ f)(a) = (g_2 \circ f)(a) = g_2(b),$$
indeed implying that $g_1 = g_2$.
- Conversely, we show the contrapositive: if $f : A \to B$ is not surjective, then f is not epic. Indeed, consider the object $X = \{0, 1\}$ with maps $g_1, g_2 : B \to X$ defined by
$$g_1(b) \coloneqq \begin{cases} 1 & b \in \operatorname{im} f, \\ 0 & b \notin \operatorname{im} f, \end{cases} \qquad \text{and} \qquad g_2(b) = 1.$$
We note that, for any $a \in A$, we have $g_1(f(a)) = 1 = g_2(f(a))$, so $g_1 \circ f = g_2 \circ g$. But f not being surjective implies that there exists $b \in B \setminus \operatorname{im} f$ so that $g_1(b) \neq g_2(b)$. It follows f is not epic.

So again, in Set, this is equivalent to surjective, but now we can talk about epic in all categories.

Warning 1.173. Monic and epic do not always turn out to mean injective and surjective. For example, in the category Ring, the map canonical map $\mathbb{Z} \hookrightarrow \mathbb{Q}$ is epic but not surjective.

To be more explicit, of course $\iota : \mathbb{Z} \hookrightarrow \mathbb{Q}$ is not surjective, but showing it is epic needs some work. Essentially this is because $\mathbb{Q}$ is the fraction field of $\mathbb{Z}$, so what $\mathbb{Q}$ does is determined by what $\mathbb{Z}$ does. More rigorously, any good ring map $\mathbb{Z} \to R$ can be uniquely lifted to a map $\mathbb{Q} \to R$.

Indeed, suppose that we have some ring R with ring homomorphisms $g_1, g_2 : \mathbb{Q} \to R$ such that $g_1 \circ \iota = g_2 \circ \iota$. Then, for any rational $m/n \in \mathbb{Q}$ with $m, n \in \mathbb{Z}$, we have[9]
$$g_1\left(\frac{m}{n}\right) = \frac{g_1(m)}{g_1(n)} = \frac{(g_1 \circ \iota)(m)}{(g_1 \circ \iota)(n)} = \frac{(g_2 \circ \iota)(m)}{(g_2 \circ \iota)(n)} = \frac{g_2(m)}{g_2(n)} = g_2\left(\frac{m}{n}\right).$$
So indeed, $g_1 = g_2$ on $\mathbb{Q}$.

Remark 1.174. Being epic can be subtle. In the category of planar graphs, it turns out that "every epimorphism is surjective" is equivalent to the Four color theorem.

1.7.3 Functors

Let's start with some examples.

Example 1.175. Here are some examples.

- We have a functor from Grp to Set by taking a group to its underlying set of objects.
- We have another functor from Set to Grp by taking any set S to the free group on S.
- Homology is a functor from topological spaces X to abelian groups $H_\bullet(X)$.

Again, let's extract our common information. For example, a topological map $X \to Y$ turns into a corresponding map of homology groups $H_\bullet(X) \to H_\bullet(Y)$. Well in fact this is true for the other examples as well: a map of groups is of course a map of sets, and a map of sets can be lifted to a map of free groups.

So here is our definition.

[9] Showing that $g(m/n) = g(m)/g(n)$ requires some care. Because certainly $g(n)g(m/n) = g(m)$, it suffices to show that $n \neq 0$ implies $g(n) \in R^\times$. But $g(n)g(1/n) = g(n/n) = g(1) = n1_R$ because ring homomorphisms send $g : 1 \mapsto 1_R$.

Definition 1.176 (Covariant functor)**.** Fix $\mathcal{A}$ and $\mathcal{B}$ categories. Then a *(covariant) functor* $F : \mathcal{A} \to \mathcal{B}$ has the following data.

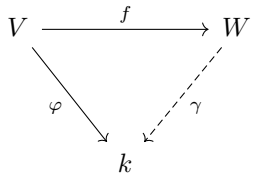
- The functor F takes objects of $\mathcal{A}$ to objects of $\mathcal{B}$.
- Given a morphism $f \in \mathrm{Mor}(A_1, A_2)$ with $A_1, A_2 \in \mathcal{A}$. Then there is a morphism $F(f) : F(A_1) \to F(A_2)$.
- Identity: given any object A, we have that $F(\mathrm{id}_A) = \mathrm{id}_{F(A)}$.
- Composition: given objects A_1, A_2, A_3 with morphisms $f \in \mathrm{Mor}(A_1, A_2)$ and $g \in \mathrm{Mor}(A_2, A_3)$, then $F(g \circ f) = F(g) \circ F(f)$.

Remark 1.177. We may skip the boring parts of the definitions because we don't care. Definitions in category theory are rather unmemorable.

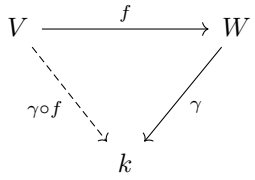
Let's have another example.

Example 1.178. Fix a group G acting on a set S. This is actually a functor from the category the one-object category representing G to Set. Namely, we take our one object to the set S, and each morphism in the one-object category correspond to some specified function on S. It turns out that the axioms of a group action correspond to the checks for being a functor.

To motivate the next definition, consider taking a vector space V over k in the category Vec_k to its dual. Namely, we take $V \in \mathrm{Vec}_k$ to $V^* := \mathrm{Hom}_k(V, k)$ its dual. Then given a map $f : V \to W$, we might want a morphism $f^* : V^* \to W^*$. Here is the diagram: given $\varphi \in V^*$, how do we induce $\gamma \in W^*$?



However, there is no useful way to take $\varphi : V \to k$ to a map $W \to k$, so we appear stuck. But conversely, it appears that given $\gamma : W \to k$, we can induce $\varphi : V \to k$ by precomposing as $\varphi := \gamma \circ f$! Here is the diagram.



So we can define the map $f^* : W^* \to V^*$ by taking $\gamma \in W^*$ to $\gamma \circ f \in V^*$. But now the direction of our morphisms is reversed, so we pick up the following definition.

Definition 1.179 (Contravariant functor)**.** A functor $F : \mathcal{A} \to \mathcal{B}$ is called *contravariant* if we "reverse" morphisms $f : A_1 \to A_2$ by $F(f) : F(A_2) \to F(A_1)$. The same definition still works, except we need to write the composition law as follows.

- Composition: given objects A_1, A_2, A_3 with morphisms $f \in \mathrm{Mor}(A_1, A_2)$ and $g \in \mathrm{Mor}(A_2, A_3)$, then $F(g \circ f) = F(f) \circ F(g)$.

So we can check that $V \mapsto V^*$ and $f \mapsto f^*$ is a contravariant functor $\mathrm{Vec}_k \to \mathrm{Vec}_k$. The composition law boils down to checking that

$$f^*(g^*\varphi) = f^*(v \mapsto \varphi g v) = (v \mapsto (\varphi g) f v) = v \mapsto \varphi(gf)v = (fg)^*(\varphi),$$

which is more annoying than interesting.

Remark 1.180. It turns out homology is a covariant functor, but cohomology is contravariant. So it goes.

1.7.4 Size Problems

We can also compose functors! For example, given $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{C}$, there is a functor $G \circ F : \mathcal{A} \to \mathcal{C}$. This lets us make the category of all categories where the objects are categories and the morphisms are functors! Except this doesn't work for size reasons, for the same reason there is no set of all sets.

There are a few ways to fix this.

- We can merely consider the category of all "small" categories, where we only consider the categories with at most some cardinal number of objects and morphisms.[10]
- We can use "classes" if we are careful.
- We can also use Grothendieck universes, which is common algebraic geometry.
- Ignore the problem completely.

We will take the last approach for our introduction here. We will have to encounter it later in life but not now.

1.7.5 Natural Transformations

Of course, we start with an example.

Example 1.181. Fix V a finite-dimensional vector space and V^* its dual. We know that $V \cong V^*$, but there is no "natural" isomorphism in general because we have to pick a basis first.

However, there is a "natural" isomorphism $V \cong V^{**}$. Namely, given an element of V, we can conically exhibit a map $(V \to k) \to k$ in V^{**}. To be explicit, we have a function $V \to V^{**}$ by

$$v \longmapsto (\varphi \mapsto \varphi v).$$

In λ-calculus, this reads $\lambda(v : V).\lambda(\varphi : V^*).\varphi v$. And we can take linear transformations $f : V \to W$ to $f^{**} : V^{**} \to W^{**}$, which looks like $f^{**} : \varphi \mapsto \varphi \circ f$.

We would like to rigorize what "natural" means. To start, we note that we have two functors $F, G : \mathrm{Vec}_k \to \mathrm{Vec}_k$ by $F : V \to V$ and $G : V \to V^{**}$, and we would like there to be a natural transformation between them. What does this mean? Well, what do we have?

- For any V, there is a map $\eta_V : F(V) \to G(V)$. Indeed, this was the map $V \mapsto V^{**}$ given in the example.
- For any morphism $f : V \to W$, the following diagram commutes.

$$\begin{array}{ccc} V & & F(V) & \xrightarrow{\eta_W} & G(V) \\ \downarrow{\scriptstyle f} & & {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ W & & F(W) & \xrightarrow[\eta_V]{} & G(W) \end{array}$$

Namely, given $v \in V$, we can track some $v \in V$ along both directions of the square, writing

$$v \overset{F(f)}{\longmapsto} fv \overset{\eta_W}{\longmapsto} (\varphi \mapsto \varphi f v)$$

while

$$v \overset{\eta_V}{\longmapsto} (\varphi \mapsto \varphi v) \overset{G(f)}{\longmapsto} (\varphi \mapsto \varphi f v).$$

[10] *The category of small categories would like to know your location.*

This is the general definition.

Definition 1.182 (Natural transformations)**.** Given two functors $F, G : \mathcal{A} \to \mathcal{B}$, we say that there is a *natural transformation* $\eta_\bullet : F \to G$ if we have the following data.

- For any $A \in \mathcal{A}$, there is a map $\eta_A : F(A) \to G(A)$.
- For any morphism $A_1 \to A_2$, the following diagram commutes.

$$\begin{array}{ccc} A_1 & & F(A_1) \xrightarrow{\eta_{A_2}} G(A_1) \\ \downarrow f & & {\scriptstyle F(f)}\downarrow \qquad\qquad \downarrow{\scriptstyle G(f)} \\ A_2 & & F(A_2) \xrightarrow[\eta_{A_1}]{} G(A_2) \end{array}$$

So the statement that V is naturally isomorphic with V^{**} turns into the fact that there is a natural transformation between $V \mapsto V$ and $V \mapsto V^{**}$.

Definition 1.183 (Natural isomrphism)**.** Fix everything as in the definition of a natural transformation. If $\eta_A : F(A) \to G(A)$ is an isomorphism, we call η a *natural isomorphism* and the functors F and G are naturally isomorphic.

So we see that $V \mapsto V$ and $V \mapsto V^{**}$ are naturally isomorphic in the category of finite-dimensional vector spaces over k.

Example 1.184. Fix $\mathcal{A}$ and $\mathcal{B}$ categories. Then the functors from $\mathcal{A}$ to $\mathcal{B}$ make a category, where objects are functors and morphisms are natural transformations. This is called a 2-category; we can even go further to 3-categories, 4-categories, and so on upwards to ∞-categories in the limit.

1.7.6 Adjoint Functors

Here is our basic example.

Example 1.185. Consider the categories of Set and Grp. We have two functors.

- The forgetful functor from $G : \mathrm{Grp} \to \mathrm{Set}$ by simply forgetting the group structure.
- The free functor from $F : \mathrm{Set} \to \mathrm{Grp}$ by sending sets to their free groups.

These are not inverses, but they are "adjoint." Namely, suppose we take a group X and set S and consider $G(X)$ and $F(S)$. Last class we noticed that we have the universal property that any function of sets $f : S \to G(X)$ induces (arguably, "lifts to") a unique morphism $g : F(S) \to X$. Here is the diagram.

$$\begin{array}{ccc} G(X) & \longleftarrow & X \\ {\scriptstyle f}\uparrow & & \vdots{\scriptstyle g} \\ S & \longrightarrow & F(S) \end{array}$$

In particular, we have a bijection between $\mathrm{Mor}(S, GX)$ and $\mathrm{Mor}(FS, X)$, and this bijection is "natural" in some sense. Rigorizing what we mean vey "natural" gives the following definition.

Definition 1.186 (Adjoints)**.** Fix functors $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{A}$. We say that (F, G) form an *adjoint pair* if and only if we have bijections

$$\tau_{A,B} : (A \to GB) \to (FA \to B)$$

causing the following two diagrams to commute: for any $\mathcal{A}$-morphism $f : A_1 \to A_2$ and $\mathcal{B}$-object B, the following must commute.

$$\begin{array}{ccc} A_1 & A_2 \to GB & \xrightarrow{\tau_{A_2,B}} FA_2 \to B \\ \downarrow f & \circ f \downarrow & \downarrow \circ Ff \\ A_2 & A_1 \to GB & \xrightarrow[\tau_{A_1,B}]{} FA_1 \to GB \end{array}$$

Additionally, there is an inverse diagram for $\tau_{A,B}^{-1} : (FA \to B) \to (A \to GB)$, which we won't write down. In this case, F is called *left adjoint* and G is called *right adjoint*.

It turns out that free functor is left adjoint with the forgetful functor; we won't check this here because it is a bit arduous.

Remark 1.187. We call these adjoint functors because they are related to adjoint linear transformations. Namely, given two linear transformations $F, G : V \to V$ quipped with an inner product $\langle \cdot, \cdot \rangle$, they are *adjoint* if and only if

$$\langle s, Gx \rangle = \langle Fs, x \rangle$$

for $s, x \in V$.

In general, right adjoints tend to be "forgetful" while left adjoints are "free." Here are some more examples of this.

Example 1.188. The following functors are adjoint.

- The forgetful functor from commutative rings CRing to sets Set.
- The free functor from a set $S \in \mathrm{Set}$ to the polynomial ring $\mathbb{Z}[S] \in \mathrm{CRing}$.

Example 1.189. The following functors are adjoint.

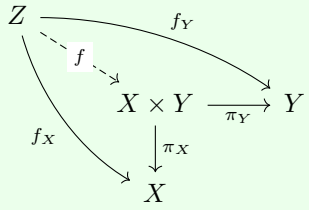
- The forgetful functor taking a complete metric space to a metric space.
- The completion functor taking a metric space X to its completion $\overline{X}$ by Cauchy sequences modded by some equivalence relation.

Here completion is left adjoint to the forgetful functor.

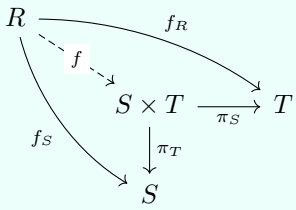
1.7.7 Products and Coproducts

Let's talk about products. The universal definition, say in Set is the set of pairs. This won't do in category theory, so here is our definition.

Definition 1.190 (Prodcuts)**.** Given objects X and Y, the *product* object $X \times Y$ is universal in the set of morphisms to X and Y. Namely, we are given maps $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ such that, given any object Z with maps $f_X : Z \to X$ and $f_Y : Z \to Y$, there is a unique map $f : Z \to X \times Y$ making the following diagram commute.



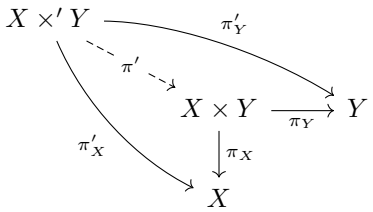
Example 1.191. We show that the product of sets S and T is indeed the product of sets; the maps $\pi_S : S \times T \twoheadrightarrow S$ and $\pi_T : S \times T \twoheadrightarrow T$ are the projections onto the corresponding coordinate. Now, suppose that we have a set R and maps $f_S : R \to S$ and $f_R : R \to T$ making the following diagram commute.



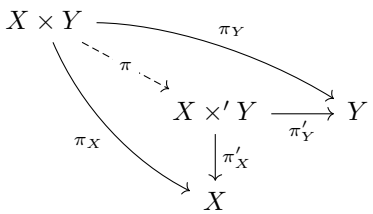
Then we define $f : R \to S \times T$ by $f(r) := (f_S r, f_T r)$. This works because $\pi_S(f_S r, f_T r) = f_S r$ and $\pi_T(f_S r, f_T r) = f_T r$, and this is forced because we must have $\pi_S(fr) = f_S r$ and $\pi_T(fr) = f_T r$.

Warning 1.192. Note that the product $X \times Y$, is not unique, but any two products are canonically isomorphic.

Indeed, suppose we have two products $X \times Y$ and $X \times' Y$ with projections $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ and $\pi'_X : X \times' Y \to X$ and $\pi'_Y : X \times' Y \to Y$. Then, we see that we get a $\pi' : X \times' Y \to X \times Y$ making the following diagram commute.



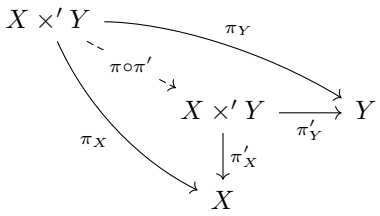
Similarly, we get a $\pi : X \times Y \to X \times' Y$ making the following diagram commute.



But now we claim that π and π' are isomorphisms, for which we show that $\pi \circ \pi' = \mathrm{id}_{X \times' Y}$ and $\pi' \circ \pi = \mathrm{id}_{X \times Y}$. We show that $\pi \circ \pi' = \mathrm{id}_{X \times' Y}$, and the other follows by symmetry. Well, for any $p \in X \times' Y$, we see

$$\pi'_\bullet(p) = (\pi_\bullet \circ \pi')(p) = \pi_\bullet(\pi' p) = \pi'_\bullet\big((\pi \circ \pi')(p)\big).$$

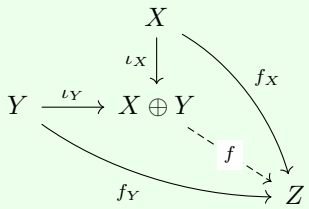
for either projection $\pi'_\bullet$. It follows that the following diagram commutes.



However, making $\mathrm{id}_{X \times' Y}$ the induced arrow also makes the above diagram commute, so we must have $\pi \circ \pi' = \mathrm{id}_{X \times' Y}$ because the induced arrow is unique! (Here we used the uniqueness of the induced arrow.)

And now let's talk about coproducts, which is defined by reversing the arrows of products.

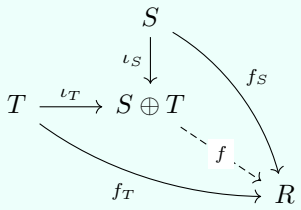
Definition 1.193 (Coproducts)**.** Given objects X and Y, the *coproduct* object $X \oplus Y$ is universal in the set of morphisms to X and Y. Namely, we are given maps $\iota_X : X \to X \oplus Y$ and $\iota_Y : Y \to X \oplus Y$ such that, given any object Z with maps $f_X : X \to Z$ and $f_Y : Y \to Z$, there is a unique map $f : X \oplus Y \to Z$ making the following diagram commute.



A similar warning as Warning 1.192 applies here, but a similar proof gives our canonical isomorphisms; we will not write it down here.

Intuitively, the coproduct is the smallest object containing X and Y. Here are some objects.

Example 1.194. The coproduct of two sets S and T is the disjoint union $S \sqcup T$. The inclusions $\iota_S : S \hookrightarrow S \sqcup T$ and $\iota_T : T \hookrightarrow S \sqcup T$ are standard. Now, suppose that we have a set R and maps $f_S : S \to R$ and $f_R : T \to R$ making the following diagram commute.



Then we define $f : S \sqcup T \to R$ by $f((s, 0)) = f_S s$ and $f((1, t)) = f_T t$. This works because $f(\iota_S s) = f_S s$ and $f(\iota_T t) = f_T s$, and this is forced by the same constraints.

Example 1.195. Given two abelian groups G_1 and G_2, the coproduct is $G_1 \times G_2$, where the inclusions $\iota_1 : G_1 \hookrightarrow G_1 \times G_2$ and $\iota_2 : G_2 \hookrightarrow G_1 \times G_2$ are defined by

$$\iota_1 : g_1 \mapsto (g_1, 0) \qquad \text{and} \qquad \iota_2 : g_2 \mapsto (0, g_2).$$

In general, the (finite) product of two abelian groups is the coproduct, which is quite remarkable and rare. For example, the product and coproduct are different in Set.

Example 1.196. More generally in groups, the coproduct of the two groups G and H, we want a group $G * H$ which is "as big as possible" given G and H as generators. So we take "reduced" words that look like

$$g_1 h_1 h_2 h_2 \cdots,$$

where $g_\bullet$ and $h_\bullet$ are nontrivial elements of G and H respectively. To show that all these "reduced" words are nontrivial, we make our words act on some giant graph. The idea is to build a Cayley graph by hand.

The point of this last example is that categorical coproducts, which look simple, are potentially very annoying.

Theme 2
Ring Rambles

One Ring to rule them all, One Ring to find them,
One Ring to bring them all and in the darkness bind them.

—J. R. R. Tolkien

2.1 September 21

All around me darkness gathers.

2.1.1 Rings

Today we do rings. We have the definition.

Definition 2.1 (Ring)**.** A *ring* is a set R with two operations $+$ and $\times$ such that $(R, +)$ is a group, $\times$ associates, and we distribute by

$$a(b+c) = ab + ac \qquad \text{and} (a+b)c = ac + bc.$$

We have some extra axioms as well.

Definition 2.2 (Commutative ring)**.** A *commutative ring* is one where multiplication commutes.

Definition 2.3 (Ring with unity)**.** A *ring with unity* is one with a multiplicative identity. We might call the multiplicative identity just "unity" or "identity."

The above two definitions are generally assumed but not by all authors.

Warning 2.4. In this course, our rings will generally be commutative with identity.

Anyways, have some examples.

Example 2.5. We have that $\mathbb{Z}$ is a ring.

Example 2.6. Any field is a ring.

Example 2.7. For R a commutative ring, the polynomials $R[x]$ form a commutative ring. If R has a multiplicative identity, then $R[x]$ has the same multiplicative identity.

Example 2.8. Given a ring R, the $n \times n$ matrices $R^{n\times n}$ form a ring. If R has multiplicative identity, then $R^{n\times n}$ has the identity matrix. However, if R is commutative, it is not necessary for $R^{n\times n}$ to be commutative.

Example 2.9. The Gaussian integers $\mathbb{Z}[i] : \{a + bi : a, b \in \mathbb{Z}\}$ make a ring.

Most of these rings are commutative with identity. Let's do some examples not containing 1.

Example 2.10. The set $C_0(\mathbb{R})$ consisting of all continuous real functions with compact support, where addition and multiplication are pointwise. However, our multiplicative identity in $\{f : \mathbb{R} \to \mathbb{R}\}$ is the $x \mapsto 1$ function, which does not have compact support.

In general, analysis has lots of natural examples like this.

Example 2.11. We can also define multiplication on $C_0(\mathbb{R})$ by

$$(f * g)(x) \coloneqq \int_{\mathbb{R}} f(y)g(x-y)\,dy$$

and makes a perfectly fine commutative ring, but there is no identity; note that this integral is surely well-defined because f and g have compact support. (The identity should be $1_{x=0}$, which is not continuous.)

The checks here are not very interesting; for example, distributivity comes down to noting

$$(f * (g + h))(x) = \int_{\mathbb{R}} f(y)(g + h)(x - y)\,dy = \int_{\mathbb{R}} f(y)g(x - y)\,dy + \int_{\mathbb{R}} f(y)h(x - y)\,dy$$

is $(f * g)(x) + (f * h)(x)$.

2.1.2 Modules

We can also define modules.

Definition 2.12 (Module)**.** A (left) *module* M over a ring R has exactly the same axioms as a vector space over a field.

- $(M, +)$ is an abelian group.
- M has a left R-action $\cdot : (R, M) \to M$, satisfying $(rs) \cdot m = r \cdot (s \cdot m)$ and the various distributive laws
$$r \cdot (m + n) = r \cdot m + r \cdot n \qquad \text{and} \qquad (r + s) \cdot m = r \cdot m + s \cdot m$$
for $r, s \in R$ and $m, n \in M$. In other words, there is a ring map $R \to \mathrm{End}(M)$.

We remark that sometimes we require $1_R m = m$ when R has an identity element, which we will also usually require.

Example 2.13. Vector spaces over a field are modules over that field.

Example 2.14. Abelian groups are $\mathbb{Z}$-modules, where the $\mathbb{Z}$-action is exponentiation.

2.1.3 Analogies

There is some correspondence between our algebraic structures, which for now are groups and rings.

- Groups act on sets in the same way that rings act on modules. There are even left and right actions in the same way that rings have left and right modules.
- There is the symmetric group S_n, which consist of all permutations of $\{1, \dots, n\}$. In rings, this is the matrix ring $R^{n\times n}$, which are all linear transformations $R^n \to R^n$.[1]
- At a high level, sets S correspond to free modules R^S, which is the module which consists of $\#S$ copies of R.
- Groups have permutation representations (which are group actions on a set), which correspond to linear representations of a ring (which are ring actions on a module).
- If a group acts on two sets A and B, then we can consider the set-theoretic union $A \cup B$ compute cardinalities as
$$|A \cup B| = |A| + |B| - |A \cap B|.$$
(Note there is a canonical way to get a group action on $A \cup B$.[2]) On the other hand, for vector spaces V and W which are subspaces of a bigger vector space X, we can compute
$$\dim(V + W) = \dim V + \dim W - \dim(V \cap W),$$
which looks quite similar.

Warning 2.15. The principle of inclusion-exclusion does not work for three vector spaces. For example, for three sets A, B, C we have
$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |C \cap A| + |A \cap B \cap C|.$$
The analogous formula for vector subspaces $U, V, W \subseteq X$ fails.

To manifest the warning, the formula fails for $U = \langle 1, 0\rangle\mathbb{R}$, $V = \langle 1, 1\rangle\mathbb{R}$, and $W = \langle 0, 1\rangle\mathbb{R}$ living in $\mathbb{R}^2$. We can compute
$$\dim(U + V + W) = \dim \mathbb{R}^2 = 2,$$
but
$$\dim U + \dim V + \dim W - \dim(U \cap V) - \dim(V \cap W) - \dim(W \cap U) + \dim(U \cap V \cap W) = 3.$$
At a high level, the problem here is that bases do not behave enough like sets.

- Regardless, (disjoint) unions of sets correspond to (direct) sums of modules.

[1] We remark that there is some care here: we want $R^{n\times n}$ acts on the left of R^n, but R acts on the left of $R^{n\times n}$.
[2] For example, if $\mathbb{Z}/2\mathbb{Z}$ swaps $A = \{1, 2\}$ and swaps $B = \{2, 3\}$, there is no good way for $\mathbb{Z}/2\mathbb{Z}$ to act on $A \cup B$.

- Given sets S and T with a G-action, we see $S \times T$ has a G-action by $g \cdot (s,t) \coloneqq (g \cdot s, g \cdot t)$, which has

$$\#(S \times T) = \#S \times \#T.$$

  This corresponds to tensor products of modules. As a quick and dirty definition, if V and W are k-vector spaces with bases $\{v_k\}_{k=1}^{\dim V}$ and $\{w_\ell\}_{\ell=1}^{\dim W}$, then we force $V \times W$ to have a basis $v_k \otimes v_\ell$. Then we see

$$\dim(V \otimes W) = \dim V \times \dim W.$$

2.1.4 Burnside Ring

Let's define an exotic ring: the Burnside ring of a group. We fix $G \coloneqq S_3$ for concreteness. We want to look at all isomorphism classes of finite sets with a G-action and make a ring, which will almost but not quite work.

For example, with sets S and T with a group action, we define

$$S + T \coloneqq S \sqcup T \qquad S \times T \coloneqq S \times T.$$

In other words, our addition is disjoint union, and our multiplication is product. To be explicit, our G-sets are as follows.

- The action of G on $S \sqcup T$ is defined by $g \cdot (s,0) \coloneqq (g \cdot s, 0)$ and $g \cdot (t,1) \coloneqq (g \cdot t, 1)$.
- The action of G on $S \times T$ is as given in the previous section.

Formally, we would have to check that that the isomorphism class of $S \sqcup T$ and $S \times T$ do not depend on the specific representative of $[S]$ and $[T]$, but this check is more annoying than difficult.

These operations obeys most of the ring actions, and multiplication even commutes! For example, our additive identity is $\varnothing$ (yes, groups can act on $\varnothing$), and the multiplicative identity is $\{*\}$ with the trivial action. However, there are no additive inverses because no operation can make our sets smaller.

To fix our additive inverse problem, we focus more closely on representatives of all transitive permutation representations of G; the point is that any G-action on a set is a disjoint union of how G acts on each orbit, which is transitive. Well, if G acts transitively on S, the order of our set S must be $\{1,2,3,6\}$ by the Orbit-stabilizer theorem. We can list the actions.

- If G acts on one element, then it is trivial.
- If G acts on two elements transitively, then one of the order-2 elements swaps, and the rest of the action can be determined from this. There is one way to do this, up to isomorphism.
- If G acts on $\{a,b,c\}$, transitively, then note that no transposition can be trivial, for then the entire conjugacy class of transpositions will be trivial, vanishing the image.

  Further, distinct transpositions must be sent to distinct transpositions in $\{1,2,3\}$, or else our action collapses to a transitive action on a two-element set. Without loss of generality, we send $(12) \mapsto (12)$ and $(23) \mapsto (bc)$ and $(13) \mapsto (ac)$ so that G is acting on S_3 on $\{a,b,c\}$.
- If G acts on six elements, it acts like S_3 on S_3 by left multiplication. Essentially, the transpositions need to all act on separate elements, else the action of G on the set will miss one of the six elements and hence not be transitive.

  From these transpositions we can determine the entire action, so there is at most one action here, up to isomorphism. Well, we can exhibit S_3 acting on S_3 by left multiplication as an action, so one certainly exists.

Label these isomorphism classes by $[1],[2],[3],[6]$. Now we see that any permutation representation of S_3 are isomorphic to some disjoint union of the above orbits; namely our elements take the form

$$a_1[1] + a_2[2] + a_3[3] + a_6[6] \qquad a_1, a_2, a_3, a_6 \geq 0.$$

To make this a ring, we just let a_1, a_2, a_3, a_6 vary over all integers.

Definition 2.16 (Burnside ring)**.** Fix G a group, and given a G-set S, let $[S]$ be the isomorphism class of S as a G-set. Now, label $\{[T]\}_{T\in\mathcal{T}}$ the isomorphism classes of transitive G-actions. Then the *Burnside ring* of G is defined as the free abelian group on $\mathcal{T}$

$$\bigoplus_{T\in\mathcal{T}} \mathbb{Z}[T]$$

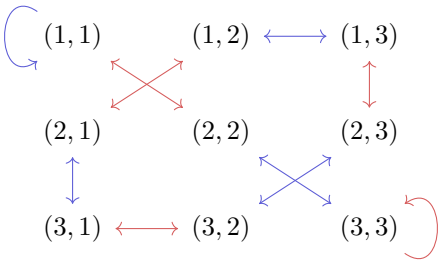
with multiplication defined as

$$[S_1]\times[S_2] \coloneqq [S_1\times S_2].$$

As an example, we'll compute $[3]\times[3]$. Well, here S_3 is acting on the following lattice.

$$\begin{array}{ccc} (1,1) & (2,1) & (3,1) \\ (1,2) & (2,2) & (3,2) \\ (1,3) & (2,3) & (3,3) \end{array}$$

For example, we can compute that (12) acts as red in the following diagram, and (23) acts as blue.



From these we see that we have at most two orbits, namely the diagonal and everything off of the diagonal. The diagonal is in fact an orbit because it is closed: for any (k,k) and $\sigma\in S_3$, then $\sigma(k,k)=(\sigma k,\sigma k)$ remains on the diagonal. It follows that the off-diagonal elements must also be closed and hence forms another orbit.

So we see that S_3 is acting transitively on a set of size 3 (the diagonal) and transitively on a set of size 6 (off the diagonal) so we find $\pi_3\times\pi_3=\pi_3+\pi_6$.

2.1.5 Group Rings

We remark that, as expected, Ring is a category of rings, where morphisms are homomorphisms are maps between rings which preserve the ring structure.

Further, modules over a fixed ring R form a category in the same way that vector spaces are; namely, our morphisms are linear transformations between R-modules.

We want some functors. Here is the group ring.

Definition 2.17 (Group ring)**.** The *group ring* $R[G]$ of a group G and ring R is defined as a free module of G with basis given by the elements of G. Then we define multiplication to distribute and then multiply as in the group:

$$\left(\sum_{g\in G} a_g g\right)\left(\sum_{h\in G} b_h h\right) \coloneqq \sum_{k\in G}\sum_{gh=k}(a_g b_h)k.$$

Remark 2.18. We can check that G acting linearly on a group is the same thing as finding a module over the group ring.

We note that the construction for a group ring also works for modules instead of groups. For example, we have the following.

Example 2.19. If we take the monoid $M = \mathbb{N}$, then the group ring $\mathbb{Z}[M]$ is the polynomial ring $\mathbb{Z}[x]$. Namely, $k \in \mathbb{N}$ gets taken to $k \mapsto x^k$.

So where are our functors? Well, we have the following.

Proposition 2.20. We claim that the (forgetful) functor $G : \mathrm{Ring} \to \mathrm{Grp}$ by $G : R \mapsto R^\times$ is right adjoint to the group ring functor $F : \mathrm{Grp} \to \mathrm{Ring}$ by $F : A \mapsto \mathbb{Z}[A]$.

Again, we are having that left adjoints appear free and right adjoints appear forgetful. At a high level, these are isomorphic because maps $G \to R^\times$ are in correspondence with maps $\mathbb{Z}[G] \to R$ because maps $\mathbb{Z}[G]$ must send group elements to elements of $\mathbb{R}^\times$.

Proof. We will actually do the checks for this because I should do this at least once in my life. The main point is the following lemma; roughly speaking this says that the group ring is the ring freely generated by a group.

Lemma 2.21. Fix A a group and R a ring with identity. Then we have the following.

(a) A morphism $\varphi : A \to R^\times$ can be uniquely lifted to a morphism $\overline{\varphi} : \mathbb{Z}[A] \to R$.

(b) Any morphism $\overline{\varphi} : \mathbb{Z}[A] \to R$ can be restricted to a morphism $\varphi : A \to R^\times$.

We remark that lifting and restriction are inverses of each other, well-defined by the uniqueness of the lifting.

Proof. We take these claims one at a time.

(a) Suppose that we have a morphism $\varphi : A \to R^\times$. Then, for any element

$$\sum_{a \in A} k_a a \in \mathbb{Z}[A],$$

properties of $\overline{\varphi}$ force that

$$\overline{\varphi}\left(\sum_{a \in A} k_a a\right) = \sum_{a \in A} k_a \varphi(a)$$

by distributing repeatedly. This shows that $\varphi : \mathbb{Z}[A] \to R$ is forced and hence unique. Conversely, the above actually works as a definition for φ, for which we have to check

$$\overline{\varphi}\left(\sum_{a \in A} k_a a \cdot \sum_{a \in A} \ell_a a\right) = \overline{\varphi}\left(\sum_{a \in A} k_a a\right) \overline{\varphi}\left(\sum_{a \in A} \ell_a a\right),$$

and

$$\overline{\varphi}\left(\sum_{a \in A} k_a a + \sum_{a \in A} \ell_a a\right) = \overline{\varphi}\left(\sum_{a \in A} k_a a\right) + \overline{\varphi}\left(\sum_{a \in A} \ell_a a\right),$$

which are true after some distributing. We do also have to check that $\overline{\varphi}(e) = e = 1e$, which holds because $\varphi : A \to R^\times$ is a group homomorphism.

(b) Conversely, suppose that we have a morphism $\overline{\varphi} : \mathbb{Z}[A] \to R$. We note that $\overline{\varphi}$ being a ring homomorphism implies that $\overline{\varphi}(a_1 a_2) = \overline{\varphi}(a_1)\overline{\varphi}(a_2)$, so $\overline{\varphi}$ restricted to A is homomorphic.

The main check is that restricting $\overline{\varphi}$ to A actually outputs into $R^\times$. Well, the multiplicative identity of $\mathbb{Z}[A]$ is $1e = e$, so for any $a \in A$, we find

$$\overline{\varphi}(a)\overline{\varphi}\left(a^{-1}\right) = \overline{\varphi}\left(aa^{-1}\right) = \overline{\varphi}(e) = 1_R,$$

so indeed $\operatorname{im}\varphi \subseteq R^\times$. ■

We now check that our functors are adjoint. There are two diagrams to check.

- Fix $\gamma : A_1 \to A_2$. We show that the following diagram commutes, for any ring R.

$$\begin{array}{ccc} A_1 & A_2 \to R^\times & \longrightarrow \mathbb{Z}[A_2] \to R \\ {\scriptstyle\gamma}\downarrow & {\scriptstyle -\circ\gamma}\downarrow & \downarrow{\scriptstyle -\circ\gamma} \\ A_2 & A_1 \to R^\times & \longrightarrow \mathbb{Z}[A_1] \to R \end{array}$$

  Here, the horizontal arrows are the ones promised by the bijection in Lemma 2.21. We check the commutativity by hand. Fix $\varphi : A_2 \to R^\times$. Following the diagram around, we see that we are showing $\overline{\varphi} \circ \gamma = \overline{\varphi \circ \gamma}$. Well, $\overline{\varphi \circ \gamma}$ is the unique morphism making the following diagram commute.

$$\begin{array}{ccc} A_1 & \xrightarrow{\varphi\circ\gamma} & R^\times \\ \downarrow & & \downarrow \\ \mathbb{Z}[A_1] & \xrightarrow[\overline{\varphi\circ\gamma}]{} & R \end{array}$$

  However, $\overline{\varphi} \circ \gamma$ makes the following diagram commute.

$$\begin{array}{ccccc} A_1 & \xrightarrow{\gamma} & A_2 & \xrightarrow{\varphi} & R^\times \\ \downarrow & & \downarrow & & \downarrow \\ \mathbb{Z}[A_1] & \xrightarrow[\gamma]{} & \mathbb{Z}[A_2] & \xrightarrow[\overline{\varphi}]{} & R \end{array}$$

  In particular, we see that $\overline{\varphi} \circ \gamma$ makes the diagram for $\overline{\varphi \circ \gamma}$ commute, so $\overline{\varphi} \circ \gamma = \overline{\varphi \circ \gamma}$ by uniqueness.

- Fix $\gamma : R_1 \to R_2$. We show that the following diagram commutes, for any group A.

$$\begin{array}{ccc} R_1 & A \to R_1^\times & \longrightarrow \mathbb{Z}[A] \to R_1 \\ {\scriptstyle\gamma}\downarrow & {\scriptstyle \gamma\circ -}\downarrow & \downarrow{\scriptstyle \gamma\circ -} \\ R_2 & A \to R_2^\times & \longrightarrow \mathbb{Z}[A] \to R_2 \end{array}$$

  Again, the horizontal arrows are coming from the bijection promised from Lemma 2.21. We check the commutativity by hand. Fix $\varphi : A \to R_1^\times$. Following the diagram around, we see that we are showing $\overline{\gamma} \circ \varphi = \overline{\gamma \circ \varphi}$. Well, $\overline{\gamma \circ \varphi}$ is the unique morphism making the following diagram commute.

$$\begin{array}{ccc} A & \xrightarrow{\gamma\circ\varphi} & R_2^\times \\ \downarrow & & \downarrow \\ \mathbb{Z}[A] & \xrightarrow[\overline{\gamma\circ\varphi}]{} & R_2 \end{array}$$

  However, $\gamma \circ \overline{\varphi}$ makes the following diagram commute.

$$\begin{array}{ccccc} A & \xrightarrow{\varphi} & R_1^\times & \xrightarrow{\gamma} & R_2^\times \\ \downarrow & & \downarrow & & \downarrow \\ \mathbb{Z}[A] & \xrightarrow[\overline{\varphi}]{} & R_1 & \xrightarrow[\gamma]{} & R_2 \end{array}$$

In particular, we see that $\gamma \circ \overline{\varphi}$ makes the diagram for $\overline{\gamma \circ \varphi}$ commute, so $\gamma \circ \overline{\varphi} = \overline{\gamma \circ \varphi}$ by uniqueness. ■

Fun with $\mathbb{C}[G]$

As an example, fix $G \coloneqq (\mathbb{Z}/2\mathbb{Z})^2$ giving the group ring $\mathbb{Z}[G]$. But we can also look at, say, $\mathbb{C}[G]$, which is a vector space with basis enumerated by $G = (\mathbb{Z}/2\mathbb{Z})^2$. We'll label $G = \{1, a, b, c\}$ for brevity.

Remark 2.22. We are using $G = (\mathbb{Z}/2\mathbb{Z})^2$ because Professor Borcherds is getting bored with cyclic groups.

However, we claim that $\mathbb{C}[G]$ is actually just four copies of $\mathbb{C}$, as a ring. Surely, $\mathbb{C}[G]$ is a vector space and splits into one-dimensional vector spaces, but that's not what we're interested in.

Roughly speaking, splitting a ring R into a product $S \times T$ corresponds with idempotent elements.

Definition 2.23 (Idempotent)**.** An element $x \in R$ is *idempotent* if and only if $x^2 = x$.

Example 2.24. Any ring R has $0 \cdot 0 = 0$, so 0 is idempotent. If R has identity 1, then $1 \cdot 1 = 1$, so 1 is also idempotent.

Example 2.25. If R is a commutative ring with identity satisfying $ab = 0$ implies $a = 0$ or $b = 0$, then $x^2 = x$ implies $x(x-1) = 0$ implies $x = 0$ or $x = 1$. So in fact 0 and 1 are only idempotents.

The reason we care about this is that the ring $S \times T$ will have the extra idempotents $(0,0), (1,1), (0,1), (1,0)$, which is a lot more that what we expect as just 0 and 1. It turns out that we can reverse: given nontrivial idempotents, then we can decompose R into a product of smaller rings.

Proposition 2.26. Suppose R is a commutative ring with identity 1 and an idempotent element $x \in R$. Then we have the direct sum

$$Rx \oplus R(1-x) \cong R.$$

Note that the rings Rx and $R(1-x)$ need not contain the unity element of R and in fact in general may not. But, for example in Rx, any $rx \in Rx$ has $rx \cdot x = rx^2 = rx$, so x serves as an identity here.

Proof of Proposition 2.26. Note that we have the map $\varphi : Rx \oplus R(1-x) \to R$ by $\varphi : (ax, b(1-x)) \mapsto ax + b(1-x)$, which we will not check is actually homomorphic, but it is. Note that φ is surjective because we can write

$$r = r1 = rx + r(1-x) = \varphi(rx, r(1-x)).$$

Further, φ is injective because it has trivial kernel: if $ax + b(1-x) = 0$, then we claim $ax = b(1-x) = 0$. Indeed, the trick is that $x^2 = x$ implies that $x(1-x) = 0$, so

$$ax = ax^2 = ax^2 + bx(1-x) = x \cdot \big(ax + b(1-x)\big) = 0.$$

Similarly,

$$b(1-x) = b(1-x)^2 = ax(1-x) + b(1-x)^2 = (1-x)\big(ax + b(1-x)\big) = 0,$$

which is what we wanted. ■

Remark 2.27. The representation if $r = ax + b(1-x)$ is only unique up to ax and $b(1-x)$, not up to a and b. In other words, it is possible for the map $R \times R \to R$ defined by $(a,b) \mapsto ax + b(1-x)$ to have a kernel, but this is not what Proposition 2.26 is claiming.

Remark 2.28. There is an analogous construction for non-commutative rings with identity: if $x \in R$ is idempotent, we can write

$$R \cong xRx \oplus xR(1-x) \oplus (1-x)Rx \oplus (1-x)R(1-x).$$

When the ring is commutative, the terms $xR(1-x)$ and $(1-x)Rx$ vanish because $x(1-x) = 0$.

With this in mind, let's find idempotents in $\mathbb{C}[G]$. We can check that the following are idempotent.

$$\left\{ \frac{1+a+b+c}{4}, \frac{1-a+b-c}{4}, \frac{1+a-b-c}{4}, \frac{1-a-b+c}{4} \right\}.$$

These are not the only idempotents (adding any of them will also give an idempotent), but they are good enough for a basis of $\mathbb{C}[G]$. Indeed, we claim that

$$\mathbb{C}[G] \cong \mathbb{C}\left[\frac{1+a+b+c}{4}\right] \oplus \mathbb{C}\left[\frac{1-a+b-c}{4}\right] \oplus \mathbb{C}\left[\frac{1+a-b-c}{4}\right] \oplus \mathbb{C}\left[\frac{1-a-b+c}{4}\right].$$

Note that each of the spaces on the right-hand side are still G-sets because they are closed under the G-action; for example, $\{1, a, b, c\}$ acting on $\frac{1-a+b-c}{4}$ will get sent to $\pm\frac{1-a+b-c}{4} \in \mathbb{C}\left[\frac{1-a+b-c}{4}\right]$ under multiplication by G. (This is the same check as $\frac{1-a+b-c}{4}$ is idempotent.)

To show the direct sum, we note that $\mathbb{C}[G]$ is four-dimensional as a $\mathbb{C}$-vector space, and the space

$$V \coloneqq \mathbb{C}\left[\frac{1+a+b+c}{4}\right] \oplus \mathbb{C}\left[\frac{1-a+b-c}{4}\right] \oplus \mathbb{C}\left[\frac{1+a-b-c}{4}\right] \oplus \mathbb{C}\left[\frac{1-a-b+c}{4}\right],$$

is at most dimension 4, so it suffices to show that the above spaces will span into all $\mathbb{C}[G]$. For this it is enough to note the natural map

$$(w,x,y,z) \longmapsto w\frac{1+a+b+c}{4} + x\frac{1-a+b-c}{4} + y\frac{1+a-b-c}{4} + z\frac{1-a-b+c}{4}.$$

is in fact bijective because this transformation corresponds to the matrix

$$\frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

which has nonzero determinant. So indeed, we see that $\mathbb{C}[G]$ splits as claimed.

Remark 2.29. It turns out that any abelian group ring over $\mathbb{C}$ will split into various copies of $\mathbb{C}$ as a ring as well, which is roughly speaking due to the representation theory of abelian groups.

Example 2.30 (Nir)**.** Take $G = \langle g \rangle \cong \mathbb{Z}/n\mathbb{Z}$. Then the group ring $\mathbb{C}[G]$ is decomposed with the idempotents $\frac{1}{n}\sum_{k=0}^{n}(\zeta^{\bullet}g)^k$, where ζ is a primitive nth root of unity. Indeed, we can compute that

$$\left(\sum_{k=0}^{n}(\zeta^a g)^k\right)\left(\sum_{\ell=0}^{n}(\zeta^b g)^\ell\right) = \sum_{k,\ell=0}^{n}\zeta^{ak+b\ell}g^{k+\ell} = \sum_{m=0}^{n}\left(\sum_{k+\ell=m}\zeta^{ak+b\ell}\right)g^m = \sum_{m=0}^{n}\left(\sum_{k=0}^{n}\zeta^{(a-b)k}\right)\zeta^{bm}g^m.$$

If $a = b$, then the internal sum evaluates to n; if $a \neq b$, then the internal sum vanishes. This shows that these elements are orthogonal idempotents.

Dirichlet Series

Let's do another example of a monoid ring, as formal series.

Example 2.31. So we have that $\mathbb{C}[\mathbb{N}]$ is the polynomials over $\mathbb{C}$, where $\mathbb{N}$ is the naturals under addition.

Example 2.32. We consider the set of formal Dirichlet series

$$\sum_{k=1}^{\infty} \frac{a_k}{k^s}.$$

Our multiplication is defined formally by

$$\left(\sum_{k=1}^{\infty} \frac{a_k}{k^s}\right)\left(\sum_{\ell=1}^{\infty} \frac{b_\ell}{\ell^s}\right) = \sum_{n=1}^{\infty}\left(\sum_{k\ell=n} a_k b_\ell\right)\frac{1}{n^s}.$$

Note that this is very different from the usual polynomial ring multiplication.

It turns out that the finite Dirichlet series are the ring $\mathbb{C}[\mathbb{N}^\times]$, where $\mathbb{N}^\times$ is the monoid of the nonzero naturals under multiplication. Then in the same way that we can make $\mathbb{C}[\mathbb{N}]$ formal by making it infinite, we can make Dirichlet series infinite.

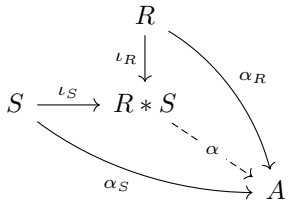
As an example of what we can do, number theorists care a lot about Dirichlet series. For example, we have

$$\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s} \qquad \text{and} \qquad \frac{1}{\zeta(s)} = \sum_{k=1}^{\infty} \frac{\mu(n)}{n^s},$$

where μ is the Möbius function.

2.1.6 Coproducts

We can also talk about categorical coproducts in R. To review, this means that we have rings R and S and want a ring $R * S$ with inclusions $\iota_R : R \to R * S$ and $\iota_S : S \to R * S$ that satisfies the universal property. Explicitly, for any other ring A with maps $\alpha_R : R \to A$ and $\alpha_S : S \to A$, there is a unique induced map $\alpha : R * S \to A$ making the following diagram commute.



However, it is an important point that what the coproduct is depends on whether we are working with commutative rings or non-commutative rings.

Example 2.33. Consider $R = S = \mathbb{Z}[x]$.

- The coproduct $\mathbb{Z}[x] * \mathbb{Z}[x]$ in the category of commutative rings is the two-variable polynomials $\mathbb{Z}[x, y]$. Indeed, any ring map $\mathbb{Z}[x] \to X$ corresponds to deciding where x should go in X, so when deciding on a map $\mathbb{Z}[x] * \mathbb{Z}[x]$, we have two decisions to make for each $\mathbb{Z}[x]$. Making two choices is the same as making choices for $\mathbb{Z}[x, y]$.
- The coproduct in the category of all rings is the "non-commutative polynomial ring," which is the ring formed over $\mathbb{Z}$ where the generators x, y do not need to commute. Namely, we have our $\mathbb{Z}$-module freely generated by

$$1, x, y, \quad x^2, xy, yx, y^2, \quad x^3, x^2y, yx^2, xy^2, y^2x, y^3, \dots$$

So this is similar to the story for groups, where things that commute are good, but things that don't commute are difficult to get a handle of. To be more explicit, the generators of the coproduct in non-commutative rings turn look like some kind of free (non-commutative) monoid generated by x, y.

2.1.7 Ideals

We work with commutative rings here. We can also define rings based off of their generators and relations. In the group story, we wanted normal subgroups, but here our story is different. Suppose that we have a surjective ring homomorphism $\varphi : R \to S$. Then $\ker \varphi$ satisfies the following.

- $\ker \varphi$ is closed under addition.
- $\ker \varphi$ is closed under multiplication by any element of R.

This defines an ideal.

Definition 2.34 (Ideal)**.** An ideal I of a ring R satisfies the following, for any $r \in R$ and $a, b \in I$.

- $a + b \in I$.
- $ra, ar \in I$.

Remark 2.35. If our rings are non-commutative, then we have to deal with left and right ideals, which might be closed under multiplication on one side but not the other.

We can show that all kernels are ideals.

Lemma 2.36. Fix R and S commutative rings with identity. Fix a ring homomorphism $\varphi : R \to S$. Then $\ker \varphi$ is an ideal.

Proof. We check the conditions one at a time.

- If $k_1, k_2 \in \ker \varphi$, then $\varphi(k_1 + k_2) = \varphi(k_1) + \varphi(k_2) = 0_S + 0_S = 0_S$, so $k_1 + k_2 \in \ker \varphi$.
- If $r \in R$ and $k \in \ker \varphi$, then
$$\varphi(rk) = \varphi(r)\varphi(k) = \varphi(r) \cdot 0_S = 0_S,$$
so $rk \in \ker \varphi$ as well. ∎

The converse is also true, using a similar construction as with quotient groups.

Lemma 2.37. Fix R a commutative ring and I an ideal. Then we can define the quotient group R/I and make R/I into a ring by

$$aI \cdot bI = (ab)I$$

for $a, b \in R$. If R has unity, then I does as well.

Proof. The quotient group R/I exists because $(R, +)$ is abelian, so I is a subgroup and hence a normal subgroup.

The main thing to check is that multiplication is well-defined. Well, suppose that $a_1 I = a_2 I$ and $b_1 I = b_2 I$ so that $a_1 - a_2 =: i_a \in I$ and $b_1 - b_2 =: i_b \in I$. Then we want to verify that

$$(a_1 b_1)I \stackrel{?}{=} (a_2 b_2)I.$$

Indeed, we see

$$a_2 b_2 = (a_1 + i_a)(b_1 + i_b) = a_1 b_1 + \underbrace{b_1 i_a + a_1 i_b + i_1 i_2}_{},$$

where the bracketed part is in I by definition of the ideal.

We will not check that all of the various ring axioms hold; they are mostly just inherited directly from R. ∎

The point is that the canonical map $R \twoheadrightarrow R/I$ by $r \mapsto rI$ goes into the kernel if and only if $r \in I$, so this map has kernel I. So indeed, any ideal can be constructed as a kernel. In this way ideals are somewhat similar to normal subgroups, in that they are the conditions we want to make quotients.

So now, to define a ring (or group) by generators and relations, we pick up some generators $\{a_k\}_{k=1}^n$, which generate a free ring (or group). Then we want to quotient by the ideal (or normal subgroup) generated by those relations. We give these constructions more explicitly as follows.

- To be formal, let $\mathrm{Free}(S)$ be the free group generated by S.

  For our construction, doing this for groups means we start with the letters $\{a_\alpha\}_{\alpha\in\lambda}$ generating the free group $\mathrm{Free}(\{a_\alpha\}_{\alpha\in\lambda})$. Then, given some relations we want to mod out by as words $\{w_\beta\}_{\beta\in\kappa}$, we note that there is a morphism $\varphi : \mathrm{Free}(\kappa) \to F$ lifting

  $$\varphi : \beta \mapsto w_\beta.$$

  Then the group G given by the letters $\{a_\alpha\}$ and relations $\{w_\beta\}$ is

  $$F/\overline{\mathrm{im}\,\varphi},$$

  which is F modulo the normal closure of $\mathrm{im}\,\varphi$. Explicitly,

  $$\overline{\mathrm{im}\,\varphi} := \bigcap_{\mathrm{im}\,\varphi \subseteq N} M,$$

  where N loops over normal subgroups of F.

- A bit more easily, we can define the ring generated by letters $\{a_\alpha\}_{\alpha\in\lambda}$ and words $\{w_\beta\}_{\beta\in\kappa}$ to be the free ring modulo

  $$\bigoplus_{\beta\in\kappa} F w_\beta,$$

  which is the ideal generated by the words $\{w_\beta\}_{\beta\in\kappa}$.

Example 2.38. Fix G generated by a, b with relations $a^2 = b^2 = (ab)^n = e$. We can check that G is D_{2n}, where a and b are some particular reflections.

In general, it is very hard to find the group given generators and relations.

Non-Example 2.39. Fix G generated by a, b, c with relations $aba^{-1} = b^2$ and $bcb^{-1} = c^2$ and $cac^{-1} = a^2$. This problem turns out to be very hard, even to determine if G is trivial or not.

Remark 2.40. In fact, there is a theorem that there is no algorithm which can in general turn a system of generators and relations into a group structure, or even if the group is trivial.

Example 2.41. Fix generators x, y with the relation $y^2 = x^3 - x$, and we look at the free polynomial ring $\mathbb{C}[x, y]$. Then we are studying

$$\frac{\mathbb{C}[x, y]}{(y^2 - x^3 + x)}.$$

Things ring can be interpreted as the polynomial functions from the curve $y^2 = x^3 - x$ to, say, $\mathbb{C}$. Namely, we want to identify two polynomials on $y^2 = x^3 - x$ if they are equal on all points of $y^2 = x^3 - x$, which is the same as modding out by polynomials which identically vanish on $y^2 = x^3 - x$.

For more related to the above example, see algebraic geometry.

2.2 September 28

He was safe, for now. But the dark thoughts would soon return.

2.2.1 Introducing Unique Factorization

We're talking factorization today.

Warning 2.42. All rings today are commutative with identity and have no zero-divisors. In other words, $ab = 0$ for a, b in our rings will imply $a = 0$ or $b = 0$.

Namely, we have the following definition.

Definition 2.43 (Integral domain)**.** A ring R is an *integral domain* if and only if it is nonzero, commutative with identity, and $ab = 0$ implies $a = 0$ or $b = 0$.

The main thing that integral domains gives us is a cancellation law for multiplication: if $ac = bc$ with $c \neq 0$, then $(a - b)c = 0$ while $c \neq 0$, so $a = b$.

Here is the standard example of unique prime factorization.

Example 2.44. For $\mathbb{Z}$, we have unique prime factorization: every positive integer is the product of positive primes, uniquely (up to permutation). Speaking more abstractly, every nonzero integer is the product of primes and units, unique up to permutation and multiplication by some unit. For example, 1 and -1 are both products of empty sets of primes.

Let's try to generalize our factorization.

Definition 2.45 (Prime)**.** Fix R a commutative ring. Then $p \in R$ is *prime* if and only if p is nonzero, not a unit, and $p \mid ab$ implies $p \mid a$ or $p \mid b$.

This is not the definition most of us are used to from elementary school. The other definition has a different name.

Definition 2.46 (Irreducible)**.** Fix R a ring with identity. Then $p \in R$ is called *irreducible* if and only if p is nonzero, not a unit, and $p = ab$ implies a or b is a unit.

Note that 1 (and units more generally) are neither prime nor irreducible. It just turns out to be more convenient that way.

Remark 2.47. Professor Borcherds thinks arguing about whether 1 is prime or not is pointless. It is not prime by definition.

We are going to talk about factorization in a few steps.

(i) We start with $\mathbb{Z}$, which is everyone's favorite.

(ii) It happens that $\mathbb{Z}$ is a Euclidean domain.

(iii) We will show all Euclidean domains are principal ideal domains.

(iv) Then we will show all principal ideal domains are unique factorization domains.

And unique factorization domains are the ones that we want.

Definition 2.48 (Unique factorization domain)**.** A ring R is a *unique factorization domain* if and only if it is an integral domain and every element can be written as a product of irreducibles, where the product is unique up to permutation and multiplication by units.

2.2.2 Euclidean Domains

So let's start with Euclidean domains.

Definition 2.49 (Euclidean)**.** A *Euclidean domain* is an integral domain R with a division algorithm. In other words, given a, b with $a \neq 0$, we can divide

$$b = aq + r$$

where $0 \leq |r| < |a|$ for some notion of $|\cdot|$. We also require that $|\cdot| : R \to \mathbb{Z}_{\geq 0}$.

Warning 2.50. The above requirements on the Euclidean function are somewhat nonstandard. More typical would be $|\cdot| : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ satisfying a division algorithm and $|a| \leq |ab|$ for any $a, b \in R$.

The exact requirements on the "norm" $|\cdot|$ are not very standard and not worth memorizing. The main point is that we can write down a statement of the division algorithm, we want 0 to be smaller than all the other elements, and we want there to be only finitely many elements of bounded norm.

Anyways, here are some examples.

Example 2.51. We have that $\mathbb{Z}$ has a division algorithm, where $|\cdot|$ is the usual absolute value.

Example 2.52. The ring $k[X]$ for a field k, where we use $\deg$ for our size function. Technically we want $|0| = 0$ and $|f| = \deg f + 1$ for $f \neq 0$ to make this work with the above definition.

Example 2.53. The Gaussian integers $\mathbb{Z}[i] = \{a+bi : a, b \in \mathbb{Z}\}$ have a division algorithm, where $|a+bi| = a^2 + b^2$.

In general, here is our idea.

Idea 2.54. A Euclidean domain is an integral domain where we cannot have infinite descending chains of elements.

2.2.3 Principal Ideal Domains

Recall the following definition.

Definition 2.55. A ring R is *principal* if and only if all ideals are generated by one element, i.e., are principal.

Lots of the rings we love are principal ideal domains: $\mathbb{Z}$, $k[X]$ for a field k, and so on. Let's see some rings which aren't.

Non-Example 2.56. The ring $k[X,Y]$ is not a principal ideal domain: take I to be (X,Y), which cannot be reduced to be a single generator. In other words, I is the ideal of polynomials with no constant term.

Indeed, suppose that $(X,Y) \subseteq (f)$ for some $f \in k[X,Y]$. Then $f \mid X$ and $f \mid Y$, and degree arguments show that $f \mid X$ implies there is some $c_x \in k^\times$ such that $f = c_x$ or $f = c_x X$. But $c_x X \nmid Y$ because $X \nmid Y$, so we must have $f \equiv c \in k^\times$. But then $(f) = k[X,Y]$, so $(f) \neq (X,Y)$.

Non-Example 2.57. The ring $\mathbb{Z}[X]$ is also not a principal ideal domain: take $I = (2,X)$. In other words, I is the ideal of polynomials with even constant term.

The proof that I is not principal is similar to before. Suppose $f \in \mathbb{Z}[X]$ has $(2,X) \subseteq (f)$. Then $f \mid 2$, so $f = \pm 1$ or $f = \pm 2$. Note $f = \pm 2$ generates an ideal missing X, so this does not work. So $f = \pm 1$, so $(f) = \mathbb{Z}[X]$, and $(f) \neq (2,X)$.

Remark 2.58 (Nir)**.** At a high level, what is happening with the above rings is that they are Noetherian of dimension 2.

Here is one step in the outline we gave at the start, which is the reason we brought up Euclidean domains to begin with.

Proposition 2.59. All Euclidean domains are principal ideal domains.

Proof. Fix R a Euclidean domain, and pick up an ideal I. Then by well-ordering we can find the minimum of

$$\{|a| : a \in I \setminus \{0\}\} \subseteq \mathbb{Z}_{\geq 0}$$

as well as some $a \in I \setminus \{0\}$ with the minimal $|a|$.

We claim $I = (a)$. In one direction, $a \in I$ implies $(a) \subseteq R$. In the other direction, take any $b \in I$. We note $a \neq 0$, so we can apply division, writing

$$b = aq + r$$

with $|r| < |a|$. But then by minimality of $|a|$, it follows that $r \notin I \setminus \{0\}$, so $r = 0$. Thus, $b = aq$, and $b \in (a)$, which shows $I \subseteq (a)$, finishing. ∎

We can ask if the converse is true: are all principal ideal domains Euclidean? Usually the answer is yes in practice, but it is false in general.

Exercise 2.60. The ring $R \coloneqq \mathbb{Z}\left[\frac{1+\sqrt{-19}}{2}\right]$ is a principal ideal domain but not a Euclidean domain.

Proof. To see that R is a principal ideal domain, see any course on algebraic number theory. The words to google are "class group" and "Minkowski bound."

To show that R is not Euclidean, we take the following lemma.

Lemma 2.61 (Universal side divisor criterion)**.** Suppose that R is a Euclidean domain. Then there exists a nonzero $a \in R \setminus R^\times$ such that the cosets $R/(a)$ can each be represented by unit or zero.

Proof. The idea is to take a to have the smallest norm, outside of units. Well-ordering implies that the set

$$\{|a| : a \in R \setminus (\{0\} \cup R^\times)\} \subseteq \mathbb{Z}_{\geq 0}$$

has a minimum, so we can find a nonzero $a \in R \setminus R^\times$ with minimal norm.

Now, pick up any coset $b + (a) \in R/(a)$. Applying division by a, we see that

$$b = aq + r$$

for some $|r| < |a|$. Then we see $b + (a) = r + (a)$ while the minimality of $|r|$ implies that $r \in \{0\} \cup R^\times$. This finishes. ∎

Remark 2.62. In fact, all elements of smallest norm (excluding 0) are units. Indeed, well-ordering implies that

$$\{|a| : a \in R \setminus \{0\}\}$$

has a smallest element. Then for any $u \in R \setminus \{0\}$ minimizing $|u|$, we claim that u is a unit. Indeed, dividing 1 by u we find

$$1 = qu + r$$

where $|r| < |u|$. But by minimality of $|u|$, we must have $r \notin R \setminus \{0\}$, so $r = 0$, implying that $1 = qu$, and u is a unit.

However, R does not satisfy the universal side divisor criterion, so it cannot be Euclidean. Indeed, the only units[3] of R are $\{\pm 1\}$ implying that we would need some $a \in R$ with $1 < \#(R/(a)) \leq 3$.

But no such a exists; the argument here is a bit technical and taken from here. The point is that, if such an a existed, there would be a ring homomorphism

$$R \twoheadrightarrow R/(a),$$

but $1 < \#(R/(a)) \leq 3$ implies that $R/(a)$ is a ring with two or three elements, of which the only options are $\mathbb{F}_2$ and $\mathbb{F}_3$.

The obstruction, now, is that R has $\theta := \frac{1+\sqrt{-19}}{2}$, which is a root of $x^2 - x + 5 = 0$, but $\mathbb{F}_2$ and $\mathbb{F}_3$ have no roots of this polynomial (this is checked by hand). Ring homomorphisms preserve polynomial equations, so no ring homomorphism may exist. ∎

2.2.4 Getting Unique Factorization

We now show that all principal ideal domains are unique factorization domains. This is done in steps.

Proposition 2.63. Fix R a principal ideal domain. Then all nonzero $a \in R \setminus R^\times$ are divisible by some irreducible. In fact, we may weaken the condition that R is a principal ideal domain to require all ascending chains of principal ideals to stabilize.

Proof. Roughly speaking, this is done by an induction-like argument. Fix $a_0 \in R \setminus \{0\}$. If $a_0 \in R$ is irreducible, we are done. Otherwise, we can factor $a_0 = a_1 b_1$ with $a_1, b_1 \in R \setminus R^\times$ and nonzero because R is an integral domain. Then we can factor a_1 further, and so on. Formally, we have the following algorithm.

[3] If $u \mid 1$, then $\overline{u} \mid 1$, so $u\overline{u} \mid 1$, and so $u = a + b\frac{1+\sqrt{-19}}{2}$ implies $\left(a + \frac{1}{2}b\right)^2 + \left(\frac{19}{2}b\right)^2 = 1$, so $a = \pm 1$ and $b = 0$ by bounding.

1. Starting with a nonzero $a_k \in R \setminus R^\times$, we may factor $a_k = a_{k+1}b_{k+1}$ where a_{k+1} and b_{k+1} are not units and not zero.
2. If a_{k+1} is irreducible, then it is an irreducible factor of a_k, which is a factor of a_0 by working our way back up the chain.
3. Otherwise, return to the first step with a_{k+1}. Any irreducible factor of a_{k+1} will also be an irreducible factor of a_k.

If this algorithm terminates, we are done. Otherwise, suppose for the sake of contradiction we can find an infinite strictly descending sequence of elements $\{a_0, a_1, \dots\}$ where

$$\frac{a_k}{a_{k+1}} \in R \setminus R^\times$$

for each k. Equivalently, we have the strictly ascending chain

$$(a_0) \subsetneq (a_1) \subsetneq (a_2) \subsetneq\subsetneq \cdots.$$

(This is strictly ascending because $(a_k) = (a_{k+1})$ implies $a_k/a_{k+1} \in R^\times$.) By hypothesis, this chain of ascending chain of principal ideals should stabilize, but we should show this is true for principal ideal domains.

Remark 2.64. The above condition is actually possible in "big" rings. For example,

$$k\left[X, X^{1/2}, X^{1/4}, \dots\right]$$

has the infinite strictly ascending chain

$$(X) \subsetneq \left(X^{1/2}\right) \subsetneq \left(X^{1/4}\right) \subsetneq \cdots.$$

Of course this is ascending, and it is strictly ascending because no $X^{1/2^n}$ is a unit by "degree" arguments.

Indeed, these infinite strictly ascending chains cannot happen for principal ideal domains for Noetherian reasons.

Lemma 2.65. Fix R a principal ideal domain and suppose that we have an ascending sequence of ideals

$$(a_0) \subseteq (a_1) \subseteq (a_2) \subseteq \cdots.$$

Then there exists some N such that $(a_n) = (a_N)$ for $n \geq N$.

Proof. The trick is to look at the ideal

$$\bigcup_{k=0}^{\infty} (a_k) = (a_0, a_1, \dots) = (b),$$

where $b \in R$ exists because all ideals are principal! But then

$$b \in (a_n)$$

for some n by the definition of a union. It follows that

$$(a_{n+1}) \subseteq \bigcup_{k=0}^{\infty} (a_k) = (b) \subseteq (a_n),$$

so we get $(a_n) = (a_{n+1})$, ∎

In light of the above lemma, there exists some N in our chain so that $(a_N) = (a_{N+1})$. But then $a_N/a_{N+1} \in R^\times$, violating the construction of our chain, which is our contradiction. ■

We continue.

Proposition 2.66 (Existence of factorizations)**.** Fix R an integral domain in which every ascending chain of principal ideals must stabilize. Every nonzero $a \in R$ is the product of irreducibles.

Proof. We do the same argument as above, factoring a nonzero $a_0 \in R$ by stripping out one irreducible at a time. Formally, we have the following algorithm, for $k \geq 0$.

1. If a_k is a unit, then we take the empty product of irreducibles and are done; i.e., our factorization is "a_k."
2. Otherwise, we know from Proposition 2.63 that a_k has an irreducible factor, say π_k.
3. Now return to the first step with $a_{k+1} \coloneqq a_k/\pi_k$. We note $a_k = \pi_k \cdot a_{k+1}$.

This algorithm creates the strictly ascending chain of ideals

$$(a_0) \subsetneq (a_1) \subsetneq (a_2) \subsetneq \cdots,$$

where the ascending is strict because $(a_k) = (a_{k+1})$ would imply that $\pi_k = a_k/a_{k+1}$ is a unit, which is not the case because π_k is irreducible.

So eventually the strictly ascending chain must stop, so there is some a_N which is a unit. So we have

$$a_0 = \pi_1 a_1 = \pi_1 \pi_2 a_2 = \cdots = a_N \prod_{k=1}^{N-1} \pi_k,$$

which after pushing the unit anywhere becomes a factorization of a_0 into irreducibles. ■

Remark 2.67. Being a product of irreducibles does not use the full power of being a principal ideal domain. We really only need to know the ring is "Noetherian," which means every ideal is finitely generated. Indeed, the meat of the above argument is showing that there are no infinite strictly ascending chain of (principal) ideals.

So the hard part is going to be showing uniqueness. The main claim will be that irreducibles are prime. It's easy to show that primes are irreducible, but the reverse is hard.

Lemma 2.68. Fix R an integral domain. Then a prime $p \in R$ is irreducible.

Proof. Being prime already gives p not a unit and nonzero. Now suppose that we can factor $p = ab$ so that we want to show one of the factors is a unit.

Well, $p \mid ab$, so p prime implies $p \mid a$ or $p \mid b$. Without loss of generality, take $p \mid a$. Then we see $p = ab$ implies

$$1 = (a/p) \cdot b,$$

so indeed, b is a unit. ■

Anyways, let's show uniqueness assuming that all irreducibles are prime.

Proposition 2.69 (Uniqueness of factorizations)**.** Fix R an integral domain in which all irreducibles are prime. Then factorization into irreducibles is unique up to permutation of the factors and multiplication by units.

Proof. Suppose we have two factorizations into irreducibles notated

$$\prod_{k=1}^{m} p_k = \prod_{\ell=1}^{n} q_\ell.$$

Without loss of generality, take $m \geq \ell$. If $m = 0$, then both sides are empty, and there is nothing to show.

Otherwise, we can pick up p_m an irreducible. By hypothesis, p_m is prime while dividing the right-hand side, so p_m divides one of the factors. Without loss of generality (permuting the elements), we take

$$p_m \mid q_n,$$

so we write $q_n = p_m u$. But now q_n is irreducible, so one of p_m or u is a unit, but it cannot be p_m, so u is the unit. This means that we can divide out

$$\prod_{k=1}^{m-1} p_k = u \prod_{\ell=1}^{n-1} q_\ell$$

to get a smaller factorization and finish by induction. Intuitively, we can just keep stripping off irreducible factors from both sides, one at a time. ■

So let's get into the meat of the proof.

Remark 2.70. According to Professor Borcherds, most of what we have been doing has been bookkeeping and has not required any ideas. What follows does.

Proposition 2.71. Fix R a principal ideal domain. Then all irreducibles are prime.

Proof. Fix p an irreducible so that we want to show p is prime. Well, suppose $p \mid ab$ so that we want $p \mid a$ or $p \mid b$. The trick is to focus on

$$(p, a) = (c),$$

where $c \in R$ exists because we live in a principal ideal domain. We see $p \in (c)$ implies $c \mid p$, so writing $p = cu$, one of c or u is a unit. We now do casework.

- If u is the unit, then $a \in (c)$ implies $c \mid a$ implies $cu \mid au$ implies $p \mid au$ implies $p \mid a$.
- Otherwise c is the unit so that $(c) = R$. We can write

  $$1 = xp + ya$$

  for some $x, y \in R$, which implies

  $$b = bxp + bya = (bx + yab/p) \cdot p,$$

  so $p \mid b$. This finishes.

So in all cases, we have $p \mid a$ or $p \mid b$, finishing. ■

This finishes the proof.

Theorem 2.72 (Unique factorization)**.** Every principal ideal domain is a unique factorization domain.

Proof. This follows from combining Proposition 2.66 and Proposition 2.69. ■

Remark 2.73. This proof is more or less in Euclid's *Elements*, but the statement was not. Euclid didn't have a good notion of multiplying more than three elements at once.

2.2.5 Gaussian Integers

Let's work with some examples now.

Theorem 2.74. The Gaussian integers $\mathbb{Z}[i]$ is a Euclidean domain and hence a unique factorization domain.

Proof. Our norm function on $\mathbb{Z}[i]$ will be

$$|x+yi| = x^2+y^2.$$

We want to show that, given $a, b \in \mathbb{Z}[i]$ with $a \neq 0$, we can write

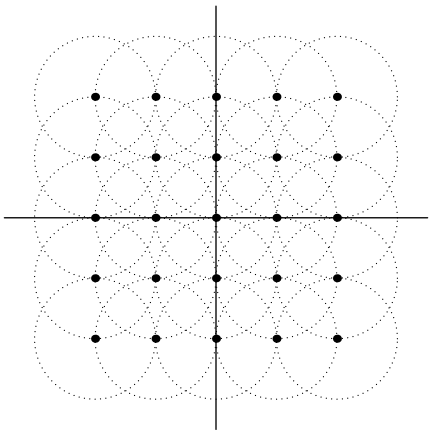
$$b = aq + r$$

where $|r| < |a|$. Equivalently, we are saying that

$$\frac{b}{a} = q + \frac{r}{a},$$

where we want $|r| < |a|$, which is equivalent to $|r/a| < 1$ because our norm is multiplicative in $\mathbb{Z}[i]$. (The norm is multiplicative in $\mathbb{C}$.) So essentially we are asking if every Gaussian integer is off by a distance of at most one from a Gaussian integer.

Well, geometrically, we place a unit circle around each Gaussian integer, as follows.



Now it's pretty clear that, for any $z \in \mathbb{C}$, we can find some $x+yi \in \mathbb{Z}[i]$ such that the distance between z and $x+yi$ is at most 1, which is what we wanted.

Formally, we can define q by dividing $b\overline{a}$ by $|b|$ coordinate-wise and rounding to get the components of q. This argument is a bit long and annoying, so we will write it out exactly once. For concreteness, set $a = a_1 + a_2 i$ and $b = b_1 + b_2 i$. Now,

$$a\overline{b} = (a_1+a_2i)(b_1-b_2i) = \underbrace{(a_1b_1+a_2b_2)}_{s_1} + \underbrace{(a_2b_1-a_1b_2)}_{s_2} i.$$

Now, we set q_1 and q_2 defined by

$$s_1 = |b|q_1 + t_1 \qquad \text{and} \qquad s_2 = |b|q_2 + t_2,$$

where $-\frac{1}{2}|b| \le t_1, t_2 \le \frac{1}{2}|b|$ by dividing in $\mathbb{Z}$. Then we take $q \coloneqq q_1 + q_2 i$ and $t \coloneqq t_1 + t_2 i$ so that

$$\frac{a}{b} = \frac{a\overline{b}}{|b|} = \frac{q|b| + t}{|b|} = q + \frac{t}{|b|}.$$

In particular, our remainder comes out to $r \coloneqq \frac{bt}{|b|} = \frac{t}{\overline{b}}$, which has norm

$$|r| = \left|\frac{t}{\overline{b}}\right| = \frac{|t|}{|b|} = \frac{t_1^2 + t_2^2}{|b|} \le \left(\frac{1}{4} + \frac{1}{4}\right)\frac{|b|^2}{|b|} < |b|.$$

This is what we wanted. ∎

So what are the primes in $\mathbb{Z}[i]$? Well, let's start with the units.

Proposition 2.75. The units of $\mathbb{Z}[i]$ are $\{\pm 1, \pm i\}$.

Proof. Of course $\{\pm 1, \pm i\}$ are units: $1 \cdot 1 = -1 \cdot -1 = i \cdot -i = 1$. Now, suppose that α is a unit so that there exists β with

$$\alpha\beta = 1.$$

Taking norms, we find that $|\alpha| \cdot |\beta| = 1$, so $|\alpha| = 1$ because the only units in $\mathbb{Z}$ are $\{\pm 1\}$. Now, letting $\alpha = x + yi$, we see $x^2 + y^2 = 1$ for size reasons. So $x^2 \le 1$, and $x \in \{-1, 0, 1\}$.

- If $x = \pm 1$, then $y = 0$, so we get $\alpha \in \{\pm 1\}$.
- If $x = 0$, then $y = \pm 1$, so we get $\alpha \in \{\pm i\}$.

This finishes the classification. ∎

Now let's classify primes.

Lemma 2.76. All primes in $\mathbb{Z}[i]$ divide a prime in $\mathbb{Z}$.

Proof. Note that any Gaussian integer α divides

$$\alpha\overline{\alpha} = |\alpha| \in \mathbb{Z},$$

so in particular any Gaussian prime π divides some integer n. If we factor n in $\mathbb{Z}$, we see

$$\pi \mid \prod_{k=1}^{N} p_k$$

for some rational primes p_k, from which it follows π divides one of the rational primes. ∎

Observe that if $x + yi$ is a Gaussian prime dividing the rational prime p, then taking norms tells us that

$$x^2 + y^2 \mid p^2.$$

In particular, $x^2 + y^2 \in \{1, p, p^2\}$, and 1 is illegal because this would imply $x + yi$ is a unit. So part of this question is if we can write p as the sum of two squares; for if we can, then

$$p = x^2 + y^2 = (x + yi)(x - yi)$$

will be the unique prime factorization of p. (We can't factor $x \pm yi$ further because, taking norms, one of the factors would have norm 1 and hence be a unit.)

So let's start factoring primes.

- We can write $2 = 1^2 + 1^2$, so $2 = (1+i)(1-i)$, but these are really the same prime because they are a multiple of i away.
- We cannot write 3 as the sum of two squares, so it is prime.
- We can write $5 = 1^2 + 2^2 = (1+2i)(1-2i)$.
- We cannot write 7 as the sum of two squares, so it is prime.

We can continue this list downwards; here is the general criterion.

Lemma 2.77. A positive prime $p \in \mathbb{Z}$ is the sum of two squares if and only if $p = 2$ or $p \equiv 1 \pmod 4$.

Proof. In one direction, if $p = x^2 + y^2$, then checking $\pmod 4$ gives $p \equiv 0, 1, 2 \pmod 4$, so p odd implies $p \equiv 1 \pmod 4$.

In the other direction, of course $2 = 1^2 + 1^2$, so we take $p \equiv 1 \pmod 4$, and we have to show that p can be written as the sum of two squares. We proceed in two steps.

1. We start by noting that $p \equiv 1 \pmod 4$ implies that $-1 \pmod p$ is a square, which is true because $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic of order $p-1$, which is divisible by 4. So, say a has order 4, and then $a^2 = -1$.
2. Now, we write $a^2 + 1 = np$ for $n \in \mathbb{Z}$, and we look at this in $\mathbb{Z}[i]$, where it factors as
$$p \mid (a+i)(a-i).$$
However, p does not divide either of those factors, so p is not prime, so p is not irreducible. So p factors as
$$p = (x+yi)(x-yi),$$
implying $p^2 = \left(x^2+y^2\right)^2$. It follows that $p = x^2 + y^2$ by positivity. ∎

Remark 2.78 (Nir). The above argument is really a more concrete version of saying that, for p an odd prime,
$$\frac{\mathbb{Z}[i]}{(p)} \cong \frac{\mathbb{Z}[x]}{(p, x^2+1)} \cong \frac{\mathbb{F}_p[x]}{(x^2+1)}.$$
Now, $p \equiv 1 \pmod 4$ if and only if x^2+1 has a root if and only if $\mathbb{F}_p[x]/\left(x^2+1\right)$ has zero-divisors if and only if $\mathbb{Z}[i]/(p)$ has zero-divisors if and only if p is not prime in $\mathbb{Z}[i]$.

Remark 2.79. This gives us an algorithm to write p as the sum of two squares. Trial and error would require about $O(\sqrt{p})$ time. Namely, we can apply the Euclidean algorithm in $\mathbb{Z}[i]$ to find the greatest common divisor of p and $a+i$, which will yield a nontrivial factor in $\mathbb{Z}[i]$ of p. (This is equivalent to doing lattice basis reduction in $\mathbb{Z}^2$ with the lattice $(a,1)\mathbb{Z} + (p,0)\mathbb{Z}$.)

Anyways, we get the following classification of primes in $\mathbb{Z}[i]$.

Theorem 2.80 (Gaussian primes). All Gaussian primes π come in one of the following forms.

- $\pi = up$ where p is a rational $3 \pmod 4$ prime.
- $\pi = a + bi$ where $p := a^2 + b^2$ is a rational prime $p = 2$ or $p \equiv 1 \pmod 4$.

Proof. We quickly check that each of the promised forms yield primes; note that none of them have norm 1, so none are units.

- $\pi = 1 + i$ is prime because it has prime norm.
- $\pi = up$ for p a rational $3 \pmod 4$ prime remains prime: if we factor $p = \alpha\beta$, then norms imply $|\alpha|, |\beta| \in \{1, p, p^2\}$. We cannot have $|\alpha| = p$ because this would make p the sum of two squares, violating Lemma 2.77.

  So one of $|\alpha|$ or $|\beta|$ is 1, implying one of α or β is a unit.
- $\pi = a + bi$ where $p := a^2 + b^2$ for p a rational $1 \pmod 4$ prime is prime because $|\pi| = p$ is prime. Indeed, if $\pi = \alpha\beta$, then $|\alpha| \cdot |\beta| = p$, so one of $|\alpha|$ or $|\beta|$ is 1.

We now check that we have all the primes. Suppose π is a rational prime; by Lemma 2.76, we may take p so that $\pi \mid p$. We have the following cases.

- If p is the sum of two squares, we can factor $p = (a + bi(a - bi)$, and we checked above that these factors are irreducible. By uniqueness, π must be one of these times a unit.
- If p is not the sum of two squares, then $p \equiv 3 \pmod 4$, so we checked above that p is prime, so π is a unit times p. ■

As an aside, we can use Gaussian integers to write general numbers as the sum of two squares.

Example 2.81. Let's do $65 = 5 \cdot 13$. We can factor $5 = (2 + i)(2 - i)$ and $13 = (3 + 2i)(3 - 2i)$. Now we have options: we could write

$$\begin{cases} 65 = (2 + i)(3 + 2i) \cdot (2 - i)(3 - 2i) = (4 + 7i)(4 - 7i) = 4^2 + 7^2, \\ 65 = (2 + i)(3 - 2i) \cdot (2 - i)(3 + 2i) = (8 - i)(8 + i) = 8^2 + 1^2. \end{cases}$$

Namely, different ways to factor 65 in $\mathbb{Z}[i]$ give different sums of squares.

2.2.6 Going Further

The theory we developed around $\mathbb{Z}[i]$ can be built for other number rings.

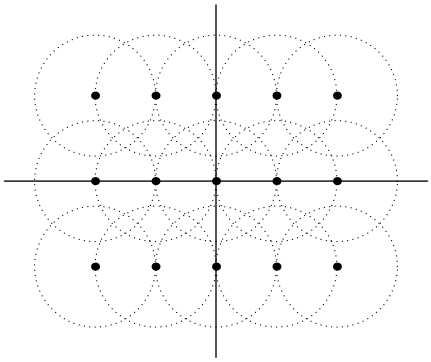
Talking $\mathbb{Z}[\sqrt{-2}]$

We start with $\mathbb{Z}[\sqrt{-2}]$.

Proposition 2.82. We have that $\mathbb{Z}[\sqrt{-2}]$ is a Euclidean domain.

Proof. Essentially the same proof as in $\mathbb{Z}[i]$ will work here. Again, embedding $\mathbb{Z}[\sqrt{-2}] \hookrightarrow \mathbb{C}$, the division algorithm comes down to showing that each $z \in \mathbb{C}$ is at most one unit away from a point on the $\mathbb{Z}[\sqrt{-2}]$ lattice.

Well, we can cover each point in $\mathbb{Z}[\sqrt{-2}]$ by some unit disk and check



So indeed, it looks like we can cover the entire plane by these disks. Again, the formal proof is somewhat technical, and I don't want to write it out again, so I won't. ■

So we get that $\mathbb{Z}[\sqrt{-2}]$ is a Euclidean domain and hence a unique factorization domain.

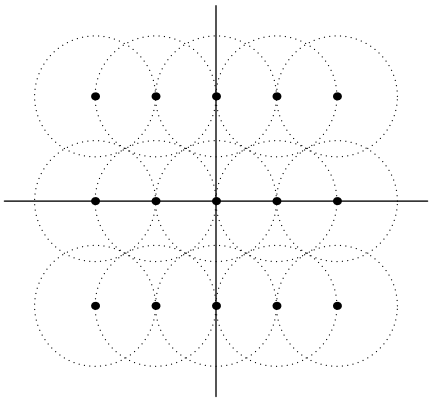
The classification of primes now has to deal with representing primes in the form $x^2 + 2y^2$. We won't write out the full proof explicitly, but here is the classification, for completeness.

Theorem 2.83 (Primes in $\mathbb{Z}[\sqrt{-2}]$)**.** A prime π in $\mathbb{Z}[\sqrt{-2}]$ comes in one of the following forms.

- $\pi = u\sqrt{2}$ for some unit u.
- $\pi = up$ for some rational prime $p \equiv \pm 3 \pmod 8$ and unit u.
- $\pi = a + b\sqrt{-2}$ where $p \coloneqq a^2 + 2b^2$ is some rational prime $\equiv \pm 1 \pmod 8$.

Talking $\mathbb{Z}[\sqrt{-3}]$

How about $\mathbb{Z}[\sqrt{-3}]$? Here, when we try to do the division algorithm and cover the plane in unit disks, it doesn't quite work. Here is the image.

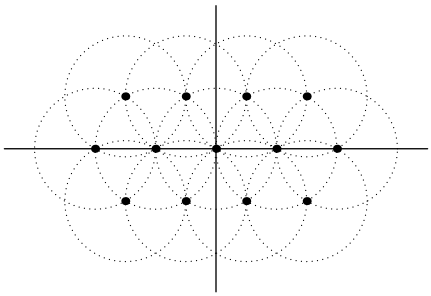


The problem here is that the closed unit disks will cover the plane, but the open ones do not; e.g., $\frac{1}{2} + \frac{1}{2}\sqrt{-3}$ is missed. This causes the entire proof to break down, and in fact $\mathbb{Z}[\sqrt{-3}]$ is not a unique factorization domain due to this problem! For example,

$$\left(1 + \sqrt{-3}\right)\left(1 - \sqrt{-3}\right) = 2 \cdot 2,$$

and we can check that all these elements are irreducible and do not differ by a unit, so this is a failure of unique factorization. In particular, these factors are irreducible but not prime.

This can be fixed by making our ring bigger: we work with $\mathbb{Z}[\omega]$ where $\omega \coloneqq \frac{1+\sqrt{-3}}{2}$ instead. This turns out to be a perfectly fine ring, isomorphic to $\mathbb{Z}[x]/\left(x^2 - x - 1\right)$. Now, when we embed $\mathbb{Z}[\omega] \hookrightarrow \mathbb{C}$, the points make a triangular lattice.



The point is that every point $z \in \mathbb{C}$ is now within one unit from a point in the lattice $\mathbb{Z}[\omega]$, so we retain our division algorithm.

Proposition 2.84. Fix $\omega := \frac{1+\sqrt{-3}}{2}$. Then $\mathbb{Z}[\omega]$ is a Euclidean domain.

Proof. This follows from the preceding discussion; as usual, imitate the proof in $\mathbb{Z}[i]$. ■

Remark 2.85 (Nir)**.** Something similar works for $\mathbb{Z}\left[\frac{1+\sqrt{-p}}{2}\right]$ where $p \in \{7, 11\}$. Notably, this does not work for $p = 19$, and it could not because we showed earlier that $\mathbb{Z}\left[\frac{1+\sqrt{-19}}{2}\right]$ is not Euclidean.

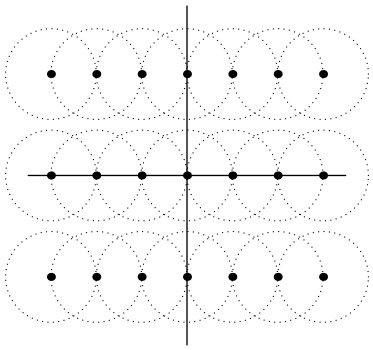
We could classify the primes in $\mathbb{Z}[\omega]$, but we did not say anything about this in class, so I will not write it out here.

Talking $\mathbb{Z}[\sqrt{-5}]$

This "make the ring bigger" algorithm does not always work. For example, in $\mathbb{Z}[\sqrt{-5}]$ we have the failure of unique prime factorization

$$\left(1+\sqrt{-5}\right)\left(1-\sqrt{-5}\right) = 2 \cdot 3.$$

For reference, here is the image of unit disks trying and failing to cover the plane.



However, $\mathbb{Z}\left[\frac{1+\sqrt{-5}}{2}\right]$ is not a good ring. One might hope that its lattice in $\mathbb{C}$ is dense enough to cover the entire plane by unit disks, but the issue is that

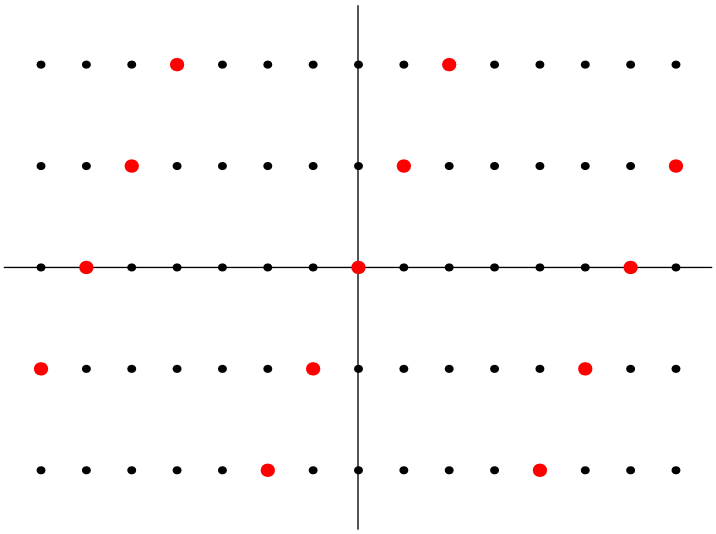
$$\mathbb{Z}\left[\frac{1+\sqrt{-5}}{2}\right] \neq \left\{a + b\frac{1+\sqrt{-5}}{2} : a, b \in \mathbb{Z}\right\},$$

because the right-hand side isn't closed under multiplication.[4]

So we have that $\mathbb{Z}[\sqrt{-5}]$ is not a principal ideal domain, and we can't easily fix it either. Well, we are promised non-principal ideal domains, so let's try to see them. Visually, $\mathbb{Z}[\sqrt{-5}]$ is a rectangular lattice; let's put our ideals in there.

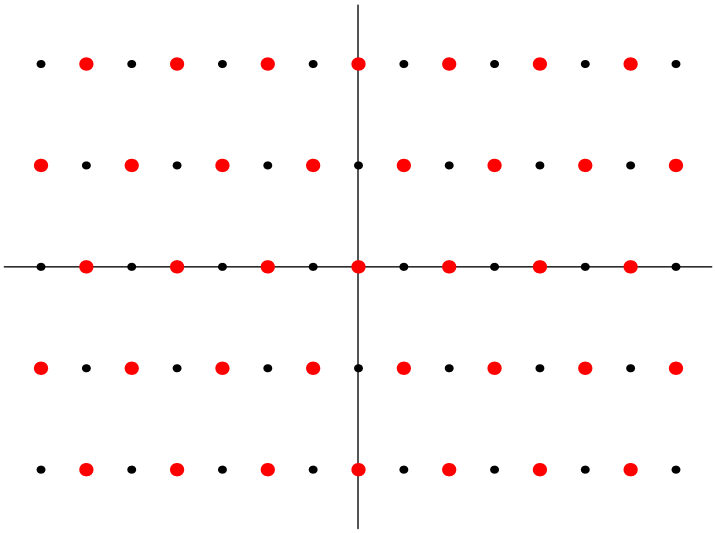
A principal ideal domain $\alpha\mathbb{Z}[\sqrt{-5}]$ will look like the $\mathbb{Z}[\sqrt{-5}]$ lattice scaled by $\sqrt{\alpha\overline{\alpha}}$ and rotated by the angle of α because that is how multiplication works in $\mathbb{C}$. For example, in red is the ideal $\left(1+\sqrt{-5}\right)$.

[4] $\left(\frac{1+\sqrt{-5}}{2}\right)^2 = \frac{-2+i\sqrt{5}}{2}$.



Note that it looks like a rectangular lattice, as we expect from $\mathbb{Z}[\sqrt{-5}]$.

In contrast, let's see a non-principal ideal domain. These tend to look like some kind of diamond lattice; for example, here is $\left(2, 1+\sqrt{-5}\right)$, in red.



Similarly, any multiple of this ideal will rotate and magnify this "diamond lattice," using the same logic as when we looked at principal ideals. So the principal and non-principal ideals really look irreconcilably different.

In $\mathbb{Z}[\sqrt{-5}]$ it happens that all ideals come with the flavor (α) or $(\alpha) \cdot \left(2, 1+\sqrt{-5}\right)$ for some $\alpha \in \mathbb{Z}[\sqrt{-5}]$. To see this, take a course on algebraic number theory. The point is that all of our ideals either look like rectangular lattices or diamond lattices. In more complicated rings, there might be more kinds of ideals.

2.3 September 30

Let's just get to the point.

2.3.1 Prime Ideals

We're having some kind of introduction to commutative algebra today.

Warning 2.86. Rings in this lecture are commutative with identity, except when explicitly said otherwise.

Let's start by talking about prime ideals.

Definition 2.87 (Prime)**.** An ideal I of a ring R is *prime* if and only if R/I is an integral domain.

Recall that an integral domain means that $ab = 0$ implies $a = 0$ or $b = 0$ in R, as well as $1 \neq 0$.[5] Similarly, we have the following.

Definition 2.88 (Maximal)**.** An ideal I of a ring R is *maximal* if and only if R/I is a field.

So we get the following for free.

Proposition 2.89. All maximal ideals are prime.

Proof. Fix I maximal in R. Then R/I is a field and hence an integral domain. ∎

These are not the usual definitions, but Professor Borcherds likes looking at quotients. Here is the usual definition of maximal.

Proposition 2.90. An ideal I of a ring R is maximal if and only if I is maximal among the set of proper ideals.

Proof. We have two implications. Note the condition $R/I \neq \{0\}$ is equivalent to $R \neq I$.

- If I is maximal among the set of ideals, we claim that R/I is a field. Indeed, for any nonzero coset $a + I \in R/I$ (i.e., $a + I \neq 0$), we note that $a \notin I$ implies
$$I \subsetneq (a) + I.$$
But then $(a) + I = R$ by maximality, so there exists $b \in R$ and $j \in I$ such that $1 = ab + j$. But then
$$(a + I)(b + I) = (ab) + I = (1 - j) + I = 1 + I,$$
so we have found a multiplicative inverse for $a + I$.
- If R/I is a field, we claim that I is maximal among the set of ideals. Suppose $J \supsetneq I$ is an ideal properly containing I, and we show $J = R$. Then there exists $a \in J \setminus I$, and we note that $a \notin I$ implies $a + I \neq 0 + I$, so there is $b \in R$ such that
$$(a + I)(b + I) = 1 + I.$$
In particular, $ab = 1 + j$ for some $j \in I \subseteq J$. Then
$$1 = ab - j \in J$$
by closure, so it follows $J = R$. ∎

And here is the usual definition for prime.

[5] Here, $0 = 1$ is equivalent to $R = \{0\}$: if $R = \{0\}$, then $1 = 0$; if $1 = 0$, then $r = 1r = 0r = 0$.

Proposition 2.91. An ideal I of a ring R is prime if and only if $I \neq R$ and $ab \in I$ implies $a \in I$ or $b \in I$.

Proof. The condition is really rephrasing that R/I is an integral domain.

- The condition that $R/I = \{0\}$ is equivalent to $a + I = 0 + I$ for all $a \in R$ is equivalent to $a \in I$ for all $r \in R$ is equivalent to $I = R$. By contraposition, $R/I \neq \{0\}$ is equivalent to I being proper.
- The condition that R/I is an integral domain is equivalent to "$(a+I)(b+I) = 0+I$ implies $a+I = 0+I$ or $b + I$" is equivalent to "$ab \in I$ implies $a \in I$ or $b \in I$," which is the listed condition. ∎

Note that there is some bad terminology here. Note that $p \in R$ is a prime element in an integral domain implies that (p) is a prime ideal, which we can check easily.[6] However, the converse is not true.

Warning 2.92. In R an integral domain, the ideal (0) is prime: if $ab = 0$, then $a = 0$ or $b = 0$ because R is an integral domain. However, 0 is not a prime element by convention, even though it generates a prime ideal.

Such is life.

2.3.2 The Spectrum

Suppose that, for a ring map $\varphi : R \to S$, suppose we have a maximal ideal $\mathfrak{m} \subseteq S$. Then is $\varphi^{-1}\mathfrak{m}$ maximal? Well, no; here is an example.

Example 2.93. Consider the embedding $\varphi : \mathbb{Z} \hookrightarrow \mathbb{Q}$. Then (0) is maximal in $\mathbb{Q}$, but $\varphi^{-1}((0)) = (0)$ is not maximal in $\mathbb{Z}$.

This is somewhat annoying. What is happening here is that we have the composite

$$R \xrightarrow{\varphi} S \twoheadrightarrow S/\mathfrak{m},$$

and we were hoping that the image $R/\varphi^{-1}\mathfrak{m}$ would be a field. But in general, subrings of fields are not fields, so have no reason to expect that $R/\varphi^{-1}\mathfrak{m}$ to be a field.

But subrings of fields do have to be integral domains, so we get the following.

Lemma 2.94. Fix $\varphi : R \to S$ a ring map with $\mathfrak{p} \subseteq S$ prime in S. Then $\varphi^{-1}\mathfrak{p}$ is prime.

Proof. We essentially repeat the above argument. Consider the composite map $\overline{\varphi}$ defined by

$$R \xrightarrow{\varphi} S \twoheadrightarrow S/\mathfrak{p}.$$

Now, $r \in R$ is in the kernel of the composite $\overline{\varphi}$ if and only if $\varphi(r) \in \mathfrak{p}$ if and only if $r \in \varphi^{-1}\mathfrak{p}$. Thus, we see

$$R/\varphi^{-1}\mathfrak{p} \cong \operatorname{im} \overline{\varphi} \subseteq S/\mathfrak{p}.$$

So now it suffices to check that a subring of an integral domain is an integral domain. Well, if $A \subseteq B$ are rings with B an integral domain, then note that $a_1 a_2 = 0$ for $a_1, a_2 \in A$ implies that $a_1 = 0$ or $a_2 = 0$ in B and hence in A. This finishes. ∎

So we have a functor from rings to sets that takes R to its set of prime ideals, which we denote $\operatorname{Spec} R$. Then we take morphisms $\varphi : R \to S$ to the map

$$\varphi^{-1} : \operatorname{Spec} S \to \operatorname{Spec} R,$$

and everything here is functorial.

[6] Note $p \mid a$ is equivalent to $a \in (p)$. So we see that "$p \mid ab$ implies $p \mid a$ or $p \mid b$" is equivalent to "$ab \in (p)$ implies $a \in (p)$ or $b \in (p)$." Additionally, p not a unit is equivalent to $(p) \neq R$.

Proposition 2.95. The function $\operatorname{Spec} : \operatorname{Ring} \to \operatorname{Set}$ taking $R \mapsto \operatorname{Spec} R$ and $\varphi : R \to S$ to $\varphi^{-1} : \operatorname{Spec} R \to \operatorname{Spec} S$ is a contravariant functor.

Proof. We see Lemma 2.94 implies that $\varphi^{-1} : \operatorname{Spec} S \to \operatorname{Spec} R$ is well-defined, so Spec is well-defined. It remains to show that Spec is functorial.

- Note that $\mathrm{id} : R \to R$ goes to $\mathrm{id}^{-1} : \operatorname{Spec} R \to \operatorname{Spec} R$, which is the identity map again because $\mathrm{id}^{-1}\mathfrak{p} = \mathfrak{p}$.
- Now suppose that $\varphi : R \to S$ and $\gamma : S \to T$ are maps of rings. Then we need to show that, given $\mathfrak{p} \in \operatorname{Spec} T$, we have
$$\varphi^{-1}\left(\gamma^{-1}\mathfrak{p}\right) = (\gamma \circ \varphi)^{-1}(\mathfrak{p}).$$
Well, $r \in \varphi^{-1}\left(\gamma^{-1}\mathfrak{p}\right)$ if and only if $\varphi(r) \in \gamma^{-1}\mathfrak{p}$ if and only if $\gamma(\varphi(r)) \in \mathfrak{p}$ if and only if $(\gamma \circ \varphi)(r) \in \mathfrak{p}$ if and only if $r \in (\gamma \circ \varphi)^{-1}\mathfrak{p}$. ∎

Remark 2.96. Grothendieck was the one who suggested that prime ideals would be better than maximal ideals because of the above functor.

2.3.3 Zariski Talk

It turns out that we can make the spectrum into a topological space. So let's think about topological spaces. If X is a topological space, we can take the set
$$\mathbb{C}(X) \coloneqq \operatorname{Mor}_{\mathrm{Top}}(X, \mathbb{C})$$
of continuous functions $X \to \mathbb{C}$. This is a ring with pointwise addition and multiplication; can we achieve all rings like this?

The answer is no, but let's try to make the answer closer to yes. Here is our question.

Question 2.97. Suppose R is a ring. How can we realize a space X so that R is the ring of continuous functions from that space to (say) $\mathbb{C}$?

Namely, fix R any ring, and we will try to find a space X with $R = \mathbb{C}(X)$. Well, we claim that any $x \in X$ induces a ring homomorphism $\mathrm{ev}_x : R \to \mathbb{C}$ by taking $\mathrm{ev}_x : r \mapsto rx$. This kernel has some nice properties.

Exercise 2.98. Fix X a topological space and $R \coloneqq \mathbb{C}(X)$. Then, given $x \in X$, we have $\mathrm{ev}_x : r \mapsto rx$ is a ring homomorphism, and the ideal $\ker \mathrm{ev}_x$ is a maximal ideal of X.

Proof. Showing that ev_x is a ring homomorphism comes down to checking the properties.

- Given $r, s \in R$, we have $\mathrm{ev}_x(r+s) = (r+s)x = rx + sx = \mathrm{ev}_x\, r + \mathrm{ev}_x\, s$ by definition of the ring addition.
- Given $r, s \in R$, we have $\mathrm{ev}_x(rs) = (rs)x = rx \cdot sx = \mathrm{ev}_x\, r \cdot \mathrm{ev}_x\, s$ again by definition of the ring multiplication.
- Lastly, the multiplicative identity in R is the $1 : z \mapsto 1$ map, which goes to $\mathrm{ev}_x\, 1 = 1$.

We now check that $\ker \mathrm{ev}_x$ is maximal. For this, we study the quotient
$$R/\ker \mathrm{ev}_x \cong \operatorname{im} \mathrm{ev}_x \subseteq \mathbb{C}.$$
So we need to show that $\operatorname{im} \mathrm{ev}_x$ is a field. We will only outline this because this is not a topology class; note it is an integral domain because it is a subring of $\mathbb{C}$.

So we have left to exhibit inverses in ev_x. Essentially, we need to know that $z \in \operatorname{im} \operatorname{ev}_x \setminus \{0\}$, then $z^{-1} \in \operatorname{im} \operatorname{ev}_x \setminus \{0\}$. However, $r \in \operatorname{im} \operatorname{ev}_x \setminus \{0\}$ implies that there exists $r \in R$ such that $rx = z$, but then

$$\left(\frac{r}{z^2}\right)(z) = \frac{1}{z},$$

so indeed, $1/z \in \operatorname{im} \operatorname{ev}_x$. The point we need to rigorize is that r/z^2 is actually a continuous function, which I assert without proof. ■

If the topology on R is good, then it turns out $\ker \operatorname{ev}_x$ will also be a closed, maximal ideal. I don't want to define a topology on $\mathbb{C}(X)$, so I won't bother elaborating on this.

Conversely, if we were to give R a nice enough topology, we could check that (closed) maximal ideals correspond to some $x \in X$. So in some sense, we could imagine recovering X as the set of closed, maximal ideals of R.

Remark 2.99. Non-closed ideals make Professor Borcherds nervous. We want ideals to have quotients, and taking quotients by non-closed spaces make the quotient space not Hausdorff, which is sad.

Now, what is the topology on X? Well, what kinds of open sets can we generate from $\mathbb{C}(X)$? The point is that we have access to continuous functions, so, say, $\mathbb{C} \setminus \{0\}$ is an open set, which makes

$$f^{-1}(\mathbb{C} \setminus \{0\})$$

an open set for any $f \in R$. These sets will turn out to make a perfectly fine basis for a topology; again details ignored.

Now, points $x \in X$ correspond with maximal ideals $\mathfrak{m} \subseteq R$, essentially behaving like kernels of special functions, so our basis element of $\{x : fx \neq 0\}$, which corresponds to $\{\mathfrak{m} : f \notin \mathfrak{m}\}$ upon associating each $x \in X$ with $\mathfrak{m} = \ker \operatorname{ev}_x$. So here is our topology: on the set of maximal ideals $\operatorname{MaxSpec} R$, we define the topology to have a basis of open sets given by

$$D_f \coloneqq \{\mathfrak{m} \in \operatorname{MaxSpec} R : f \notin \mathfrak{m}\},$$

where f is some element of R.

But this topology has the "concrete" $\mathbb{C}(X)$ part in sight! So we can do this more generally, still working with maximal ideals, creating a topology $\operatorname{MaxSpec} R$ out of the maximal ideals of our ring R. But we want to work with prime ideals

Definition 2.100 (Zariski)**.** Fix R a ring. Then we define the *Zariski* topology on $\operatorname{Spec} R$ to have open sets defined by the basis elements

$$\overline{V}(f) \coloneqq \{\mathfrak{p} \in \operatorname{Spec} R : f \notin \mathfrak{p}\}$$

for any particular $f \in R$. The closed set $V(f)$ might be called the "vanishing set" of f.

It's not too hard to check that the set $\overline{V}(f)$ actually forms a basis. Indeed, given any two $\overline{V}(f_1)$ and $\overline{V}(f_2)$, we can check that

$$\overline{V}(f_1 f_2) = \overline{V}(f_1) \cap \overline{V}(f_2).$$

In the analogy, this is saying that if f_1 and f_2 both fail to vanish at a point, then $f_1 f_2$ fails as well. Anyways, this comes down to checking that $f_1 f_2 \notin \mathfrak{p}$ if and only if $f_1 \notin \mathfrak{p}$ and $f_2 \notin \mathfrak{p}$, which is true: forwards because $\mathfrak{p}$ is an ideal and backwards because $\mathfrak{p}$ is prime.

Anyways, for all of our hard work, we get the following.

Proposition 2.101. The function $\operatorname{Spec} : \operatorname{Ring} \to \operatorname{Top}$ taking $R \mapsto \operatorname{Spec} R$ and $\varphi : R \to S$ to $\varphi^{-1} : \operatorname{Spec} S \to \operatorname{Spec} R$ is a contravariant functor.

Proof. We have shown that $R \mapsto \operatorname{Spec} R$ is well-defined, but we do not yet know that $\varphi^{-1} : \operatorname{Spec} S \to \operatorname{Spec} R$ is a continuous function given $\varphi : R \to S$ is a ring homomorphism. It suffices to check that the pre-image of a basis element $\overline{V}(f)$ is open. Namely, we want to show that

$$\left(\varphi^{-1}\right)^{-1}\left(\overline{V}(f)\right)$$

is open. Now, fix a prime $\mathfrak{q} \in \operatorname{Spec} S$ with $\mathfrak{q} \in \left(\varphi^{-1}\right)^{-1}\left(\overline{V}(f)\right)$. This condition is equivalent to $\varphi^{-1}\mathfrak{q} \in \overline{V}(f)$ is equivalent to $f \notin \varphi^{-1}\mathfrak{q}$ is equivalent to $\varphi f \notin \mathfrak{q}$. So we find that

$$\left(\varphi^{-1}\right)^{-1}\left(\overline{V}(f)\right) = \{\mathfrak{q} : \varphi f \notin \mathfrak{q}\} = \overline{V}(\varphi f).$$

This finishes. ∎

2.3.4 Making Ideals

Anyways, let's do an example.

Example 2.102. In the zero ring, we might want (0) to be a prime ideal, but it is not proper. So $\operatorname{Spec}\{0\} = \varnothing$. Thankfully, trivialities correspond to trivialities.

This is somewhat troublesome: are there any maximal ideals? The way that maximal ideals are usually constructed is by Zorn's lemma.

Axiom 2.103 (Zorn's lemma)**.** Fix X a nonempty partially ordered set. Further, suppose that any totally ordered subset has an upper bound in X. In other words, for any ascending chain

$$a_0 \leq a_1 \leq a_2 \leq \cdots,$$

there exists $a \in X$ with $a_\bullet \leq a$ for each $a_\bullet$. Then X has a maximal element.

Be careful with what "maximal" means; these need not be unique.

Definition 2.104 (Maximal)**.** An element m of a partially ordered set is *maximal* if and only if $m \leq x$ implies $m = x$ for $x \in X$.

Zorn's lemma requires the axiom of choice, which is somewhat annoying. Roughly speaking, the proof is as follows.

Proof of Axiom 2.103. The idea is to apply "transfinite induction."

- We can start with any $a_0 \in X$.
- If a_0 is maximal, we are done; otherwise, we can find $a_1 > a_0$.
- If a_1 is maximal, we are done; otherwise, we can find $a_2 > a_1$.
- Then we can continue down the line, and if we never find our element, we have an ascending chain

  $$a_0 < a_1 < a_2 < \ldots,$$

  which gives some a_ω bigger than everyone by the ascending chain condition on X.
- If a_ω is maximal, we are done; otherwise, we can find $a_{\omega+1} > a_\omega$.
- If $a_{\omega+1}$ is maximal, we are done; otherwise we can find $a_{\omega+2} > a_{\omega+1}$.
- This process could theoretically continue to all ordinals, adding 1 to the index and using the ascending chain condition to overcome limit ordinals. However, there is an absolute limit: there are ordinals with size larger than $\#X$, so the process will have to stop before then.

We note that this requires making infinitely many choices, which is where the axiom of choice is required. ■

Anyways, we can now show that maximal ideals exist.

Proposition 2.105. Fix R a nonzero ring. Then R has a maximal ideal.

Proof. We use Zorn's lemma on the collection $\mathcal{P}$ of proper ideals so that a maximal ideal will be the same as maximal element in this partially ordered set. We see $\mathcal{P}$ is nonempty because $(0) \neq R$ is a proper ideal.

So now we check the ascending chain condition on $\mathcal{P}$. Suppose that we have an ascending chain of proper ideals

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots.$$

Then we can check that

$$I \coloneqq \bigcup_{k=1}^{\infty} I_k$$

is an ideal[7] and is proper because $1 \notin I$ because $1 \notin I_k$ for any $k \in \mathbb{N}$. So I is an upper bound for our chain in $\mathcal{P}$.

To finish, we see that Zorn's lemma gives an ideal $\mathfrak{m}$ which is maximal among the set of proper ideals, which is maximal by Proposition 2.90. ■

We can actually do a little better than this. Here are some variations.

Proposition 2.106. Given any proper R-ideal J, we can find a maximal ideal $\mathfrak{m}$ containing J.

Proof. Note that Proposition 2.105 was essentially a proof that we can find a maximal ideal containing the ideal (0), so we essentially repeat the argument from there, verbatim, instead using the partially ordered set $\mathcal{P}_J$ of proper ideals which contain J. This collection is nonempty because it contains J. ■

For the next variation, we pick up the following definition.

Definition 2.107. Fix R a ring. A subset $S \subseteq R$ is *multiplicative* if and only if $1 \in S$, and S is closed under multiplication. In other words, $x, y \in S$ implies $xy \in S$.

Proposition 2.108. Suppose $S \subseteq R$ satisfies $1 \in S$ and is closed under multiplication. Then any maximal element among the proper ideals disjoint from S is prime, and such elements exist if $0 \notin S$.

Proof. To show that such an element exists, we use Zorn's lemma on the collection $\mathcal{P}_S$ of proper ideals disjoint from S. Essentially the same argument as in Proposition 2.105 will again work here: $\mathcal{P}_S$ is nonempty because $0 \notin S$, and for any ascending chain

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$$

has upper bound given by

$$I \coloneqq \bigcup_{k=1}^{\infty} I_k,$$

which is proper and disjoint from S because each of the I_N are.

Now, take $\mathfrak{p}$ in the set of proper ideals disjoint from S. To show that $\mathfrak{p}$ is prime, suppose that $a \notin \mathfrak{p}$ and $b \notin \mathfrak{p}$. Then $(a) + \mathfrak{p}$ and $(b) + \mathfrak{p}$ are ideals properly containing $\mathfrak{p}$ and hence must intersect S by maximality. But now

$$\big((a) + \mathfrak{p}\big)\big((b) + \mathfrak{p}\big) = (ab) + (a)\mathfrak{p} + (b)\mathfrak{p} + \mathfrak{p}^2 \supseteq (ab) + \mathfrak{p}$$

will also contain an element of S, so we must have $ab \notin \mathfrak{p}$ because $\mathfrak{p} \cap S = \varnothing$. This finishes. ■

In general, any maximal ideal of some collection of ideals tend to be prime.

[7] Given $a, b \in I$ and $r \in R$, we can find N such that $a, b \in I_N$. Then $ra \in I_N \subseteq I$ and $a + b \in I_N \subseteq I$ because I_N is an ideal.

2.3.5 Examples of the Spectrum

Anyways, let's do more examples of the spectrum.

Example 2.109. For R a field, the only ideals are 0 or R, so $\operatorname{Spec} R = \{(0)\}$, which is extraordinarily nice.

Example 2.110. For $R = \mathbb{C}[x]$, we note that $\mathbb{C}[x]$ is a principal ideal domain, so any ideal takes the form (f) for some $f \in \mathbb{C}[x]$. This ideal will be prime if and only if $f = 0$ or f is an irreducible polynomial by unique prime factorization. But over $\mathbb{C}$, irreducibles only look like $c(x-\alpha)$ for some $c, \alpha \in \mathbb{C}$, so we see

$$\operatorname{Spec} \mathbb{C}[x] = \{(0)\} \cup \{(x-\alpha) : \alpha \in \mathbb{C}\}.$$

Note that we can correspond $(x-\alpha) \in \mathbb{C}[x]$ with $\alpha \in \mathbb{C}$ in some sense.

The above example seems to recover the complex plane from $\operatorname{Spec} \mathbb{C}[x]$, but the topology on $\operatorname{Spec} \mathbb{C}[x]$ is very bad. Recall that our basis consisted of closed sets

$$V(f) = \{\mathfrak{p} \in \operatorname{Spec} \mathbb{C}[x] : f \in \mathfrak{p}\},$$

for various $f \in \mathbb{C}[x]$. Unravelling this for our example, we see that $f \in (0)$ is equivalent to $f = 0$, and $f \in (x-\alpha)$ is equivalent to $x - \alpha \mid f$ is equivalent to $f(\alpha) = 0$. So our vanishing sets are as follows.

- If $f = c\prod_{k=1}^{n}(x-\alpha_k)$ for $c, \{\alpha_k\}_{k=1}^{n} \subseteq \mathbb{C}$, then $V(f) = \{(x-\alpha_1), \dots, (x-\alpha_n)\}$, which corresponds to some finite set of points in $\mathbb{C}$.
- If $f \equiv 0$, then $V(f) = \operatorname{Spec} \mathbb{C}[x]$, which corresponds to all of $\mathbb{C}$.
- If $f \equiv c$ for $c \in \mathbb{C}^\times$, then $V(f) = \varnothing$.

In particular, we have the cofinite topology, which isn't even Hausdorff. In general, most of our spectrums are sad like this.

Example 2.111. For $R = \mathbb{Z}$, we have

$$\operatorname{Spec} \mathbb{Z} = \{(0)\} \cup \{(p) : p \text{ prime}\}.$$

Again, let's think about the topology. Well, given $n \in \mathbb{Z}$, a prime p "vanishes" at n if and only if $n \notin (n)$, which corresponds to $p \nmid n$. So we find that our vanishing sets are empty ($n = 1$), some finite set of primes ($n = \prod p_\bullet$), or everything ($n = 0$).

Remark 2.112. In the above examples, we tend to have sane primes in addition to the weird prime ideal (0) which lives inside of all the other primes. This is an example of a "generic" point.

2.3.6 Localization

Our first example of localization is as follows.

Example 2.113. Take $\mathbb{Z}$ and force all nonzero elements to have inverses by making it $\mathbb{Q}$.

This turns out to be a very useful, general operation. Here is our idea.

Idea 2.114. Fix R a ring with S a subset, we want to define a ring $S^{-1}R$ to be R where the elements of S have inverses.

This can be done, essentially, by taking S and looking at the quotient

$$R\left[S^{-1}\right] = \frac{R\left[\{t_s\}_{s\in S}\right]}{(\{(st_s - 1)\}_{s\in S})}.$$

This is a very nice, universal way of forcing inverses, but we have no idea what it looks like. For example, is the canonical embedding $R \hookrightarrow R\left[S^{-1}\right]$ is injective? Well, it often isn't. Is $R\left[S^{-1}\right]$ even nonzero? These questions are not at all obvious because we have a huge ring modded out by a huge ideal.

Ignoring Zero-Divisors

So we want a little more control over our inverses, in the same way that we did for $\mathbb{Q}$. So let's try to imitate $\mathbb{Q}$. Here is our first attempt:

Definition 2.115 (Localization for integral domains)**.** Fix R an integral domain and S a multiplicative subset not containing 0. Then we define the *localization* $S^{-1}R$ to be the set of pairs (r, s) (denoted r/s) modulo the equivalence relation $r_1/s_1 \equiv r_2/s_2$ if and only if $r_1s_2 = r_2s_1$. Then we define addition by

$$\frac{r_1}{s_1} + \frac{r_2}{s_2} := \frac{r_1s_2 + r_2s_1}{s_1s_2} \qquad \text{and} \qquad \frac{r_1}{s_1} \cdot \frac{r_2}{s_2} := \frac{r_1r_2}{s_1s_2}.$$

We note that the multiplication in the denominator is why we want S to be a multiplicative subset.

There are lots of things to check here; we will outline the things we have to check.

Lemma 2.116. Fix R a ring and S a multiplicative subset not containing any zero-divisors. The relation $\equiv$ above on $R \times S$ is an equivalence relation.

Proof. We check the conditions one at a time.

- Reflexive: for $r/s \in S^{-1}R$, we note that $r/s \equiv r/s$ is implied by $rs = rs$.
- Symmetric: note that $r_1/s_1 \equiv r_2/s_2$ implies $r_1s_2 = r_2s_1$ implies $r_2s_1 = r_1s_2$ implies $r_2/s_2 \equiv r_1/s_2$.
- Transitive: note that $r_1/s_1 \equiv r_2/s_2$ and $r_2/s_2 \equiv r_3/s_3$ implies $r_1s_2 = r_2s_1$ and $r_2s_3 = r_3s_2$. Then

$$s_2 \cdot r_1s_3 = (r_1s_2)s_3 = (r_2s_1)s_3 = (r_2s_3)s_1 = (r_3s_2)s_1 = s_2 \cdot r_3s_1,$$

  so we are done after applying the cancellation law to get rid of the s_2. ∎

Remark 2.117. We had to use that R is an integral domain in the transitivity check to cancel s_2. Namely, transitivity need not be transitive when S has zero-divisors. For example, in $R = \mathbb{Z}/12\mathbb{Z}$, we can take $S = \{1, 2, 4, 8\}$ so that

$$\frac{3}{2} = \frac{6}{4} = \frac{3}{4},$$

but $\frac{3}{2} \neq \frac{3}{4}$.

Lemma 2.118. Fix R an integral domain and S a multiplicative subset not containing any zero-divisors. Then $S^{-1}R$ is a ring.

Proof. This is identical to the proof that $\mathbb{Q}$ is a ring: we apply brute force to each of the checks to be a ring, and they all work. We won't write them out here. ∎

Lemma 2.119. Fix R an integral domain and S a multiplicative subset not containing any zero-divisors. Then the canonical map $R \hookrightarrow S^{-1}R$ is injective.

Proof. Suppose $r_1, r_2 \in R$ have $r_1/1 = r_2/1$. Then $r_1 = 1r_1 = 1r_2 = r_2$ by definition. ∎

Not Ignoring Zero-Divisors

What if our multiplicative set S does have zero-divisors? To deal with the problem, we set

$$I \coloneqq \{a \in R : as = 0 \text{ for some } s \in S\}.$$

We quickly check that I is an ideal.

- If $a_1, a_2 \in I$, then $a_1 s_1 = 0$ and $a_2 s_2 = 0$ for some $s_1, s_2 \in S$. Then
$$(a_1 a_2)(s_1 + s_2) = (a_1 s_1)s_2 + (a_2 s_2)s_1 = 0s_2 + 0s_1 = 0.$$
- If $a \in I$ and $r \in R$, then $as = 0$ for some $s \in S$, so $(ar)s = a(rs) = a0 = 0$.

The point is that, in R/I, the set S/I has no zero-divisors because we have decided to kill all problematic elements: if $(r + I)(s + I) = I$, then $rs \in I$, so there is $s' \in S$ such that $rss' = 0$, so $r(ss') = 0$, so $r \in I$ because S is multiplicative. So now we can define

$$S^{-1}R \coloneqq (S/I)^{-1}(R/I).$$

This construction took two steps: take the quotient and then localize. However, it is possible to do this by making a trickier equivalence relation.

Definition 2.120 (Localization)**.** Fix R a ring and S a multiplicative subset. Then we define the *localization* $S^{-1}R$ to be the set of pairs (r, s) (denoted r/s) modulo the equivalence relation $r_1/s_1 \equiv r_2/s_2$ if and only if $sr_1s_2 = sr_2s_1$ for some $s \in S$. Then we define addition by

$$\frac{r_1}{s_1} + \frac{r_2}{s_2} \coloneqq \frac{r_1 s_2 + r_2 s_1}{s_1 s_2} \qquad \text{and} \qquad \frac{r_1}{s_1} \cdot \frac{r_2}{s_2} \coloneqq \frac{r_1 r_2}{s_1 s_2}.$$

We note that the multiplication in the denominator is why we want S to be a multiplicative subset.

We check that the trickier equivalence relation creates the same localization as $(S/I)^{-1}(R/I)$.

Proposition 2.121. Fix R, S, I as above. Then, given pairs $(r_1, s_1), (r_2, s_2) \in R \times S$ has the following equivalent relations.

- We write $(r_1, s_1) \equiv (r_2, s_2)$ if and only if there exists $s \in S$ such that $sr_1s_2 = sr_2s_1$.
- We write $(r_1, s_1) \equiv' (r_2, s_2)$ if and only if $(r_1 + I)(s_2 + I) = (r_2 + I)(s_1 + I)$.

In particular, $\equiv$ is an equivalence relation because $\equiv'$ is.

Proof. Suppose that $(r_1, s_1) \equiv (r_2, s_2)$, which is equivalent to the existence of $s \in S$ such that $sr_1s_2 = sr_2s_1$. Then this is equivalent to

$$s(r_1 s_2 - r_2 s_1) = 0,$$

which is equivalent to $r_1 s_2 - r_2 s_1 \in I$. Moving arithmetic into R/I, we have $r_1 s_2 - r_2 s_1 \in I$ is equivalent to

$$(r_1 + I)(s_2 + I) = (r_2 + I)(s_1 + I)$$

after some rearranging, and this is equivalent to $(r_1, s_1) \equiv' (r_2, s_2)$. ∎

So, for example, we get the following for free.

Lemma 2.122. Fix R, S, I as above. Then $S^{-1}R$ is a well-defined ring where the map $R \to S^{-1}R$ has kernel I.

Proof. Proposition 2.121 tells us that $S^{-1}R$ is in bijection with $(S/I)^{-1}(R/I)$ by sending

$$\frac{r}{s} \longmapsto \frac{r+I}{s+I}.$$

(Explicitly, the equivalence relations on $R \times S$ are the same, so we are getting the same equivalence classes.) Because the addition and multiplication laws are defined in the same way, we in fact have that $S^{-1}R \cong (S/I)^{-1}(R/I)$.

So our work with S not containing zero-divisors tells us that $S^{-1}R$ is a ring for free, and the kernel of $R \to S^{-1}R$ is the kernel of the composite map

$$R \twoheadrightarrow R/I \to (S/I)^{-1}(R/I).$$

Here the first map has kernel I and the second map has trivial kernel, so the composite's kernel is I. ■

Remark 2.123. Professor Borcherds does not like the one-step construction because the equivalence relation is somewhat unintuitive.

2.3.7 Localizing at a Prime

Let's try and use localization for something. Given a space X, we can look back at $\mathbb{C}(X)$ and note that we actually have lots of possible functions: for each open set $U \subseteq X$, we can define the continuous functions $U \to \mathbb{C}$ as the space $\mathbb{C}(U)$. This has some nice properties.

- Given two open sets $U_1 \subseteq U_2$, we can take functions in $\mathbb{C}(U_2)$ and "restrict" them to $\mathbb{C}(U_1)$. This gives us a function $\mathrm{Res}_{U_1,U_2} : \mathbb{C}(U_2) \to \mathbb{C}(U_1)$.
- Given continuous functions on a family $\{U_\alpha\}_{\alpha\in\lambda}$ of open sets such that the continuous functions behave nicely with restriction, we can build a larger continuous function on $\bigcup_{\alpha\in\lambda} U_\alpha$.
- Given an open cover of U named $\{U_\alpha\}_{\alpha\in\lambda}$, then if two continuous functions are identically equal on each U_α, then they are equal on U.

In other words, $\mathbb{C}(-)$ is a sheaf of rings, but we won't use this.

Now, back in the analogy, take R a ring with $\operatorname{Spec} R$ the spectrum, as usual. Now, take U to be an open subset of $\operatorname{Spec} R$, and we want to imagine what the analogue of $\mathbb{C}(U)$ should be. To start, we should take U as a basis element $\overline{V}(f)$.

Let's check what $\overline{V}(f)$ means back in our example. Well, if U is the open sets where a fixed function $f \in \mathbb{C}(U)$ doesn't vanish, then really the only truly obvious function we have added here is f^{-1}. So in general, we define the ring of $U = \overline{V}(f)$ to be

$$R\left(f^{-1}\right),$$

where we are just inverting out by that function f. One might think that we've added different functions with the extra power to invert at a point, but, algebraically speaking, we don't have access to these.

Remark 2.124. The secret to making algebraic geometry easier is to ignore all open sets which are not the basis elements $\overline{V}(f)$.

Anyways, here is an example.

Example 2.125. Suppose we live in $\mathbb{Z}$, but we want to kill the prime 2 because 2 has been really messing up your day. Well, the idea is to take the open set

$$U = \overline{V}(2) = \{\mathfrak{p} : 2 \notin \mathfrak{p}\}$$

so that we get $\mathbb{Z}[1/2]$, effectively killing the prime (2).

Example 2.126. Conversely, suppose we live in $\mathbb{Z}$, and we want to focus on 2 alone. We start by ignoring $3, 5, 7$, which means we want the open set

$$\overline{V}(105) = \{\mathfrak{p} : 105 \notin \mathfrak{p}\} = \{\mathfrak{p} : 3 \notin \mathfrak{p} \text{ and } 5 \notin \mathfrak{p} \text{ and } 7 \notin \mathfrak{p}\},$$

and we get $\mathbb{Z}[1/3, 1/5, 1/7]$. To keep killing more primes, we take the direct limit of this process for all open sets U containing 2. At the end of this process, we get

$$\mathbb{Z}[1/3, 1/5, 1/7, \ldots],$$

which is called the localization of $\mathbb{Z}$ at the prime of (2).

The above example can be generalized.

Definition 2.127 (Localization at a prime)**.** Fix R a ring and $\mathfrak{p}$ a prime. Then $S \coloneqq R \setminus \mathfrak{p}$ is multiplicative, so we define $R_{\mathfrak{p}} \coloneqq S^{-1}R$ to be the *localization at* $\mathfrak{p}$.

Theme 3
Module Monologue

It is my experience that proofs involving matrices can be shortened by 50% if one throws the matrices out.

—Emil Artin

3.1 October 5

Are you feeling nervous? Are you having fun?

3.1.1 Modules

Today we talk about modules. Here is the definition.

Definition 3.1 (Module)**.** A (left) module M over a ring R is an abelian group with a "(left) ring action." In other words, we have an operation $\cdot : R \times M \to M$ satisfying some linearity axioms, as follows; fix $r, s \in R$ and $m, n \in M$.

- Distributive: $r(m+n) = rm + rn$.
- Distributive: $(r+s)m = rm + sm$.
- Associative: $(rs)m = r(sm)$.
- Identity: $1_R m = m$.

As usual, there is also a notion of right modules and two-sided modules, and this distinction matters for non-commutative rings.

Example 3.2. Vector spaces are modules over fields. The field action is the scalar multiplication.

Example 3.3. Abelian groups are modules over $\mathbb{Z}$. The $\mathbb{Z}$-action on an abelian group G is exponentiation by $n \cdot g \mapsto g^n$.

Example 3.4. Ideals are equivalent to R-submodules of R. Indeed, we have that left/right/two-sided ideals are left/right/two-sided R-submodules. I will not do this check because I am lazy; the main point is that closure of I under multiplication by R is the same thing as closure of I under the R-action.

We also have maps between modules.

Definition 3.5 (Module homomorphism)**.** Fix M and N left modules over R. Then $\varphi \in \mathrm{Hom}_R(M,N)$ is a group homomorphism $\varphi : M \to N$ such that

$$(r_1 m_1 + r_2 m_2)\varphi = r_1(m_1\varphi) + r_2(m_2\varphi)$$

where $r_1, r_2 \in R$ and $m_1, m_2 \in M$.

Explicitly, if M and N are left modules, then $\varphi \in \mathrm{Hom}_R(M,N)$ should be written on the right because the linearity condition requires

$$(rm)\varphi = r(m\varphi)$$

for $r \in R, m \in M, \varphi \in \mathrm{Hom}_R(M,N)$. What is bad here is that writing on the other side gives $r\varphi(m) = \varphi(rm)$, which requires a switching of variables. This distinction matters for non-commutative rings, but I will largely ignore this and continue to write functions on the left.

We note the following.

Proposition 3.6. If M and N are (left) modules over a commutative ring R, then $\mathrm{Hom}_R(M,N)$ is an R-module, where the action is

$$(r\varphi)(m) \coloneqq r \cdot \varphi(m).$$

Proof. We have that $\mathrm{Hom}_R(M,N)$ is an abelian group where addition is done pointwise; indeed, it is a subgroup of $\mathrm{Hom}(M,N)$ closed under the subgroup test because $\varphi, \gamma \in \mathrm{Hom}_R(M,N)$ still has $(\varphi - \gamma)$ an R-module homomorphism.

Lastly we have to check that the R-action makes $\mathrm{Hom}_R(M,N)$ into an R-module. This is relatively unenlightening. For example, we can check that

$$\big((r_1 + r_2)\varphi\big)(m) = (r_1 + r_2)\varphi(m) = r_1\varphi(m) + r_2\varphi(m) = (r_1\varphi + r_2\varphi)(m)$$

for any $r_1, r_2 \in R, \varphi \in \mathrm{Hom}_R(M,N), m \in M$. I won't do the other checks out of lazy. ∎

We remark that if R is not commutative, then $\mathrm{Hom}_R(M,N)$ is merely an abelian group, not an R-module.

We continue with our examples.

Definition 3.7 (Opposite ring)**.** Fix R a ring. Then we define the *opposite ring* R^{op} to have elements r^{op} for $r \in R$ where our operations are defined by

$$r^{\mathrm{op}} + s^{\mathrm{op}} \coloneqq (r+s)^{\mathrm{op}} \qquad \text{and} \qquad r^{\mathrm{op}} \cdot s^{\mathrm{op}} \coloneqq (sr)^{\mathrm{op}}$$

This forms a ring, which can be checked by hand. In other words, the underlying abelian group is the same for R and R^{op}, but the ring multiplication is flipped.

The point of the above definition is the following example.

Example 3.8. If M is a left R-module, then M is a right R^{op}-module by $m \cdot r^{\mathrm{op}} := r \cdot m$. All of the distributivity axioms come for free from M being a left R-module, and the associativity axiom holds because

$$(mr^{\mathrm{op}})\, s^{\mathrm{op}} = s(r(m)) = (sr)m = m(sr)^{\mathrm{op}} = m\left(r^{\mathrm{op}} s^{\mathrm{op}}\right).$$

Warning 3.9. Left and right modules can be very different for a particular ring, namely when non-commutative.

Explicitly, some rings have $R \cong R^{\mathrm{op}}$, but not all. Of course this is true when R is commutative; here is a less trivial example.

Proposition 3.10. Fix R a ring and G a group so that $R[G]$ is the group ring. Then $R[G] \cong R[G]^{\mathrm{op}}$.

Proof. The idea is to consider the map $G \to R[G]^{\mathrm{op}}$ by

$$g \mapsto 1_R g^{-1}$$

and use the universal property to lift this to a map $\varphi : R[G] \to R[G]^{\mathrm{op}}$. Explicitly,

$$\varphi\left(\sum_{g\in G} r_g g\right) := \sum_{g\in G} r_g g^{-1}.$$

We now check that φ is an isomorphism of rings. This is not terribly interesting, but we will do it anyways.

- We see that φ preserves addition because

$$\varphi\left(\sum_{g\in G} r_g g + \sum_{g\in G} s_g g\right) = \varphi\left(\sum_{g\in G} (r_g + s_g) g\right) = \sum_{g\in G} (r_g + s_g) g^{-1} = \sum_{g\in G} r_g g^{-1} + \sum_{g\in G} s_g g^{-1},$$

  which is what we need.

- We see that φ preserves multiplication because

$$\varphi\left(\sum_{x\in G} r_x x \times \sum_{y\in G} s_y y\right) = \varphi\left(\sum_{g\in G}\left(\sum_{xy=g} r_x s_y\right) g\right) = \sum_{g\in G}\left(\sum_{xy=g} r_x s_y\right) g^{-1},$$

  but in the opposite ring, we have

$$\varphi\left(\sum_{y\in G} s_y y\right)\varphi\left(\sum_{x\in G} r_x x\right) = \left(\sum_{y\in G} s_y y^{-1}\right)\left(\sum_{x\in G} r_x x^{-1}\right) = \sum_{g\in G}\left(\sum_{xy=g} r_x s_y\right)\left(y^{-1}x^{-1}\right),$$

  which is indeed $\sum_{g\in G}\left(\sum_{xy=g} r_x s_y\right) g^{-1}$.

- We see that φ preserves identity because $\varphi(1e) = 1e^{-1} = 1e$.

- We see that φ is surjective because, for any $\sum r_g g \in R[G]^{\mathrm{op}}$, we have

$$\varphi\left(\sum_{g\in G} r_{g^{-1}} g\right) = \sum_{g\in G} r_{g^{-1}} g^{-1} = \sum_{g\in G} r_g g.$$

- We see that φ is injective because, if $\sum r_g g^{-1} = \sum s_g g^{-1}$, then $r_g = s_g$ for each g. ∎

3.1.2 Hom Is Left Exact

Suppose that we have an exact sequence of R-modules as follows.

$$0 \longrightarrow A \xrightarrow{\varphi} B \xrightarrow{\gamma} C \longrightarrow 0$$

Given a a fixed R-module M, we can look at the following sequence.

$$0 \longrightarrow \mathrm{Hom}(M, A) \xrightarrow{\varphi\circ -} \mathrm{Hom}(M, B) \xrightarrow{\gamma\circ -} \mathrm{Hom}(M, C) \longrightarrow 0$$

Most of this sequence is exact but not all.

Proposition 3.11. Suppose that we have an exact sequence of R-modules

$$0 \longrightarrow A \xrightarrow{\varphi} B \xrightarrow{\gamma} C \longrightarrow 0$$

Then, for any R-module M,

$$0 \longrightarrow \mathrm{Hom}(M, A) \xrightarrow{\varphi\circ -} \mathrm{Hom}(M, B) \xrightarrow{\gamma\circ -} \mathrm{Hom}(M, C)$$

is exact.

Proof. We have two things to check

- Exact at $\mathrm{Hom}(M, A)$: we have to show that $\mathrm{Hom}(M, A) \to \mathrm{Hom}(M, B)$ by $f \mapsto \varphi \circ f$ is injective. Well, suppose that $\varphi \circ f_1 = \varphi \circ f_2$ for $f_1, f_2 \in \mathrm{Hom}(M, A)$. Then for any $m \in M$, we have

  $$\varphi(f_1 m) = \varphi(f_2 m),$$

  so $f_1(m) = f_2(m)$ because φ is injective. So indeed, $f_1 = f_2$.

- Exact at $\mathrm{Hom}(M, B)$: we have to show that the kernel of $\mathrm{Hom}(M, B) \to \mathrm{Hom}(M, C)$ by $g \mapsto \gamma \circ g$ is exactly the image of $f \mapsto \varphi \circ f$.

  In one direction, if $\varphi \circ f$ is in the image of $\mathrm{Hom}(M, A) \to \mathrm{Hom}(M, B)$, then for any $m \in M$ we have

  $$(\gamma \circ \varphi \circ f)(m) = (\gamma \circ \varphi)(fm) = 0(fm) = 0,$$

  so indeed, $\varphi \circ f$ is in the kernel of $\gamma \circ -$.

  In the other direction, fix any $g \in \mathrm{Hom}(M, B)$ in the kernel of $\gamma \circ -$ so that $\gamma \circ g = 0$. This is equivalent to, for any $m \in M$, having

  $$\gamma(g(m)) = 0,$$

  which is equivalent to $g(m) \in \ker \gamma$, which is equivalent to $g(m) \in \mathrm{im}\, \varphi$ by exactness. Now, φ is injective, so each $g(m)$ has a unique lift into A, letting us define

  $$f(m) \coloneqq \varphi^{-1}(g(m)).$$

  There is some check here to make sure $f \in \mathrm{Hom}(M, A)$, which is not very interesting.[1] The point is that $g = \varphi \circ f$, so g is in the image of $\varphi \circ -$. ∎

However, $\mathrm{Hom}(M, -)$ does not always produce sequences always exact at the end.

[1] Note $f(r_1 m_1 + r_2 m_2)$ is the unique element such that $\varphi(f(r_1 m_1 + r_2 m_2)) = g(r_1 m_1 + r_2 m_2)$, but because $g(r_1 m_1 + r_2 m_2) = r_1 g(m_1) + r_2 g(m_2)$, we see that $r_1 f(m_1) + r_2 f(m_2)$ goes to the same place under φ.

Example 3.12. Consider the short exact sequence of $\mathbb{Z}$-modules

$$0 \to \mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0$$

and take $M := \mathbb{Z}/2\mathbb{Z}$. Then applying $\operatorname{Hom}(\mathbb{Z}/2\mathbb{Z}, -)$, we note there are no nontrivial maps $\mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}$ (the image of 1 must double to 0, but the only element of additive order dividing 2 is 0 itself).

On the other hand, $\operatorname{Hom}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z}$ by sending $1 \mapsto 0$ or $1 \mapsto 1$, so the sequence

$$0 \to \operatorname{Hom}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}) \to \operatorname{Hom}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}) \to \operatorname{Hom}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}) \to 0$$

becomes the sequences

$$0 \to 0 \to 0 \to \mathbb{Z}/2\mathbb{Z} \to 0,$$

which is not exact at the end, sadly.

Remark 3.13. The high-level way to see Proposition 3.11 is that $\operatorname{Hom}(M, -)$ is right adjoint (to tensor), so Hom preserves limits, so Hom is left exact.

Similarly, we can continue fixing an R-module N and apply $\operatorname{Hom}(-, N)$. This turns the sequence

$$0 \longrightarrow A \xrightarrow{\varphi} B \xrightarrow{\gamma} C \longrightarrow 0$$

into the sequence

$$0 \longleftarrow \operatorname{Hom}(A, N) \xleftarrow{-\circ\varphi} \operatorname{Hom}(B, N) \xleftarrow{-\circ\gamma} \operatorname{Hom}(C, N) \longleftarrow 0$$

where the arrows are still composition as labeled.

Warning 3.14. The above sequence of maps has the arrows reversed.

In this case, we have the following.

Proposition 3.15. Suppose that we have an exact sequence of R-modules

$$0 \longrightarrow A \xrightarrow{\varphi} B \xrightarrow{\gamma} C \longrightarrow 0$$

Then, for any R-module N, the sequence

$$\operatorname{Hom}(A, N) \xleftarrow{-\circ\varphi} \operatorname{Hom}(B, N) \xleftarrow{-\circ\gamma} \operatorname{Hom}(C, N) \longleftarrow 0$$

is exact.

Proof. This is essentially the same as Proposition 3.11. We have two things to check.

- Exact at $\operatorname{Hom}(C, N)$: essentially, we have to show that the kernel of $- \circ \gamma$ is trivial. So suppose that we have $f \in \operatorname{Hom}(C, N)$ such that $f \circ \gamma = 0$. The, for any $c \in C$, we note that the surjectivity of γ promises $b \in B$ such that $\gamma b = c$, implying
$$f(c) = (f \circ \gamma)(b) = 0,$$
so f is the zero map. So indeed, $\ker(f \mapsto f \circ \gamma) = \{0\}$.
- Exact at $\operatorname{Hom}(B, N)$: we have to show that a map $g \in \operatorname{Hom}(B, N)$ has $g \circ \varphi = 0$ if and only if $g = f \circ \gamma$ for some $f \in \operatorname{Hom}(C, N)$.

In one direction, suppose that $g = f \circ \gamma$ for some $f \in \mathrm{Hom}(M, C)$. Then we have that

$$g \circ \varphi = (f \circ \gamma) \circ \varphi = f \circ (\gamma \circ \varphi) = f \circ 0 = 0,$$

where $\gamma \circ \varphi = 0$ because $\mathrm{im}\,\varphi \subseteq \ker \gamma$ by exactness.

In the other direction, suppose that $g \circ \varphi = 0$. Then, $g(\mathrm{im}\,\varphi) = \{0\}$, so $\mathrm{im}\,\varphi \subseteq \ker g$. In particular, $\ker \gamma \subseteq \mathrm{im}\,\varphi$ by exactness, so $\ker \gamma \subseteq \ker g$. It follows that $g : B \to N$ can be made into a well-defined map

$$\overline{g} : B/\ker \gamma \to N$$

such that $B \to N$ is the same as $B \twoheadrightarrow B/\ker \gamma \to N$. Now, $\mathrm{im}\,\gamma \cong B/\ker \gamma$, so we have the sequence of maps

$$B \overset{\gamma}{\twoheadrightarrow} \mathrm{im}\,\gamma \cong B/\ker \gamma \overset{\overline{g}}{\to} N$$

whose composite is equal to g by pushing through elements. Letting $f : C \to N$ be the composite $C = \mathrm{im}\,\gamma \cong B/\ker \gamma \to N$, we find that $g = f \circ \gamma$, which is exactly what we wanted. ■

And again, we don't have to be fully exact.

Example 3.16. Consider the short exact sequence of $\mathbb{Z}$-modules

$$0 \to \mathbb{Z} \overset{\times 2}{\to} \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0$$

and take $N \coloneqq \mathbb{Z}/2\mathbb{Z}$. Then applying $\mathrm{Hom}(-, N)$, we see that $\mathrm{Hom}(\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z}$ (we send $1 \mapsto 0$ or $1 \mapsto 1$), but also $\mathrm{Hom}(\mathbb{Z}/2\mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z}$ as discussed last time.

So the sequence

$$0 \leftarrow \mathrm{Hom}(\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}) \leftarrow \mathrm{Hom}(\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}) \leftarrow \mathrm{Hom}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}) \leftarrow 0$$

is

$$0 \leftarrow \mathbb{Z}/2\mathbb{Z} \leftarrow \mathbb{Z}/2\mathbb{Z} \leftarrow \mathbb{Z}/2\mathbb{Z} \leftarrow 0,$$

which cannot be exact for size reasons: the left end would have to have size $2/2 = 1$. And indeed, we can verify that the $\times 2$ mapping is losing surjectivity at the end.

The lack of these exactness turns out to be a huge problem in algebra. The entire field of homological algebra is dedicated to fixing this problem.

Remark 3.17. The short exact sequence

$$0 \to \mathbb{Z} \to \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0$$

is a good universal counterexample to various statements.

3.1.3 Free Modules

We have the following definition.

Definition 3.18 (Free)**.** An R-module M is *free* if it is the direct sum of some number of copies of R.

We have the following sequence of propositions.

Proposition 3.19. Suppose that we have a split short exact sequence of R-modules

$$0 \to A \to B \to C \to 0.$$

Then $B \cong A \oplus C$, canonical up to the choice of lift $C \to B$.

Proof. This requires some care. Label $\iota : A \to B$, $\pi : B \to C$, and $\rho : C \to B$, where $\rho \circ \pi = \mathrm{id}_C$ is our lift of π. Here is the diagram.

$$0 \longrightarrow A \xrightarrow{\iota} B \underset{\rho}{\overset{\pi}{\rightleftarrows}} C \longrightarrow 0$$

Now we note that we have a map $\varphi : A \oplus C \to B$ by

$$\varphi : (a, c) \mapsto \iota a + \rho c.$$

We show that this map is an R-module isomorphism. Note that it is an R-module homomorphism by using the universal property of $A \oplus C$ on the morphisms $\iota : A \to B$ and $\rho : C \to B$, so the main obstruction is showing the isomorphism. We have two things to check.

- We show that φ is injective. Indeed, suppose that $\varphi((a, c)) = 0$ so that $\iota a + \rho c = 0$ and

  $$\iota(a) = \rho(-c).$$

  Applying π to both sides implies that $0 = -c$ because $\operatorname{im} \iota \subseteq \ker \pi$ by exactness (!). Thus, $c = 0$, implying $\iota(a) = 0$, so $a = 0$ because $\ker \iota = \{0\}$. Thus, $\ker \varphi = \{(0, 0)\}$.
- We show that φ is surjective. Indeed, fix any $b \in B$. We start by taking $c \coloneqq \pi b$ and observe that

  $$\pi(b - \rho c) = \pi(b) - (\pi \circ \rho)(c) = c - c = 0,$$

  so $b - \rho c \in \ker \pi$. But $\ker \pi \subseteq \operatorname{im} \iota$ by exactness (!), so $b - \rho c = \iota a$ for some $a \in A$. Thus,

  $$\varphi : (a, c) \mapsto \iota a + \rho c = b,$$

  which is what we needed. ∎

The reason we bring this up is to talk about free modules.

Proposition 3.20. If C is a free R-module in the short exact sequence

$$0 \to A \to B \to C \to 0,$$

then this short exact sequence splits so that $B \cong A \oplus C$.

Proof. Label our short exact sequence as follows.

$$0 \longrightarrow A \xrightarrow{\iota} B \xrightarrow{\pi} C \longrightarrow 0$$

Now, find a basis $\{c_\alpha\}_{\alpha \in \lambda}$ for C, and lift each element to $\{b_\alpha\}_{\alpha \in \lambda}$ in B along π so that $\pi : b_\alpha \mapsto c_\alpha$. This induces a map $\rho : C \to B$ defined by

$$\rho\left(\sum_{\alpha \in \lambda} r_\alpha c_\alpha\right) \coloneqq \sum_{\alpha \in \lambda} r_\alpha b_\alpha,$$

where we might have to mumble something about the universal property of free objects. (Here, all but finitely many of the r_α vanish.) Then we note that, for any element $\sum_\alpha r_\alpha c_\alpha \in C$, we have

$$(\pi \circ \rho)\left(\sum_{\alpha \in \lambda} r_\alpha c_\alpha\right) = \pi\left(\sum_{\alpha \in \lambda} r_\alpha b_\alpha\right) = \sum_{\alpha \in \lambda} r_\alpha \pi(b_\alpha) = \sum_{\alpha \in \lambda} r_\alpha c_\alpha,$$

so $\pi \circ \rho = \mathrm{id}_C$, so the short exact sequence splits due to this map. So indeed, $B \cong A \oplus C$. ∎

In particular, free modules make $\mathrm{Hom}(M, -)$ into an exact functor.

Proposition 3.21. Fix a split short exact sequence of R-modules

$$0 \to A \to B \to C \to 0.$$

Then we have the short exact sequence

$$0 \to \mathrm{Hom}(M, A) \to \mathrm{Hom}(M, B) \to \mathrm{Hom}(M, C) \to 0.$$

In particular, we get this whenever C is free.

Proof. Set $B \cong A \oplus C$ as induced by the short exact sequence and label our sequence by

$$0 \longrightarrow A \xrightarrow{\iota} A \oplus C \xrightarrow{\pi} C \longrightarrow 0$$

and

$$0 \longrightarrow \mathrm{Hom}(M, A) \xrightarrow{\iota \circ -} \mathrm{Hom}(M, A \oplus C) \xrightarrow{\pi \circ -} \mathrm{Hom}(M, C) \longrightarrow 0$$

At this point we only have to check that $\mathrm{Hom}(M, A \oplus C) \to \mathrm{Hom}(M, C)$ by $f \mapsto \pi \circ f$ is surjective. Indeed, fix any $g \in \mathrm{Hom}(M, C)$, and we can lift it to $m \mapsto gm \mapsto (0, gm)$, which is what we needed. ∎

We remark that, given a free module M generated by $\{m_\alpha\}_{\alpha \in \lambda}$, then we can describe $\mathrm{Hom}_R(M, M)$ essentially just using matrices: if $f \in \mathrm{Hom}_R(M, M)$, then we can describe

$$f(m_\alpha) = \sum_{\beta \in \lambda} a_{\alpha\beta} m_\beta,$$

which extends to just matrix multiplication. For example, if M is finitely generated by $\{m_k\}_{k=1}^n$, we can write

$$f(m_\ell) = \sum_{k=1}^{n} a_{k\ell} m_k$$

so that f corresponds to the matrix

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}.$$

Of course, we have to be careful about the direction here.

Warning 3.22. If M is a left module, then matrix multiplication should (in a moral sense) happen on the right, as discussed earlier.

Namely,

$$f\left(\sum_{\ell=1}^{n} r_\ell m_\ell\right) = \sum_{k=1}^{n} \sum_{\ell=1}^{n} r_\ell a_{k\ell} m_k.$$

corresponds to the multiplication

$$\begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} r_1 a_{11} + \cdots + r_n a_{1n} \\ \vdots \\ r_1 a_{n1} + \cdots + r_n a_{nn} \end{bmatrix}.$$

3.1.4 Ranks of Free Modules

We would like to define the rank of an R-module. The goal is for $\operatorname{rank} R^n = n$ at the end, but this needs to depend only on the module itself. This is harder than it looks.

Non-Example 3.23. Take $R = \{0\}$. Then the rank is not well-defined because $R^n \cong \{0\}$ always.

So there is something to do when trying to define the rank.

Example 3.24. If R is a field, then we can use dimension, which we are assured is well-defined.

Example 3.25. If R is a nonzero commutative ring, then given a module M, we can take a maximal ideal $\mathfrak{m}$ (which exists because $R \neq \{0\}$) and measure

$$\operatorname{rank} M \coloneqq \dim_{R/\mathfrak{m}} M/\mathfrak{m}M.$$

Proposition 3.26. The above rank is well-defined. Explicitly, if R is a nonzero commutative ring with $\mathfrak{m}$ any maximal ideal, then $n = \dim_{R/\mathfrak{m}} R^n/\mathfrak{m}R^n$.

Proof. Note that it suffices to show that $n = \dim_{R/\mathfrak{m}} R^n/\mathfrak{m}R^n$, which will tell us that the rank does not depend on $\mathfrak{m}$. The main point is to show that

$$R^n/\mathfrak{m}R^n \cong (R/\mathfrak{m})^n$$

as $R/\mathfrak{m}$-vector spaces, which will finish because $(R/\mathfrak{m})^n$ is n n-dimensional $R/\mathfrak{m}$-vector space.

Indeed, we note that we have an R-module homomorphism

$$\varphi : R^n \to (R/\mathfrak{m})^n$$

by taking $(r_1, \dots, r_n) \mapsto (r_1 + \mathfrak{m}, \dots, r_n + \mathfrak{m})$. (That φ is actually an R-module homomorphism comes down to checking that $R \twoheadrightarrow R/\mathfrak{m}$ is a ring map, which is true.) Further, φ is surjective because we can lift any $(r_1 + \mathfrak{m}, \dots, r_n + \mathfrak{m})$ back up to some $(r_1, \dots, r_n)$.

Lastly, we note that $(r_1, \dots, r_n) \in \ker \varphi$ if and only if $r_\bullet + \mathfrak{m} = 0 + \mathfrak{m}$ for each $r_\bullet$, which is equivalent to $r_\bullet \in \mathfrak{m}$ for each $r_\bullet$. So the kernel is exactly $\mathfrak{m}^n = \mathfrak{m}R^n$. So we have an R-module isomorphism

$$\overline{\varphi} : R^n/\mathfrak{m}R^n \to (R/\mathfrak{m})^n$$

induced by φ. To make this is an $R/\mathfrak{m}$-module isomorphism, we first note that $R^n/\mathfrak{m}R^n$ is indeed an $R/\mathfrak{m}$-module where the $R/\mathfrak{m}$-action is induced by R: the thing to check here is that the action is well-defined, for which it suffices to note $r_1 + \mathfrak{m} = r_2 + \mathfrak{m}$ implies

$$r_1 (v + \mathfrak{m}R^n) = r_1 v + \mathfrak{m}R^n = r_2 v + \underbrace{(r_1 - r_2)}_{\in \mathfrak{m}} v + \mathfrak{m}R^n = r_2 v + \mathfrak{m}R^n,$$

so the action is well-defined up to coset of $R/\mathfrak{m}$. So we still have that

$$\overline{\varphi} : R^n/\mathfrak{m}R^n \to (R/\mathfrak{m})^n$$

is an isomorphism of abelian groups, and the $R/\mathfrak{m}$-action is preserved because the $R/\mathfrak{m}$-action is induced by the R-action, which is preserved. This finishes. ∎

Example 3.27. Take $R \coloneqq k^{n\times n}$. Then R is a finite-dimensional vector space over k, so may define $\operatorname{rank} R^n \coloneqq \dim_k R^n / \dim_k R$. The point here is that we have managed to define rank for a special non-commutative ring.

So it looks like maybe the rank is always well-defined for nonzero rings? This turns out to be true for "small" rings in some sense, but of course not for general rings because algebra is terrible.

Proposition 3.28. We can construct a nonzero ring R so that $R \cong R \oplus R$.

Proof. Fix S a ring with a free, right module M so that $M \cong M \oplus M$ with $M \neq \{0\}$. For example, $M = \mathbb{Z}^{\oplus \mathbb{N}}$ as a $\mathbb{Z}$-module will do the trick.

Now we define $R \coloneqq \operatorname{End}_S(M)$. Checking that this is a ring is annoying, so we will not do it in detail. The point is that we can add endomorphisms, and multiplication is composition. (In particular, R is not commutative.) Closure under addition and composition are a matter of writing out what we need to check.

Visually, if M is free over S, then R is the ring of matrices with infinite rows and columns with only finitely many nonzero elements. Here, R acts on the left of M in order to preserve the S-action on the right, and we can check that M is a left R-module.

- $\varphi(m_1 + m_2) = \varphi m_1 + \varphi m_2$ because these are endomorphisms.
- $(\varphi_1 + \varphi_2)m = \varphi_1 m + \varphi_2 m$ by definition of addition in R.
- $1_R m = \operatorname{id}_M m = m$.

The idea is to study how R behaves with the isomorphism $\varphi : M \to M \oplus M$. Because this is an isomorphism, it must be a homomorphism in each of the coordinates, so $a \coloneqq \pi_1 \circ \varphi$ and $c \coloneqq \pi_2 \circ \varphi$ (where $\pi_\bullet$ are the projections) must be homomorphisms. So we have $a, c \in \operatorname{End}_S(M) =: R$ with

$$\varphi : m \mapsto (am, cm).$$

Conversely, because φ is an isomorphism, we have a map $M \oplus M \to M$ which again must be a homomorphism in both coordinate by using the inclusions defining $M \oplus M$. So by the universal property of $\oplus$, we have $b, d \in \operatorname{End}_S(M) =: R$ with

$$\varphi^{-1}(m, n) \mapsto bm + dn.$$

Composing as $\varphi \circ \varphi^{-1} = \operatorname{id}_{M\oplus M}$ and $\varphi^{-1} \circ \varphi = \operatorname{id}_M$, we find

$$(m, n) = \left(\varphi \circ \varphi^{-1}\right)(m, n) = \varphi(bm + dn) = (abm + adn, cbm + cdn),$$

and

$$m = \left(\varphi^{-1} \circ \varphi\right)(m) = \varphi^{-1}(am, cm) = bam + dcm.$$

Comparing componentwise, we see that

$$ab = 1, \qquad ad = 0, \qquad cb = 0, \qquad cd = 1, \qquad ba + dc = 1.$$

We remark that if R were commutative, $ab = 1$ and $cd = 1$ and $ba + dc = 1$ would imply that $1 = 2$ and $0 = 1$, forcing R to be the zero ring.

Anyways, the point is that we have $R \cong R \oplus R$ by

$$\gamma : r \mapsto (ar, cr) \qquad \text{and} \gamma^{-1} : (r, s) \mapsto br + ds.$$

Essentially directly from the above computations we can check that

$$\left(\gamma \circ \gamma^{-1}\right)(r, s) = \gamma(br + ds) = (abr + ads, cbr + cds) = (r, s),$$

and

$$\left(\gamma^{-1} \circ \gamma\right)(r) = \gamma^{-1}(ar, cr) = bar + dcr = (ba + dc)r = r.$$

So we have a group isomorphism $R \cong R \oplus R$. To make this an R-module homomorphism, we have R act on itself on the right by multiplication, which is safe because γ and γ^{-1} only ever multiple on the left. For example, γ is an R-module homomorphism because, for $x, r \in R$, we have

$$\gamma(x) \cdot r = (ax, cx) \cdot r = (axr, cxr) = \gamma(xr),$$

and similar works for γ^{-1}. Thus, $R \cong R \oplus R$ as (right) R-modules. If we wanted left R-modules, we could switch the directions of everything above. ■

Remark 3.29. It is also true that $R \cong R^{2\times 2}$, but we've seen enough weird properties of this ring for today.

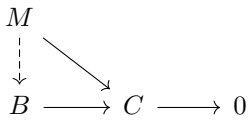
So our rank is not always well-defined for the above R. This is why people (or at least I) don't like modules over non-commutative rings.

3.1.5 Projective Modules

Recall from Proposition 3.21 that free R-modules C had the nice property of preserving the exactness of

$$B \to C \to 0$$

upon applying $\mathrm{Hom}(M, -)$. In other words, for any map $M \to C$, we can lift it to a map $M \to B$ so that $\mathrm{Hom}(M, B) \to \mathrm{Hom}(M, C)$ is surjective. Here is the diagram.



This property of "lifting surjections" is so nice that it has a name.

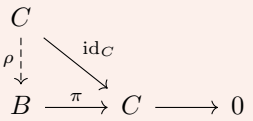
Definition 3.30 (Projective)**.** A module M is *projective* if it has the above property.

Example 3.31. Any free module is projective. Roughly speaking, this is by Proposition 3.21.

Remark 3.32 (Nir)**.** Here is a quick way to be convinced that "projective" is a good idea to care about: projective is what makes Proposition 3.20 work. Indeed, if

$$0 \longrightarrow A \xrightarrow{\iota} B \xrightarrow{\pi} C \longrightarrow 0$$

is a short exact sequence with C projective, then the surjection $\pi : B \twoheadrightarrow C$ induces $\rho : C \to B$ so that $\pi \circ \rho = \mathrm{id}_C$ by projectivity, as in the following diagram.



So indeed, our original short exact sequence splits.

Here is one way for us to generate lots of projective modules.

Proposition 3.33. Fix M a projective R-module. Then if $M \cong P \oplus Q$ (as R-modules), then both P and Q are projective. We may call P and Q the "split/direct summands."

Proof. We show that P is projective, and Q projective will follow by symmetry (because $M = Q \oplus P$). Fix any surjection $\varphi : B \twoheadrightarrow C$ with a map $f : P \to C$ so that we lift f to $\overline{f}$ making the following diagram commute.

$$\begin{array}{ccccc} & & & & P \\ & & \overset{\overline{f}?}{\swarrow} & & \downarrow f \\ B & \xrightarrow[\varphi]{} & C & \longrightarrow & 0 \end{array}$$

Now, $f : P \to C$ induces a composite map g by $M \twoheadrightarrow P \to C$ using the canonical projection $M \twoheadrightarrow P$ by $(p, q) \mapsto p$, so because we have a map $g : M \to C$, this lifts to a map $\overline{g}$ making the following diagram commute.

$$\begin{array}{ccccc} M & \twoheadrightarrow & P & & \\ \overline{g}\downarrow & & \downarrow f & & \\ B & \xrightarrow[\varphi]{} & C & \longrightarrow & 0 \end{array}$$

However, we also have a canonical inclusion $P \hookrightarrow M$ by $p \mapsto (p, 0)$, so we have induced a map $\overline{f}$ by $P \hookrightarrow M \to B$. We claim that this is the map we want. Indeed, we know that $(\varphi \circ \overline{g})(p, q) = f(p)$ by construction of $\overline{g}$, so

$$(\varphi \circ \overline{f})(p) = (\varphi \circ \overline{g})(p, 0) = f(p),$$

which is exactly what we need. ∎

Note that we are not claiming that general submodules P of free modules M are projective: we need the short exact sequence

$$0 \to P \to M \to M/P \to 0$$

split.

Here is another nice property of projective modules: this is the converse of Proposition 3.33.

Proposition 3.34. Fix M a projective R-module. Then $M \oplus N$ is free for some R-module N. If M is finitely generated, we may let $M \oplus N$ be finitely generated.

Proof. Fix F any free module which can surject onto M, and let $\pi : F \twoheadrightarrow M$ be our surjection. (For example, the free module $\bigoplus_{m \in M} Rm$ generated by the letters of M would do the trick. If M is finitely generated, use the corresponding F.) Then the idea is to lift $\mathrm{id}_M : M \to M$ along the surjection $\pi : F \to M$ to some $\rho : M \to F$. Here is the diagram.

$$\begin{array}{ccccc} M & & & & \\ \rho\vdots & \searrow^{\mathrm{id}_M} & & & \\ F & \xrightarrow[\pi]{} & M & \longrightarrow & 0 \end{array}$$

The point here is that the short exact sequence

$$0 \to \ker \pi \hookrightarrow F \twoheadrightarrow M \to 0$$

will split due to ρ: by construction of ρ, we have that $\pi \circ \rho = \mathrm{id}_M$, which is exactly the condition to make this short exact sequence split. Thus, $M \oplus \ker \pi \cong F$ is free, which is what we wanted. ∎

Remark 3.35 (Nir). Collecting our facts about projective modules, we have the following criteria for an R-module C, which we claim are equivalent.

(i) For each surjection $\pi : B \twoheadrightarrow C$ and map $\varphi : M \to C$, there exists a map $\overline{\varphi} : M \to B$ so that $\varphi = \pi \circ \overline{\varphi}$. (I.e., the induced map $\operatorname{Hom}(M, B) \xrightarrow{\pi \circ -} \operatorname{Hom}(M, C)$ is surjective.)

(ii) Every short exact sequence $0 \to A \to B \to C \to 0$ splits.

(iii) There exists a module N such that $C \oplus N$ is free.

From the above discussion, we already know (i) implies (ii) as well as (iii) implies (i). We can also see that (ii) implies (iii) by considering the short exact sequence

$$0 \to \ker \pi \hookrightarrow \bigoplus_{c \in C} Rc \xrightarrow{\pi} C \to 0,$$

which must split and gives $C \oplus \ker \pi$ free.

3.1.6 Examples of Projective Modules

We've been providing some theory on projective modules, but most of what we've done would only produce free modules as our examples. So it looks like projective might mean free, but here is an example saying no.

Example 3.36. Note that R is a free R-module, so if we can decompose $R = A \oplus B$ into R-modules, then A and B will be projective. For example, fix $R = \mathbb{Z}/6\mathbb{Z} = \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}$ implies that $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$ must then be projective $\mathbb{Z}/6\mathbb{Z}$-modules. (We technically have to check that $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$ are $\mathbb{Z}/6\mathbb{Z}$-modules, and they are, induced by the $\mathbb{Z}$-action.)

However, $\mathbb{Z}/2\mathbb{Z}$ is not free over $\mathbb{Z}/6\mathbb{Z}$ because it would have to have dimension strictly between 0 and 1, which is impossible.

One potential complaint is that the above example more or less has zero-divisors built into R: if we can decompose $R \cong A \oplus B$ into two nonzero R-modules, then[2]

$$(a, 0) \cdot (0, b) = (0, 0)$$

for $a \in A \setminus \{0\}$ and $b \in B \setminus \{0\}$ forces R to have zero-divisors.

So here is an example where R is an integral domain.

Exercise 3.37. Fix $R := \mathbb{Z}[\sqrt{-5}]$ and $\mathfrak{p} := \left(2, 1 + \sqrt{-5}\right)$ a non-principal R-ideal. Then $\mathfrak{p}$ is a projective but not free module.

Proof. We start by showing that $\mathfrak{p}$ is not free. Roughly speaking, this comes from the fact $\mathfrak{p}$ is not a principal ideal. Quickly, we see that $\mathfrak{p}$ is not freely generated by zero elements because $\mathfrak{p} \neq \{0\}$, and $\mathfrak{p}$ is not freely generated by one element because $\mathfrak{p}$ is not principal.

Now, supposing that $\mathfrak{p}$ is generated by some set $\{z_\alpha\}_{\alpha \in \lambda} \subseteq \mathfrak{p}$ with $\#\lambda \geq 2$, and we show the $z_\bullet$ do not freely generate. Indeed, the trick is that

$$\left(\mathrm{N}(z_\beta)\overline{z_\alpha}\right) z_\alpha + \left(-\mathrm{N}(z_\alpha)\overline{z_\beta}\right) z_\beta = 0 \tag{$*$}$$

for any $\alpha, \beta \in \lambda$ distinct elements. If $z_\alpha = 0$ or $z_\beta = 0$, then of course the $z_\bullet$ do not freely generate. Otherwise, $(*)$ tells us that the map

$$\bigoplus_{\alpha \in \lambda} Rz_\alpha \twoheadrightarrow \mathfrak{p}$$

[2] Technically we are forcing some multiplication structure here.

has kernel, making the $z_\bullet$ still not freely generate.

We now check that $\mathfrak{p}$ is projective. To start, we note that we have a surjection $\pi : R \oplus R \twoheadrightarrow \mathfrak{p}$ by

$$\pi : (r, s) \mapsto 2r + \left(1 + \sqrt{-5}\right) s.$$

But in fact we can split π with $\rho : \mathfrak{p} \to R \oplus R$ by

$$\rho : x \mapsto \left(-x, \frac{1 - \sqrt{-5}}{2} x\right).$$

This is well-defined because $\rho(2) = \left(-2, 1 - \sqrt{-5}\right) \in R \oplus R$, and $\rho\left(1 + \sqrt{-5}\right) = \left(-1 - \sqrt{-5}, 3\right) \in R \oplus R$. Further, we can compute

$$(\pi \circ \rho)(x) = \pi\left(-x, \frac{1 - \sqrt{-5}}{2} x\right) = -2x + \left(1 + \sqrt{-5}\right)\left(\frac{1 - \sqrt{-5}}{2}\right) x = -2x + 3x = x,$$

so indeed, $\pi \circ \rho = \mathrm{id}_{\mathfrak{p}}$. The point is that the short exact sequence

$$0 \to \ker \pi \hookrightarrow R \oplus R \twoheadrightarrow \mathfrak{p} \to 0$$

splits, so $\mathfrak{p} \oplus \ker \pi \cong R \oplus R$. It follows that $\mathfrak{p}$ is projective by Proposition 3.33. ∎

We continue with the examples.

Exercise 3.38. Consider the Möbius strip X as a line bundle over S^1; let π be our standard projection $X \twoheadrightarrow S^1$. I will not TeX a diagram of this, out of laziness. To get our module, we define

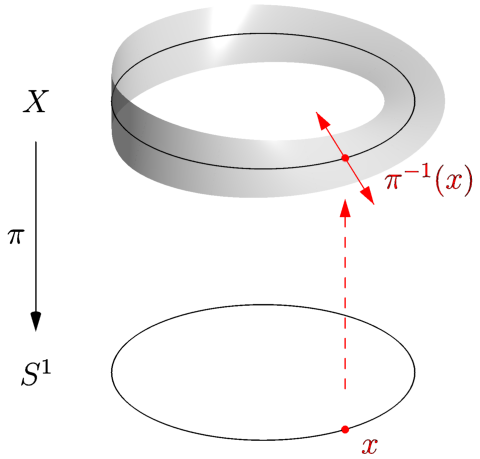
$$R \coloneqq \{\text{continuous functions } r : S^1 \to \mathbb{R}\}$$

and

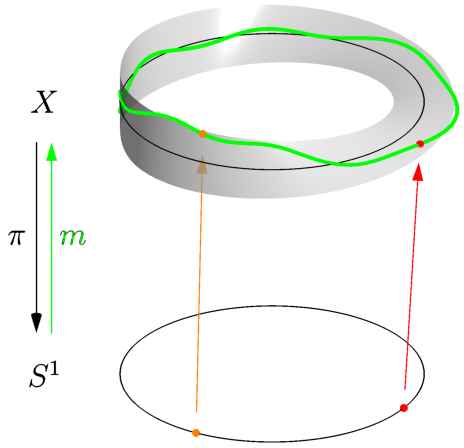
$$M \coloneqq \{\text{continuous functions } m : S^1 \to X \text{ such that } \pi \circ \rho = \mathrm{id}_{S^1}\}.$$

Then M is a projective but not free R-module.

Proof. Here is the image of the Möbius strip X projecting on S^1 by $\pi : X \to S^1$. We have highlighted the fiber of a particular point $x \in S^1$ and explicitly note that it is a (one-dimensional) vector space.



Now, an element of M is a "global section" of X, which means it is a continuous function $m : S^1 \to X$ such that $S^1 \xrightarrow{m} X \xrightarrow{\pi} S^1$ is the identity. For example, here is such a section.

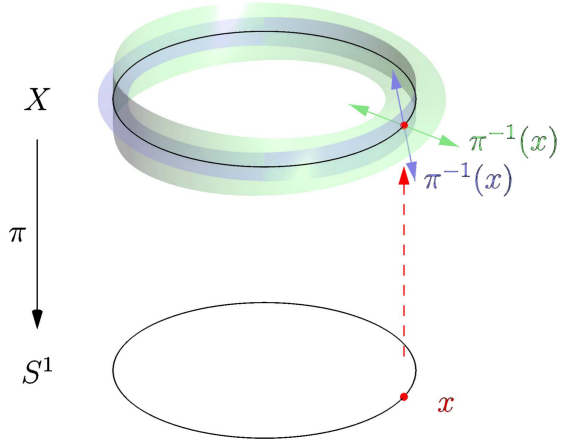


Now, for any $r \in R$ and $m \in M$, we define rm to act pointwise. For example, if $r \equiv 2$, then rm essentially stretches r by 2.

To start, we claim that

$$M \oplus M \cong R \oplus R,$$

which will give the needed projectivity. Essentially, $M \oplus M$ allows us two copies of the Möbius strip, which we can lay orthogonally as in the following diagram.



But having two axes at each point is merely assigning an $\mathbb{R}^2$ to each point of S^1, so this is the vector bundle $S^1 \times \mathbb{R}^2$. Now, a section $S^1 \to S^1 \times \mathbb{R}^2$ is pretty much just a pair of functions $S^1 \to \mathbb{R}$, which precisely describes $R \oplus R$. So indeed, $M \oplus M \cong R \oplus R$.

We now check that M is not free. The main point is that, due to the twisting, going around the edge of the Möbius returns us to the opposite side, so any continuous section $m : S^1 \to X$ must intersect the central S^1 somewhere.[3] In terms of the line bundle, we are saying that all global sections vanish somewhere.

Now suppose for the sake of contradiction M is free. Because we already know that

$$M \oplus M \cong R \oplus R,$$

and because R is a commutative ring, we see that ranks are well-defined, so M free implies that M is generated by a single element, say m. But m vanishes at some x_0, so anything in Rm will also vanish at x_0, so $Rm \neq M$, which is a contradiction. ■

[3] Thinking about m as the composite $[0, 2\pi) \to S^1 \to X$, we are saying $m(0) = -m(2\pi)$ because walking around the loop flips the sign.

Remark 3.39. Nobody seems to be able to understand the above example in lecture. I am no exception.

In general, the above examples have been taking the theme of finding a module M such that $M \oplus M \cong R \oplus R$. What this means is that our projective modules look "locally" like a free module but not globally, in the sense that we can add enough copies to get the free module.

3.1.7 Stably Free Modules

Recall from Proposition 3.34 that, if P is a projective R-module, then $P \oplus Q$ is free for some Q. We might hope to be able to constrain Q in some way; of course it must be projective by Proposition 3.33, but perhaps we can do more.

Proposition 3.40. Fix P a projective R-module. Then $P \oplus Q$ is free for some free R-module Q.

Proof. We start by using Proposition 3.34 to get some N such that $P \oplus N$ is free. Then the trick is to study

$$Q \coloneqq \bigoplus_{k \in \mathbb{Z}} (N \oplus P).$$

On one hand, $N \oplus P \cong P \oplus N$ is free, so Q is free because it is the direct sum of free modules. On the other hand, we can write

$$P \oplus Q = P \oplus \bigoplus_{k \in \mathbb{Z}} (N \oplus P) = \bigoplus_{k \in \mathbb{Z}} (P \oplus N).$$

In other words,

$$P \oplus Q = P \oplus (N \oplus P) \oplus (N \oplus P) \oplus \cdots = (P \oplus N) \oplus (P \oplus N) \oplus \cdots.$$

Anyways, the point is that $P \oplus Q$ is free because it is the direct sum of the free modules $P \oplus N$. So Q is a free module, and $P \oplus Q$ is a free module. ∎

Remark 3.41. This is similar to the "proof" that $0 = 1$ by

$$0 = (1 + -1) + (1 + -1) + \cdots = 1 + (-1 + 1) + (-1 + 1) + \cdots = 1.$$

We might hope to make Q of finite rank, but this is not always possible. Such modules have a name.

Definition 3.42 (Stably free)**.** Fix an R-module M. Then we say that M is *stably free* if and only if there exists $n \in \mathbb{N}$ such that $M \oplus R^n$ is free and finitely generated.

As promised, not all projective modules are stably free.

Proposition 3.43. There exist projective modules which are not stably free.

Proof. As above, take M the global sections Möbius strip form a R-module as defined above. Intuitively, M is not stably free because adding any finite number of R to the Möbius band cannot "untwist" the Möbius band, implying that it will never be free.

Explicitly, if $M \oplus R^n$ were free for some $n \geq 0$, then, comparing ranks of

$$(M \oplus R^n) \oplus (M \oplus R^n) \cong (M \oplus M) \oplus R^{2n} \cong R^{2n+2},$$

we still would need $M \oplus R^n$ generated by $n + 1$ element. But we need n generators for R^n, so we only have one degree of freedom for M, which fails for the same reasons as before. (Please don't ask me to rigorize this.) ∎

But nontrivial stably free modules do exist.

Proposition 3.44. There exist stably free modules which are not free modules.

Proof. Take the tangent bundle of S^2, where the total space consists of the tangent space of each point on S^2. Now our ring R consists of continuous functions $S^2 \to \mathbb{R}$ and our module M is vector fields on S^2, where the action is again just scalar multiplication.

We have the following checks.

- We see that M is stably free (and hence projective) because
$$M \oplus R \cong R^3.$$
Indeed, what is happening is that the extra $\oplus R$ will encode the orthogonal bundle to the tangent bundle, so now are essentially associating a full $\mathbb{R}^3$ to each point on S^2, which is exactly R^3.
- We check that M is not free. This follows from the Hairy ball theorem, which tells us that every vector field in M must vanish somewhere.

  Indeed, suppose for the sake of contradiction M was free. Then, comparing ranks of $M \oplus R \cong R^3$, we see that M must have rank 2, with generators we name m_1, m_2, which vanish at (say) $x_1, x_2 \in S^2$. But then, any R-linear combination
$$r_1 m_1 + r_2 m_2$$
will have $(r_1 m_1 + r_2 m_2)(x_1) = r_2 m_2(x_1)$ parallel to $m_2(x_1)$ at x_1. But there are vector fields which are perpendicular to $m_2(x_1)$ at x_1, so we have not covered all of M, which is our contradiction. ∎

We remark that if we use a torus $S^1 \times S^1$ instead of the sphere, then the module is free because we no longer have access to the Hairy ball theorem.

3.2 October 7

There must be some way out of here.

3.2.1 Tensor Products over Abelian Groups

Today we're going to define tensor products. For now, we work in the category of abelian groups.

To begin with we have the following warning.

Warning 3.45. The tensor product is not $A \times B$, and in fact has little to do with this product. Instead, the tensor product is arguably closer to $A \oplus B$.

Anyways, the main idea behind tensor products is the following.

Idea 3.46. Tensor products turn bilinear maps $A \times B \to C$ into linear maps from $A \otimes B \to C$.

In particular, we should start by defining "bilinear."

Definition 3.47 (Bilinear)**.** Fix abelian groups A, B, C. A map $\varphi : A \times B \to C$ is *bilinear* if and only if
$$f(a_1 + a_2, b) = f(a_1, b) + f(a_2, b) \qquad \text{and} \qquad f(a, b_1 + b_2) = f(a, b_1) + f(a, b_2)$$
for each $a_1, a_2, a \in A$ and $b_1, b_2, b \in B$.

Equivalently, we see that

$$f(a_1 + a_2, b) = f(a_1, b) + f(a_2, b)$$

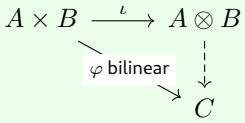
for any $a_1, a_2 \in A$ and $b \in B$ is merely asserting that $b \mapsto f(-, b)$ is a function $B \to \operatorname{Hom}(A, C)$. Similarly,

$$f(a, b_1 + b_2) = f(a, b_1) + f(a, b_2)$$

for $a \in A$ and $b_1, b_2 \in B$ is asserting that $a \mapsto f(a, -)$ is a function $A \to \operatorname{Hom}(B, C)$.

We defined "bilinear" so that we could define tensor products.

Definition 3.48 (Tensor products)**.** The *tensor product* $A \otimes B$ of two abelian groups A and B to be "universal" as an abelian group equipped with a bilinear map $\iota : A \times B \to A \otimes B$. Explicitly, for any bilinear map $\varphi : A \times B \to C$, there exists a unique induced homomorphism (!) $A \otimes B \to C$ making the following diagram commute.



We remark that tensor products are unique up to isomorphism, using a fairly typical argument.

As usual, we start by showing that tensor products actually exist.

Proposition 3.49. Given two abelian groups A and B, their tensor product $A \otimes B$ exists.

Proof. Essentially, we want the "largest" abelian group with a bilinear map from $A \times B$. To start off, we'll say that we send $\iota : (a, b) \mapsto a \otimes b \in A \otimes B$, and then we mod out by the minimal relations which will make ι bilinear.

Explicitly, $A \times B$ has a subgroup generated by

$$G \coloneqq \big\langle (a_1, b) + (a_2, b) - (a_1 + a_2, b) \qquad \text{and} \qquad (a, b_1) + (a, b_2) - (a, b_1 + b_2) \big\rangle.$$

(For example, take the image from the free abelian group on $(A \times B)^6$.) Now we define

$$A \otimes B \coloneqq \frac{A \times B}{G}.$$

It remains to show the universal property. Well, suppose we have a bilinear map $\varphi : A \times B \to C$. Then by hypothesis on φ, we know that

$$\varphi\big((a_1, b) + (a_2, b) - (a_1 + a_2, b)\big) = 0 \qquad \text{and} \qquad \varphi\big((b_1, a) + (b_2, a) - (b_1 + b_2, a)\big) = 0.$$

As these elements generate G, we see $G \subseteq \ker \varphi$, so we have a unique induced homomorphism from $A \otimes B = (A \times B)/G$ to C, as requested. ∎

The above proof does establish existence, but, as seems to be the case a lot, we have just taken a huge thing modulo a huge thing, so it is not even obvious if the tensor product is nonzero zero. And (unlike with localization) there is actually danger here! For example,

$$\mathbb{Z}/91\mathbb{Z} \otimes \mathbb{Z}/119\mathbb{Z} \neq 0 \qquad \text{but} \qquad \mathbb{Z}/91\mathbb{Z} \otimes \mathbb{Z}/120\mathbb{Z} = 0.$$

So in practice, to actually compute the tensor product, we use the universal property and notably not the explicit construction. Here are some examples of doing this by hand.

Proposition 3.50. Fix A any abelian group. Then $\mathbb{Z} \otimes A \cong A$.

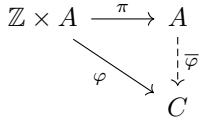
By symmetry, we remark that $A \otimes \mathbb{Z} \cong A$ as well.

Proof. The main point is that we can take elements of the form $k \otimes a$ and turn them into $1 \otimes a^k$, which projects nicely into A. Formally, we show that A satisfies the universal property of $\mathbb{Z} \otimes A$. We define the needed inclusion $\iota : \mathbb{Z} \times A \to A$ by

$$(k, a) \mapsto a^k$$

for any $k \in \mathbb{Z}$ and $a \in A$.

Now, for any bilinear map $\varphi : \mathbb{Z} \times A \to C$, we have to show that there exists a unique $\overline{\varphi}$ making the following diagram commute.



To start, fix any bilinear map $\varphi : \mathbb{Z} \times A \to C$ for some abelian group C. Then, for $a_1, a_2 \in A$, we see

$$\varphi(1, a_1 + a_2) = \varphi(1, a_1) + \varphi(1, a_2)$$

because φ is bilinear, so $\varphi(1, -) \in \operatorname{Hom}(A, C)$. And we can check that the needed diagram commutes because, for any $(k, a) \in \mathbb{Z} \times A$,

$$\varphi(k, a) = \varphi(\underbrace{1 + \cdots + 1}_{k}, a) = \underbrace{\varphi(1, a) + \cdots + \varphi(1, a)}_{k} = k\varphi(1, a) = k\overline{\varphi}(a) = \overline{\varphi}\left(a^k\right),$$

which is $(\overline{\varphi} \circ \iota)((k, a))$, as needed.

We now show uniqueness. Suppose that some $\overline{\varphi} : A \to C$ makes the given diagram commute. Then we find that

$$\overline{\varphi}(a) = (\overline{\varphi} \circ \iota)(1, a) = \varphi(1, a)$$

uniquely determines $\overline{\varphi}$. ∎

For completeness, we observe that the induced isomorphism $A \cong \mathbb{Z} \otimes A$ is by $a \otimes k \mapsto a^k$, which we see by applying the universal property to the canonical bilinear map $\mathbb{Z} \times A \to \mathbb{Z} \otimes A$. The inverse map is $a \mapsto a \otimes 1$.

Remark 3.51 (Nir)**.** In fact, we remark that a homomorphism of abelian groups $\varphi : A \to B$ will remain unchanged after taking $\mathbb{Z} \otimes -$ and applying the above isomorphism. Indeed, the induced morphism $\varphi : \mathbb{Z} \otimes A \to \mathbb{Z} \otimes B$ is by

$$\varphi_1(k \otimes a) = k \otimes \varphi(a),$$

which can be checked to be homomorphic. But applying the isomorphism $x \otimes k \mapsto x^k$, we get $\varphi_2 : A \to B$ which satisfies $\varphi_2\left(a^k\right) = \varphi(a)^k$ for each a, k, which is true and the exact same φ morphism we had before.

Proposition 3.52. Fix abelian groups A, B, C. Then we have the "distributive" law

$$(A \oplus B) \otimes C \cong (A \otimes C) \oplus (B \otimes C).$$

Proof. The point is that bilinear maps from $(A \oplus B) \times C \to X$ are the same as a pair of bilinear maps $A \times C \to X$ and $B \times C \to X$. Formally we show that $(A \otimes C) \oplus (B \otimes C)$ satisfies the universal property of $(A \oplus B) \otimes C$.

To start off, we note that we have the map $\iota : (A \oplus B) \times C \to (A \otimes C) \oplus (B \otimes C)$ by

$$\iota : ((a, b), c) \mapsto (a \otimes c, b \otimes c).$$

This map is bilinear, roughly by construction. For example,

$$\iota((a_1, b_1) + (a_2, b_2), c) = (a_1 \otimes c + a_2 \otimes c, b_1 \otimes c + b_2 \otimes c) = (a_1 \otimes c, b_1 \otimes c) + (a_2 \otimes c, b_2 \otimes c),$$

and the other side is similar.

It remains to show the universal property. Suppose that we have any bilinear map $\varphi : (A \oplus B) \times C \to X$ so that we want to exhibit a unique linear map $\overline{\varphi} : (A \otimes C) \oplus (B \otimes C) \to X$ making the following diagram commute.

$$\begin{array}{ccc} (A \oplus B) \times C & \xrightarrow{\iota} & (A \otimes C) \oplus (B \otimes C) \\ & \searrow^{\varphi} & \downarrow{\scriptstyle \overline{\varphi}} \\ & & X \end{array}$$

We start by showing the uniqueness of $\overline{\varphi}$. Indeed, for any $a \in A, b \in B, c \in C$, we can push $((a,b),c) \in (A \oplus B) \times C$ through the diagram to see the following.

$$\begin{array}{ccc} \big((a,b),c\big) & \xmapsto{\iota} & (a \otimes c, b \otimes c) \\ & \searrow^{\varphi} & \downarrow{\scriptstyle \overline{\varphi}} \\ & & \varphi\big((a,b),c\big) \end{array}$$

Namely, we must have

$$\overline{\varphi}(a \otimes c, b \otimes c) = \varphi((a,b),c)$$

for each $a \in A, b \in B, c \in C$. It follows that

$$\overline{\varphi}(a \otimes c_1, b \otimes c_2) = \overline{\varphi}(a \otimes c_1, b \otimes c_1) + \overline{\varphi}(a \otimes 0, b \otimes (c_2 - c_1)) = \varphi((a,b),c_1) + \varphi((0,b),c_2 - c_1),$$

so indeed, $\overline{\varphi}$ is uniquely determined. More simply this is $\varphi((a,0),c_1) + \varphi((0,b),c_2)$ after some rearranging.

It remains to show that $\overline{\varphi}$ is actually well-defined. Well, by projecting on the a coordinate, we see that φ induces a bilinear map $\varphi_A : A \times C \to X$ by

$$\varphi_A(a,c) = \varphi((a,0),c).$$

Similarly, we get a bilinear map $\varphi_B : B \times C \to X$ by $\varphi_B(b,c) = \varphi((0,b),c)$. We will not check that these are bilinear explicitly.

The point is that our bilinear maps φ_A and φ_B induces linear maps $\overline{\varphi_A} : A \otimes C \to X$ (by $a \otimes c \mapsto \varphi((a,0),c)$) and $\overline{\varphi_B} : B \otimes C \to X$ (by $b \otimes c \mapsto \varphi((0,b),c)$), so we have the following diagram.

$$\begin{array}{ccccc} & & A \otimes C & & \\ & & \downarrow{\scriptstyle \iota_A} & \searrow^{\overline{\varphi_A}} & \\ B \otimes C & \xrightarrow{\iota_B} & (A \otimes C) \oplus (B \otimes C) & \overset{\overline{\varphi}}{\dashrightarrow} & X \\ & \searrow_{\overline{\varphi_B}} & & \nearrow & \end{array}$$

Namely, we have an induced $\overline{\varphi}$ defined by

$$\overline{\varphi}(a \otimes c_1, b \otimes c_2) = \overline{\varphi_A}(a \otimes c_1) + \overline{\varphi_B}(b \otimes c_2) = \varphi_A(a,c_1) + \varphi_B(b,c_2),$$

which is indeed $\varphi_A((a,0),c_1) + \varphi_B((0,b),c_2)$. So this map does exist. ∎

Remark 3.53 (Nir)**.** The second part of the proof can be stated in terms of bijections between Hom sets and show the uniqueness and existence simultaneously. However, the above proof feels more concrete to me.

Example 3.54. We have that

$$\mathbb{Z}^m \otimes \mathbb{Z}^n = (\underbrace{\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}}_{m}) \otimes \mathbb{Z}^n \cong \underbrace{(\mathbb{Z} \otimes \mathbb{Z}^n) \oplus \cdots \oplus (\mathbb{Z} \otimes \mathbb{Z}^n)}_{m} \cong (\mathbb{Z}^n)^m \cong \mathbb{Z}^{mn}.$$

3.2.2 Tensor Is Right Exact

In general, we might want to compute tensor products with quotients. This would involve taking the short exact sequence

$$0 \to A \to B \to C \to 0$$

to a sequence

$$0 \to A \otimes M \to B \otimes M \to C \otimes M \to 0.$$

The best possible world would make this sequence short exact. Well, at least part of the sequence is exact.

Theorem 3.55. If

$$0 \to A \to B \to C \to 0$$

is a short exact sequence, then

$$A \otimes M \to B \otimes M \to C \otimes M \to 0$$

is exact.

Proof. This is difficult to do with the specific construction we provided for the tensor product. So we use category theory, which makes this result trivial but not obvious. The main point is the following lemma.

Lemma 3.56. Fix B an abelian group. Then the tensor functor $- \otimes B$ is left adjoint to the hom functor $\mathrm{Hom}(B, -)$.

Proof. We note that $- \otimes B$ is actually a functor because a map $f : A_1 \to A_2$ will induce a map $\varphi : A_1 \otimes B \to A_2 \otimes B$ by

$$a_1 \otimes b \mapsto f(a_1) \otimes b.$$

Less explicitly, we have a bilinear map defined as the composite $A_1 \times B \xrightarrow{f} A_2 \times B \to A_2 \otimes B$, which will induce a map $A_1 \otimes B \to A_2 \otimes B$, defined as above.

Anyways, the main idea for the adjunction is that, for any abelian groups A, C,

$$\mathrm{Hom}(A \otimes B, C) \cong \mathrm{Bilinear}(A \times B, C) = \mathrm{Hom}(A, \mathrm{Hom}(B, C)),$$

where the last step is by currying, and these isomorphisms are exactly we need for the lemma. We will establish these isomorphisms, but we will not actually show the coherence laws for the adjunction because I'm lazy.

Namely, linear maps $A \otimes B \to C$ are in canonical bijection with bilinear maps $A \times B \to C$ by definition of $\otimes$. In fact this is a group isomorphism, where the operation on $\mathrm{Bilinear}(A \times B, C)$ is pointwise addition. Indeed, we are homomorphic because $\varphi_1, \varphi_2 \in \mathrm{Bilinear}(A \times B, C)$ have

$$(\overline{\varphi_1} + \overline{\varphi_2})(a \otimes b) = \overline{\varphi_1}(a \otimes b) + \overline{\varphi_2}(a \otimes b) = (\varphi_1 + \varphi_2)(a, b).$$

This establishes the isomorphism $\mathrm{Hom}(A \otimes B, C) \cong \mathrm{Bilinear}(A \times B, C)$.

Now, currying says that bilinear maps $\varphi : A \times B \to C$ are really curried homomorphisms: given $a \in A$, define $\varphi_a \in \mathrm{Hom}(B, C)$ by $\varphi_a(b) \coloneqq \varphi(a, b)$. Then φ_a is indeed in $\mathrm{Hom}(B, C)$ because

$$\varphi_a(b_1 + b_2) = \varphi(a, b_1 + b_2) = \varphi(a, b_1) + \varphi(a, b_2) = \varphi_a(b_1) + \varphi_a(b_2).$$

But further, the map $a \mapsto \varphi_a$ is itself a group homomorphism in $\mathrm{Hom}(A, \mathrm{Hom}(B, C))$; indeed, we have

$$\varphi_{a_1 + a_2}(b) = \varphi(a_1 + a_2, b) = \varphi(a_1, b) + \varphi(a_2, b) + \varphi_{a_1}(b) + \varphi_{a_2}(b).$$

So we have a map of sets $\mathrm{Bilinear}(A \times B, C) \to \mathrm{Hom}(A, \mathrm{Hom}(B, C))$. In fact, this is homomorphic because the sum $\varphi_1 + \varphi_2$ will have

$$\big((\varphi_1)_a + (\varphi_2)_a\big)(b) + (\varphi_2)_a(b) = \varphi_1(a,b) + \varphi_2(a,b) = (\varphi_1 + \varphi_2)(a,b).$$

And our map is injective because $\varphi \in \mathrm{Bilinear}(A \times B, C)$ going to the zero map in $[Hom](A, \mathrm{Hom}(B, C))$ would mean that $\varphi(a,b) = \varphi_a(b) = 0(b) = 0$ for each $(a,b) \in A \times B$. So the set map has trivial kernel.

Lastly, our maps is surjective because $\varphi_\bullet \in \mathrm{Hom}(A, \mathrm{Hom}(B, C))$ can be induced by a $\varphi \in \mathrm{Bilinear}(A \times B, C)$ by $\varphi(a,b) \coloneqq \varphi_a(b)$. We see that φ is indeed bilinear because

$$\varphi(a_1 + a_2, b) = \varphi_{a_1+a_2}(b) = \varphi_{a_1}(b) + \varphi_{a_2}(b) = \varphi(a_1, b) + \varphi(a_2, b)$$

and

$$\varphi(a, b_1 + b_2) = \varphi_a(b_1 + b_2) = \varphi_a(b_1) + \varphi_a(b_2) = \varphi(a, b_1) + \varphi(a, b_2).$$

This finishes the isomorphism $\mathrm{Bilinear}(A \times B, C) \cong \mathrm{Hom}(A, \mathrm{Hom}(B, C))$. ■

Now that we know $- \otimes B$ is a left adjoint, we pick up the following fact about left adjoints.

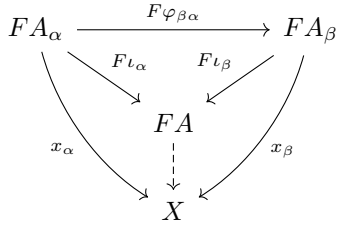
Lemma 3.57. Left adjoints preserve colimits. In other words, fix categories $\mathcal{A}, \mathcal{B}$ and an adjoint pair $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{A}$. Then suppose that we objects $\{A_\alpha\}_{\alpha \in \lambda}$ with commuting maps $\varphi_{\beta\alpha} : A_\alpha \to A_\beta$. (Given α, β, there might be no $\varphi + \beta\alpha$, or there might even be multiple.) Then

$$F\left(\varinjlim A_\alpha\right) \cong \varinjlim F(A_\alpha),$$

supposing that the colimit on the left exists.

Proof. We outline the proof that $F\left(\varinjlim A_\alpha\right)$ satisfies the universal property of $\varinjlim F(A_\alpha)$. For concreteness, set $A \coloneqq \varinjlim A_\alpha$, and let ι_α be the promised map $A_\alpha \to A$.

Now, fix any object $X \in \mathcal{B}$ with maps $x_\alpha : FA_\alpha \to X$ which commute with the $\varphi_{\beta\alpha}$ (i.e., $x_\beta \circ F\varphi_{\beta\alpha} = x_\alpha$). Here is our diagram, where we need to show that there is a unique induced arrow.



Well, $\mathrm{Hom}(FA, X) \cong \mathrm{Hom}(A, GX)$ (naturally) by the adjunction. But by definition of A as a colimit, we see that $\mathrm{Hom}(A, GX)$ is in natural isomorphism with commuting tuples of morphisms as in

$$\mathrm{Hom}(A, GX) \cong \left\{ \{a_\alpha\}_{\alpha \in \lambda} \in \prod_{\alpha \in \lambda} \mathrm{Hom}(A_\alpha, GX) : a_\beta \circ \varphi_{\beta\alpha} = \varphi_\beta \right\}.$$

But commuting tuples of morphisms in $\mathrm{Hom}(A_\alpha, GX)$ can be pushed back to $\mathrm{Hom}(FA_\alpha, X)$ by the adjunction again, and the fact that the adjunction natural means that the morphisms will commute afterwards as needed. So we have

$$\left\{ \{a_\alpha\}_{\alpha \in \lambda} \in \prod_{\alpha \in \lambda} \mathrm{Hom}(A_\alpha, GX) : a_\beta \circ \varphi_{\beta\alpha} = \varphi_\beta \right\} \cong \left\{ \{b_\alpha\}_{\alpha \in \lambda} \in \prod_{\alpha \in \lambda} \mathrm{Hom}(FA_\alpha, X) : b_\beta \circ F\varphi_{\beta\alpha} = F\varphi_\beta \right\}.$$

So in total,

$$\mathrm{Hom}(FA, X) \cong \left\{ \{b_\alpha\}_{\alpha \in \lambda} \in \prod_{\alpha \in \lambda} \mathrm{Hom}(FA_\alpha, X) : b_\beta \circ F\varphi_{\beta\alpha} = F\varphi_\beta \right\},$$

which is exactly what we need for FA to be the colimit of the FA_α. This finishes. ■

And now we can realize right-exactness as a special kind of colimit.

Lemma 3.58. Suppose that the functor of abelian categories $F : \mathcal{A} \to \mathcal{B}$ preserves colimits. (For example, F might be a left adjoint.) Then F is right exact.

Proof. The point is the right short exact sequence

$$A' \xrightarrow{\iota} A \xrightarrow{\pi} A'' \longrightarrow 0$$

is equivalent to saying that A'' is the colimit of the following diagram.

$$A' \underset{0}{\overset{\iota}{\rightrightarrows}} A$$

Indeed, the right short exact sequence is equivalent to $A'' \cong A/\operatorname{im}\iota$ by using the Homomorphism theorem, and $A/\operatorname{im}\iota$ is the colimit of the above: for any X with maps $A' \to X$ and $A \to X$ making the above commute, surely there is at most one map $A/\operatorname{im}\iota \to X$, and this map exists because $A \to X = A' \to A \to X$ implies that A' vanishes under $A \to X$.

Thus, because F preserves colimits, it will preserves quotients in the above way. Explicitly, if

$$A' \xrightarrow{\iota} A \xrightarrow{\pi} A'' \longrightarrow 0$$

is right exact, then

$$FA' \xrightarrow{\iota} FA \xrightarrow{\pi} FA'' \longrightarrow 0$$

will be right exact. ∎

Remark 3.59. We also have the dual statement that right adjoints preserves limits, which implies right adjoints preserve left exactness. For example, we could just move everything into an opposite category and repeat the proofs above.

Now Theorem 3.55 follows by stringing the above lemmas together. ∎

Remark 3.60. As promised, category theory is a nice tool for making trivial results trivial. However, it is not obvious that the result is trivial.

Category theory also gives us some other nice properties. For example, we have the following, practically for free.

Proposition 3.61. Given abelian groups $\{A_\alpha\}_{\alpha\in\lambda}$, we have

$$\left(\bigoplus_{\alpha\in\lambda} A_\alpha\right) \otimes B \cong \bigoplus_{\alpha\in\lambda} (A_\alpha \otimes B)$$

Proof. Direct sums are colimits (where there are no commuting morphisms to worry about), so this follows directly from Lemma 3.56 and Lemma 3.57. ∎

Proposition 3.62. We can show that $\operatorname{Hom}(B, -)$ is left exact.

Proof. The point is that right adjoints preserve limits, so $\mathrm{Hom}(B, -)$ preserves limits. Then, as before, we show that

$$0 \longrightarrow A' \xrightarrow{\iota} A \xrightarrow{\pi} A''$$

is left exact if and only if A' is the limit of the following diagram.

$$A \underset{0}{\overset{\pi}{\rightrightarrows}} A''$$

Indeed, the left short exact sequence is equivalent to $A' \cong \ker \pi$ by using the Homomorphism theorem. And this is equivalent to being the limit of the above diagram because for any X with maps $X \to A$ and $X \to A''$ causing everything to commute, we see that $X \to A$ must map into $\ker \pi \cong A'$, so the induced map exists and is unique by restricting the image.

So we see that

$$0 \longrightarrow A' \xrightarrow{\iota} A \xrightarrow{\pi} A''$$

is left exact if and only if A' is the limit of some diagram if and only if $F(A')$ is the limit of a similar diagram if and only if

$$0 \longrightarrow FA' \xrightarrow{F\iota} FA \xrightarrow{F\pi} FA''$$

is left exact. ∎

Similarly, $\mathrm{Hom}(B, -)$ preserves products in the same way that $- \otimes B$ preserves direct sums.

3.2.3 Back To Examples

Let's go back to examples.

Example 3.63. To compute $\mathbb{Z}/2\mathbb{Z} \otimes \mathbb{Z}/2\mathbb{Z}$, we look at the short exact sequence

$$0 \to \mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \to \mathbb{Z}\mathbb{Z}/2\mathbb{Z} \to 0.$$

Taking $- \otimes \mathbb{Z}/2\mathbb{Z}$ and using Remark 3.51 to keep track of the morphisms, we get right exact sequence

$$\mathbb{Z}/2\mathbb{Z} \xrightarrow{\times 2} \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0,$$

so $\mathbb{Z}/2\mathbb{Z} \otimes \mathbb{Z}/2\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z}$. (Namely, the $\mathbb{Z}/2\mathbb{Z} \xrightarrow{\times 2} \mathbb{Z}/2\mathbb{Z}$ at the front is the zero map.)

From the above example, we notice that the full sequence

$$0 \to \mathbb{Z}/2\mathbb{Z} \xrightarrow{\times 2} \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0$$

is not short exact, at the very least for size reasons but more immediately because the first $\mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ is not injective (it is the zero map). So indeed, tensor products do not preserve left exactness.

Example 3.64. Let's compute $\mathbb{Z}/2\mathbb{Z} \otimes \mathbb{Z}/3\mathbb{Z}$. Again, take

$$\mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0,$$

and we apply $- \otimes \mathbb{Z}/3\mathbb{Z}$. This gives us

$$\mathbb{Z}/3\mathbb{Z} \xrightarrow{\times 2} \mathbb{Z}/3\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \otimes \mathbb{Z}/3\mathbb{Z} \to 0.$$

However, the $\xrightarrow{\times 2}$ is surjective, so $\mathbb{Z}/2\mathbb{Z} \otimes \mathbb{Z}/3\mathbb{Z} \cong 0$.

So nonzero tensor products can give 0, sadly. Here is the general case.

Exercise 3.65. Fix m, n positive integers. Then $\mathbb{Z}/m\mathbb{Z} \otimes \mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/\gcd(m,n)\mathbb{Z}$.

Proof. Again, consider the exact sequence

$$\mathbb{Z} \xrightarrow{\times m} \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \to 0$$

and apply $- \otimes \mathbb{Z}/n\mathbb{Z}$ to get

$$\mathbb{Z}/n\mathbb{Z} \xrightarrow{\times m} \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \otimes \mathbb{Z}/n\mathbb{Z} \to 0.$$

Taking the quotient, our tensor product is

$$\frac{\mathbb{Z}/n\mathbb{Z}}{\operatorname{im} \xrightarrow{\times m}} \cong \frac{\mathbb{Z}/n\mathbb{Z}}{m\mathbb{Z}/n\mathbb{Z}} \cong \frac{\mathbb{Z}}{m\mathbb{Z}+n\mathbb{Z}} \cong \frac{\mathbb{Z}}{\gcd(m,n)\mathbb{Z}},$$

which is what we wanted. ∎

This gets us tensor products for finitely generated abelian groups by distributing Proposition 3.52 repeatedly while applying Exercise 3.65 to each distributed factor. The actual statement is somewhat obnoxious because a prime can appear multiple times, which is annoying to keep track of, so we will not write this out explicitly.

What about groups which are not finitely generated?

Example 3.66. We compute $\mathbb{Z}/n\mathbb{Z} \otimes \mathbb{Q}$ for n a positive integer. Well, take

$$\mathbb{Z} \xrightarrow{\times n} \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \to 0.$$

Applying $- \otimes \mathbb{Q}$, we get

$$\mathbb{Q} \xrightarrow{\times n} \mathbb{Q} \to \mathbb{Z}/n\mathbb{Z} \otimes \mathbb{Q} \to 0.$$

But now $\mathbb{Q} \xrightarrow{\times n} \mathbb{Q}$ is surjective, so $\mathbb{Z}/n\mathbb{Z} \otimes \mathbb{Q} \cong 0$. In general, $A \otimes \mathbb{Q}$ for A a finite group will vanish.

However, if we want to work more closely with $\mathbb{Q}$, we should realize it as a colimit. We claim that $\mathbb{Q}$ behaves as the colimit of the system

$$\mathbb{Z} \xrightarrow{\times 1} \mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \xrightarrow{\times 3} \mathbb{Z} \xrightarrow{\times 4} \cdots.$$

To see this, observe that this is the same system as

$$\mathbb{Z} \to \mathbb{Z} \to \frac{1}{2}\mathbb{Z} \to \frac{1}{6}\mathbb{Z} \to \cdots.$$

Indeed, $\mathbb{Q}$ is the colimit of this system because, for any A with maps $\frac{1}{n!}\mathbb{Z} \to A$ which commute nicely, we can induce the unique map $\mathbb{Q} \to A$ by taking any $\frac{p}{q}$ and running it through $\frac{1}{q!}\mathbb{Z} \to A$. This map is well-defined because the map $\frac{1}{n}\mathbb{Z} \to A$ commute nicely.

Namely, if we have an abelian group A and want to compute $A \otimes \mathbb{Q}$, then it is the colimit of the diagram

$$A \xrightarrow{\times 1} A \xrightarrow{\times 2} A \xrightarrow{\times 3} A \xrightarrow{\times 4} \cdots.$$

Now let's do some computations.

Example 3.67. We compute $\mathbb{Q} \otimes \mathbb{Q}$. From our work above, this will be the colimit of the diagram

$$\mathbb{Q} \xrightarrow{\times 1} \mathbb{Q} \xrightarrow{\times 2} \mathbb{Q} \xrightarrow{\times 3} \mathbb{Q} \xrightarrow{\times 4} \cdots.$$

However, each of the $\xrightarrow{\times n}$ maps are isomorphisms, so we can just embed all these groups into $\mathbb{Q}$. Explicitly, for any abelian group G with maps from the above system, we have a unique map $\mathbb{Q} \to G$ commuting with the above maps by using the leftmost $\mathbb{Q} \to G$, and this commutes because we had isomorphisms. So $\mathbb{Q} \otimes \mathbb{Q} \cong \mathbb{Q}$.

The above example was nice because applying $\otimes$ didn't lose injectivity, but we are not always so lucky.

Example 3.68. If we wanted to compute $\mathbb{Z}/2\mathbb{Z} \otimes \mathbb{Q}$, then we are computing the colimit of the diagram

$$\mathbb{Z}/2\mathbb{Z} \xrightarrow{\times 1} \mathbb{Z}/2\mathbb{Z} \xrightarrow{\times 2} \mathbb{Z}/2\mathbb{Z} \xrightarrow{\times 3} \mathbb{Z}/2\mathbb{Z} \xrightarrow{\times 4} \cdots,$$

but every other map is the zero map (and notably not injective!), so we just get $\mathbb{Z}/2\mathbb{Z} \otimes \mathbb{Q} \cong 0$.

Explicitly, for any abelian group G with maps from the above system, we see that commuting with the zero maps forces each $\mathbb{Z}/2\mathbb{Z} \to G$ to be the zero map: $\mathbb{Z}/2\mathbb{Z} \xrightarrow{\times(2n)} \mathbb{Z}/2\mathbb{Z} \to G$ is the zero map, and then $\mathbb{Z}/2\mathbb{Z} \xrightarrow{\times(2n+1)} \mathbb{Z}/2\mathbb{Z} \to G$ is the zero map because $\mathbb{Z}/2\mathbb{Z} \to G$ is zero by the previous case.

So that covers abelian groups pretty well. Here are some last exercises.

- Compute $\mathbb{Q} \otimes \mathbb{Q}/\mathbb{Z}$.
- Compute $\mathbb{Q}/\mathbb{Z} \otimes \mathbb{Z}/n\mathbb{Z}$.
- Compute $\mathbb{Q}/\mathbb{Z} \otimes \mathbb{Q}/\mathbb{Z}$.

3.2.4 Tensor Products over Commutative Rings

Our definition in general commutative rings is roughly the same as for abelian groups.

Definition 3.69 (Bilinear)**.** Fix R a commutative ring and (left) R-modules A, B, C. Then $f : A \times B \to C$ is *bilinear* if and only if, for each $a, a_1, a_2 \in A$ and $b, b_1, b_2 \in B$, we have

$$f(a, b_1 + b_2) = f(a, b_1) + f(a, b_2) \qquad \text{and} \qquad f(a_1 + a_2, b) = f(a_1, b) + f(a_2, b).$$

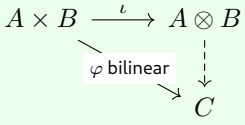
Additionally, we require, for each $r \in R$,

$$f(ra, b) = f(a, rb) = rf(a, b).$$

Observe that the second condition was automatic for $\mathbb{Z}$-modules by inducting off of the first condition. But general rings do not have access to such an induction, so we want to say this explicitly to more closely emulate an R-module homomorphism.

Anyways, we define tensor products by universal property again.

Definition 3.70 (Tensor products)**.** Fix R a commutative ring. Then for R-modules A and B, then we take the *tensor product* $A \otimes_R B$ to be "universal" as an R-module equipped with a bilinear map $\iota : A \times B \to A \otimes B$. Explicitly, for any bilinear map $\varphi : A \times B \to C$, there exists a unique induced homomorphism (!) $A \otimes B \to C$ making the following diagram commute.



We can quickly show that tensor products exist.

Proposition 3.71. Fix R a commutative ring. Then for R-modules A and B, $A \otimes_R B$ exists.

Proof. This construction is essentially the same as with abelian groups. Define N as the submodule of $A \oplus B$ generated by the elements

$$\begin{cases} (a, b_1) + (a, b_2) - (a, b_1 + b_2), \\ (b, a_1) + (b, a_2) - (b, a_1 + a_2), \\ (ra, b) - (a, rb), \end{cases}$$

for any $a, a_1, a_2 \in A$ and $b, b_1, b_2 \in B$ and $r \in R$. Then we define $A \otimes_R B \coloneqq (A \oplus B)/R$, with an R-action defined by $r(a \otimes b) \coloneqq (ra) \otimes b$. We will omit the checks here because they are essentially the same as in $\mathbb{Z}$-modules, though we do note that we add the condition $(ra, b) = (a, rb)$ because of the added condition to being bilinear. ■

We have many of the same properties. We will outline the properties bu tno more; they are pretty much the same as for $\mathbb{Z}$-modules.

Proposition 3.72. Fix R a commutative ring and A an R-module. Then we have $A \otimes_R R \cong A$.

Outline. We can show A satisfies the universal property $A \otimes_R R$ in essentially the same way as in $\mathbb{Z}$. At a high level, for any R-module X, we see from the adjunction (written out below) that

$$\mathrm{Hom}_R(A \otimes_R R, X) \cong \mathrm{Hom}_R(A, \mathrm{Hom}_R(R, X)),$$

but $\mathrm{Hom}_R(R, X) \cong X$ by tracking where 1 goes. So $\mathrm{Hom}_R(A \otimes_R R, X) \cong \mathrm{Hom}_R(A, X)$ for all R-modules X, so we are done by the Yoneda lemma. ■

If we actually track everything through, then again, the isomorphism $A \to A \otimes_R R$ is $a \mapsto a \otimes 1$, and the inverse mapping is $a \otimes r \mapsto ra$.

Proposition 3.73. We have that $- \otimes A$ is left adjoint to $\mathrm{Hom}_R(A, -)$, so $- \otimes_R A$ is right exact.

Proof. This is essentially the same proof as for abelian groups, so we won't say much here. We will remark that the extra bilinear condition on $B \otimes_R A$ corresponds to needing

$$\varphi(ra) = r\varphi(a)$$

for an R-module homomorphism $\varphi : A \to B$. ■

Example 3.74. Fix M a module and I an ideal of a commutative ring R, and we compute $M \otimes (R/I)$. For this we have the exact sequence

$$I \to R \to R/I \to 0$$

which becomes

$$M \otimes I \to M \otimes R \to M \otimes (R/I) \to 0$$

after applying $M \otimes -$. Tracking our quotient through, we see $M \otimes I \to M$ by $m \otimes i \mapsto im$, which surjects onto IM, so $M \otimes I \cong IM$. So we have $M \otimes (R/I) \cong M/IM$ here.

3.2.5 Tensor Products Over General Rings

In commutative rings R, we had the very nice property that $M \otimes_R N$ was an R-module for R-modules M and N by the linearity in the bottom. However, in general rings, the relations

$$rm \otimes n = m \otimes rn = r(m \otimes n)$$

are a bit fuzzy because it moves r from the outside left to the inside left, which are different! So in general rings, we should take

$$m \otimes rn = mr \otimes n,$$

where M is a right R-module and N is a left R-module, but now there is no good eay to make $M \otimes_R N$ is not an R-module, so $M \otimes_R N$ is merely an abelian group. We say this again.

Warning 3.75. For general rings, the functor $M \otimes_R -$ for general rings takes right R-modules to abelian groups, not R-modules to R-modules.

But we still have our definition as follows.

Definition 3.76 (Tensor product, I)**.** Fix M a right R-module and N a left R-module. The *tensor product* $M \otimes_R N$ takes funny bilinear maps $f : A \times B \to C$ (satisfying $f(ar, b) = f(a, rb)$) to linear maps $f : A \otimes B \to C$.

If we want to make this good again, we should take bimodules.

Definition 3.77 (Tensor products, II)**.** Fix M and N R-bimodules. Then the *tensor product* $M \otimes_R N$ imposes the conditions

$$ar \otimes b = a \otimes rb, \qquad r(a \otimes b) = (ra) \otimes b, \qquad a \otimes (br) = (a \otimes b)r,$$

where the last two laws turn $M \otimes_R N$ into an R-bimodule.

Note that the above roughly just includes the commutative case because right R-modules can be turned into right R-modules (by $r \cdot m \coloneqq mr$) when R is commutative.

3.2.6 More Applications and Examples

Let's have some fun.

Example 3.78. Fix k-vector spaces V and W, and we study $V \otimes W$. We claim that $W \cong k^{(\dim V)(\dim W)}$. Tangibly, we can fix bases $\{v_\alpha\}_{\alpha \in I}$ and $\{w_\beta\}_{\beta \in J}$ for V and W respectively, and then $V \otimes W$ will have basis given by

$$\{v_\alpha \otimes w_\beta\}_{(\alpha,\beta) \in I \times J}.$$

Checking linear independence is nontrivial, but we can see this because tensor products preserve direct sums, which implies

$$V \otimes W = \left(\bigoplus_{\alpha \in I} kv_\alpha \right) \otimes \left(\bigoplus_{\beta \in J} kw_\beta \right) \cong \bigoplus_{\alpha \in I} \left(kv_\alpha \otimes \bigoplus_{\beta \in J} kw_\beta \right) \cong \bigoplus_{(\alpha,\beta) \in I \times J)} k(v_\alpha \otimes w_\beta).$$

There is some work to track through the isomorphisms, but we have more or less done this in the notation above.

Example 3.79. Fix a (finite-dimensional) k-vector space V, and we study $W \coloneqq V \otimes V \otimes V \otimes V^*$, where V^* is the dual space. Then if V has a basis $\{v_k\}_{k=1}^{\dim V}$, then W has a basis

$$v_a \otimes v_b \otimes v_c \otimes v_*^d,$$

totaling to a dimension of $(\dim V)^4$. Again, these elements span W by looking component-wise, and these elements are linearly independent, roughly speaking, because there isn't a way to combine them meaningfully. Alternatively, we could just inductively apply the previous example.

In differential geometry, we might omit everything except the coefficients of the basis in the above example because they are a mess.

Tensor products also help out category theory (which is perhaps unsurprising).

Proposition 3.80. The coproduct of two commutative rings R and S is $R \otimes S$, where $R \otimes S$ is a ring with multiplication defined by extending

$$(r_1 \otimes s_1)(r_2 \otimes s_2) \coloneqq (r_1 r_2) \otimes (s_1 s_2)$$

linearly.

Proof. We already have that $R \otimes S$ is an abelian group (because we took the tensor product in $\mathbb{Z}$-modules), so checking that it is a ring only needs to worry about the multiplication law. Showing that multiplication is well-defined is surprisingly annoying; we do this in steps.

(i) We know that we have a bilinear map $R \times S \to R \otimes S$ by $(r, s) \mapsto r \otimes s$ is bilinear, and the distributive law in R and S promise that, for given $(r_0, s_0) \in R \times S$, the map $\mu_{(r_0,s_0)} : (r, s) \mapsto (r_0 r) \otimes (s_0 s)$ is still bilinear:

$$\mu_{(r_0,s_0)}(r_1+r_2, s) = (r_0(r_1+r_2)) \otimes (s_0 s) = (r_0 r_1) \otimes (s_0 s) + (r_0 r_2) \otimes (s_0 s) = \mu_{(r_0,s_0)}(r_1, s) + \mu_{(r_0,s_0)}(r_2, s),$$

and similarly,

$$\mu_{(r_0,s_0)}(r, s_1+s_2) = (r_0 r) \otimes (s_0(s_1+s_2)) = (r_0 r) \otimes (s_0 s_1) + (r_0 r) \otimes (s_0 s_2) = \mu_{(r_0,s_0)}(r, s_1) + \mu_{(r_0,s_0)}(r, s_2).$$

(ii) Because $\mu_{(r_0,s_0)} : R \times S \to R \otimes S$ is bilinear, it induces a linear map $R \otimes S \to R \otimes S$ by $r \otimes s \mapsto (r_0 r) \otimes (s_0 s)$.

(iii) In fact, we claim that $(r_0, s_0) \mapsto \mu_{(r_0,s_0)}$ is itself a bilinear map $R \times S \to \operatorname{Hom}(R \otimes S, R \otimes S)$. Indeed, we have to check that

$$\mu_{(r_1+r_2,s)}(r_0 \otimes s_0) = ((r_1 + r_2) r_0) \otimes (s s_0) = (r_1 r_0) \otimes (s s_0) + (r_2 r_0) \otimes (s s_0),$$

and

$$\mu_{(r,s_1+s_2)}(r_0 \otimes s_0) = (r r_0) \otimes ((s_1 + s_2) s_0) = (r r_0) \otimes (s_1 s_0) + (r r_0) \otimes (s_2 s_0)$$

and then these extend out to all of $R \otimes S$.

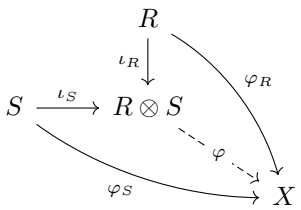
(iv) So because $\mu_\bullet : R \times S \to \operatorname{Hom}(R \otimes S, R \otimes S)$ is a bilinear map, we have a linear map $R \otimes S \to \operatorname{Hom}(R \otimes S, R \otimes S)$ by

$$(r_0 \otimes s_0) \mapsto ((r \otimes s) \mapsto (r_0 r \otimes s_0 s)),$$

which is exactly what we wanted.

To finish checking that $R \otimes S$ is a ring, associativity is inherited from R and S. Our identity is $1 \otimes 1$. The right distributive law holds because $\mu_\bullet$ is a group homomorphism, and then we can get the left distributive law because multiplication is commutative.

Now we have to actually check that $R \otimes S$ is the coproduct. To start, we see that we have inclusions $\iota_R : R \to R \otimes S$ and $\iota_S : S \to R \otimes S$ by $r \mapsto r \otimes 1$ and $s \mapsto s \otimes 1$ respectively. To show the universal property, fix X a ring with maps $\varphi_R : R \to X$ and $\varphi_S : S \to X$.



For the universal property, we need to induce φ uniquely.

- We start by showing it is unique; it suffices to show that $\varphi(r \otimes s)$ is forced for $r \in R$ and $s \in S$. Well, we see
$$\varphi(r \otimes s) = \varphi\big((r \otimes 1) \cdot (1 \otimes s)\big) = \varphi(r \otimes 1) \cdot \varphi(1 \otimes s) = (\varphi \circ \iota_R)(r)(\varphi \circ \iota_S)(s) = \varphi_R(r)\varphi_S(s),$$
which is now indeed forced.
- We now show that φ exists. Indeed, we note that the maps φ_R and φ_S induce a bilinear map $(r, s) \mapsto \varphi_R(r)\varphi_S(s)$; we won't write out the check that this is bilinear this time, but it comes from the distributive laws in X.

  The point is that the bilinear map $R \times S \to X$ induces a linear map $\varphi : R \otimes S \to X$ by
$$\varphi(r \otimes s) = \varphi_R(r)\varphi_S(s).$$
We have to actually show that φ is a ring map; we are already given that it is a group homomorphism. Then
$$\varphi((r_1r_2) \otimes (s_1s_2)) = \varphi_R(r_1r_2)\varphi_S(s_1s_2) = \big(\varphi_R(r_1)\varphi_S(s_1)\big)\big(\varphi_R(r_2)\varphi_S(s_2)\big)$$
shows φ respects multiplication, and we can see $\varphi(1 \otimes 1) = \varphi_R(1)\varphi_S(1) = 1 \cdot 1 = 1$, so φ also preserves the identity. This finishes. ■

Asn an aside, we note that we can do something similar for R-algebras.

Definition 3.81 (Algebra)**.** Given a ring R, an *R-algebra* is a commutative ring with an R-action.

Proposition 3.82. The tensor product is the coproduct in the category of R-algebras.

Proof. We omit this proof because I don't want to think about algebras. ■

In algebraic geometry, algebras are roughly schemes, and then the their tensor product is the "fiber product" of the schemes.

Remark 3.83. Rigorizing the above sentence takes about five hours of book-keeping.

Example 3.84. Fix $R = \mathbb{C}$ and $A = \mathbb{C}[x]$ with $B = \mathbb{C}[y]$ which are R-algebras. Well, these are really R-vector spaces, where A has a basis $\{x^k\}_{k \in \mathbb{N}}$ and B has a basis $\{y^\ell\}_{\ell \in \mathbb{N}}$, so $A \otimes_R B$ has a basis (as an R-module) $x^k \otimes y^\ell$, which is $\mathbb{C}[x, y]$.

Let's keep working with the above example. Taking spectrums, we see
$$\operatorname{Spec} \mathbb{C}[x] \approx \mathbb{C} \cup \{\infty\},$$
where ∞ corresponds to the zero ideal. What about $\operatorname{Spec} \mathbb{C}[x, y]$? Well, certainly some of its primes look like $(x - \alpha)$ or $(y - \beta)$ or $(x - \alpha, y - \beta)$, which correspond to (α, ∞) or (β, ∞) or (α, β) respectively.

But there are lots of other primes in $\operatorname{Spec} \mathbb{C}[x, y]$ to keep track of. For example,
$$\left(x^2 + y^2 - 1\right)$$
is not from anyone in $\operatorname{Spec} A$ or $\operatorname{Spec} B$. So in general, we do not always have
$$\operatorname{Spec} A \times \operatorname{Spec} B \stackrel{?}{=} \operatorname{Spec}(A \otimes_R B),$$
which is sad. Making these actually equal requires some care to redefine the product on the left.

3.2.7 Group Actions

We continue. Let's talk about representations.

Example 3.85. We compute $V \coloneqq \mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$. Viewing as an $\mathbb{R}$-vector space, V has a basis $1 \otimes 1$ and $1 \otimes i$ and $i \otimes 1$ and $i \otimes i$. Viewing V a ring, the elements

$$\frac{1 \otimes 1 - i \otimes i}{2} \qquad \text{and} \qquad \frac{1 \otimes i + 1 \otimes i}{2}$$

are orthogonal idempotents (I won't check this explicitly), so we have a decomposition of rings (!)

$$\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C}\left[\frac{1 \otimes 1 - i \otimes i}{2}\right] \oplus \mathbb{C}\left[\frac{1 \otimes i + 1 \otimes i}{2}\right],$$

and we can check that the $\mathbb{R}$-dimension on both sides is $2 \cdot 2 = 2 + 2$.

Remark 3.86. Tensor products of fields like this come up in algebraic number theory quite a bit.

For our story here, fix G a group which acts on the vector spaces V and W. Then G acts on $V \oplus W$ pointwise, and in fact G acts on $V \otimes W$ by

$$g(v \otimes w) = (gv) \otimes (gw)$$

for $g \in G$ and $v \in V$ and $w \in W$. Indeed, the map $\mu_g : V \times W \to V \otimes W$ by $(v, w) \mapsto (gv \otimes gw)$ is bilinear because $(v, w) \mapsto (v \otimes w)$ is, and $g \mapsto \mu_g$ is a group homomorphism because G acts on V and W. (We won't write these out.)

Example 3.87. Take $G = \mathbb{Z}/n\mathbb{Z}$, we can take $V = W = \mathbb{C}$ as $\mathbb{C}$-vector spaces, where the G-action on V is given by $g \cdot z \coloneqq ze^{2\pi iag/n}$ for some fixed $a \in \mathbb{Z}$, and the G-action on W is given by $g \cdot z \coloneqq ze^{2\pi ibg/n}$ for some fixed $b \in \mathbb{Z}$. Then

$$g(v \otimes w) = (gv) \otimes (gw) = e^{2\pi i(a+b)g/n}(v \otimes w)$$

is our G-action on $V \otimes W$.

Recall from earlier that we had the Burnside ring of (equivalence classes of) sets with a G-action. The above ideas let us maybe define an arithmetic on (equivalence classes of) linear representations of G. Here we define

$$V + W \coloneqq V \oplus W \qquad \text{and} \qquad V \times W \coloneqq V \otimes W.$$

From earlier we had a distributive law

$$(A \oplus B) \otimes C \cong (A \otimes C) \oplus (B \otimes C).$$

We even have nice association

$$(A \otimes B) \otimes C \cong A \otimes (B \otimes C),$$

which is simply by $(a \otimes b) \otimes c \mapsto a \otimes (b \otimes c)$. So we have most of what we need for a ring!

But again, we have no subtraction, but we can do a similar construction as with the Burnside ring, where we just forced a subtraction to exist. This requires some care because it is possible for $A \oplus R \cong B \oplus R$ while $A \ncong B$ for general rings R, as we saw last class. Regardless, this is still possible, and gives us the representation ring.

Definition 3.88. Fix a field k. The *representation ring* of a group G is the ring more or less generated by the k-linear representations of G with addition given by $\oplus$ and multiplication given by $\otimes$.

3.3 October 12

I am a small boat, and these are big waves.

3.3.1 Duality for Vector Spaces

Today is module miscellany. Recall duality for vector spaces.

Definition 3.89 (Vector space duality)**.** Given a k-vector space V, we define

$$V^* \coloneqq \mathrm{Hom}_k(V, k).$$

Then we know there is a natural map $V \to V^{**}$, given by

$$v \mapsto (\varphi \mapsto \varphi v),$$

which is notably canonical. This is an isomorphism if V is finite-dimensional because we can check $V \to V^{**}$ is injective[4] and $\dim V = \dim V^* = \dim V^{**}$ shows that we are bijective for size reasons.

However, these size reasons are no longer valid for infinite-dimensional vector spaces, so we might have V^* larger than V.

Non-Example 3.90. Fix a k-vector space V with basis given by $\{v_n\}_{n\in\mathbb{N}}$ such that

$$V = \bigoplus_{n\in\mathbb{N}} kv_n.$$

Now we claim that $V^* \cong k^{\mathbb{N}}$, which is of strictly larger cardinality that $V \cong k^{\oplus\mathbb{N}}$. Indeed, we associate $\{a_n\}_{n\in\mathbb{N}}$ with the linear map

$$\sum_{n=1}^{\infty} k_n v_n \mapsto \sum_{n=1}^{\infty} a_n k_n.$$

The sum converges because all but finitely many terms are nonzero. Now, certainly each $\{a_n\}_{n\in\mathbb{N}}$ is a linear map, and all linear maps take this form by tracking where each individual $v_\bullet$ goes.

In analysis, we usually put a topology on V, which makes things better to name.

Before going into the next example, we take the following definition.

Definition 3.91 (L^p spaces)**.** Fix X an integrable space. Then, for a real number $p > 0$, we define the space $L^p(X)$ to consist of integrable functions $f : X \to \mathbb{R}$ such that

$$\left(\int_X |f(x)|^p \, dx\right)^{1/p} \in \mathbb{R}.$$

In line with this definition, we define $L^\infty(X)$ to consist of bounded integrable functions.

Our example will take $X = \mathbb{N}$, where $\int_X dx$ turns into $\sum_{n\in\mathbb{N}}$.

Example 3.92. Fix $V \coloneqq C_0(\mathbb{N})$, which consists all (continuous) sequences $\mathbb{N} \to \mathbb{R}$ which tend to 0, and we note we have a topology induced by

$$\sup_n |c_n|$$

where $\{c_n\}_{n\in\mathbb{N}} \in V$.

[4] If $\varphi \mapsto \varphi v$ is the same as $\varphi \mapsto \varphi w$ for each $\varphi \in V^*$, then fixing a basis $\{\beta_\alpha\}_{\alpha\in\lambda}$, φ projecting onto a basis element detects $v = w$

Let's talk through some of the duals of $V = C_0(\mathbb{N})$.

- We see $V^* \cong L^1(\mathbb{N})$. Indeed, for each $\{d_n\}_{n\in\mathbb{N}} \in L^1(\mathbb{N})$, we have the linear map

$$\{c_n\}_{n\in\mathbb{N}} \mapsto \sum_{n\in\mathbb{N}} c_n d_n,$$

  which converges because the c_k are bounded. In one direction, certainly all of these are linear maps. In the other direction, suppose $T : V \to \mathbb{R}$ is a linear transformation. Not by caring convergence too much, we note that

$$T\left(\sum_{k=1}^{\infty} c_k\right) = \sum_{k=1}^{\infty} c_k \underbrace{T\left(\{1_{\ell=k}\}_{\ell\in\mathbb{N}}\right)}_{=:d_k},$$

  where this works over finite sums and extends to infinite sums in the limit. In particular, because the left-hand side must converge, the right-hand side needs to converge as well, so $d_k \to 0$ as $k \to \infty$.

  It remains to show that $\{d_k\}_{k\in\mathbb{N}} \in L^1(\mathbb{N})$. Suppose for the sake of contradiction that

$$\sum_{k\in\mathbb{N}} |d_k| = \infty.$$

  Then, we set $m = 0$ and say that for each $m \geq 1$, there exists $n_m \geq n_{m-1}$ such that

$$\sum_{k=n_{m-1}+1}^{n_m} |d_k| > 1.$$

  Now we set $c_k = \operatorname{sgn}(d_k)\frac{1}{m}$, where $n_{m-1} < k \leq n_m$. The point is that

$$\sum_{k=1}^{\infty} d_k c_k = \sum_{m=1}^{\infty} \sum_{k=n_{m-1}+1}^{n_m} d_k c_k = \sum_{m=1}^{\infty} \frac{1}{m} \sum_{k=n_{m-1}+1}^{n_m} |d_k| < \sum_{m=1}^{\infty} \frac{1}{m} = \infty,$$

  which contradicts the $\{d_k\}_{k\in\mathbb{N}}$ defining an element of V^*.

- We see $V^{**} \cong L^\infty(\mathbb{N})$ using a similar argument as above. I am too lazy to work this out.
- Continuing V^{***} has to do with "contents" of $\mathbb{N}$, but at this point, we need the axiom of choice to find an example which isn't from $V^* = L^1(\mathbb{N})$. Here we'll stop.

So the above examples show a strictly ascending chain of double duals even though, say, V and V^{**} do have the same cardinality.

We also have the following general result for L^p spaces.

Exercise 3.93. Fixing $p \in (0, \infty)$, we have that $L^p(\mathbb{N})^* = L^q(\mathbb{N})$, where $q \in \mathbb{R}$ is chosen with $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. Again, the point is that we can associate $\{d_n\}_{n\in\mathbb{N}} \in L^q(\mathbb{N})$ with the linear map

$$\{c_n\}_{n\in\mathbb{N}} \mapsto \sum_{n\in\mathbb{N}} c_n d_n.$$

Indeed, this converges by Hölder's inequality: we have

$$\sum_{n\in\mathbb{N}} |c_n d_n| \leq \left(\sum_{n\in\mathbb{N}} |c_n|^p\right)^{1/p} \left(\sum_{n\in\mathbb{N}} |c_n|^q\right)^{1/q} < \infty.$$

And of course these functions are linear. We will omit the proof that this mapping is bijective because I fear it is nontrivial, and I am lazy. ∎

3.3.2 Duality for General Rings

Now we turn to generalizing duality from vector spaces because we're algebraists.

Definition 3.94 (Duality for free and projective modules)**.** Fix M an R-module over a commutative ring R, and we can define the *dual module* $F^* := \mathrm{Hom}_R(F, R)$.

Remark 3.95. We use commutative rings because noncommutative rings make Professor Borcherds nervous.

Note that this definition makes sense because $\mathrm{Hom}_R(F, R)$ is an R-module by

$$(r\varphi)(x) = r \cdot \varphi(x)$$

for $r \in R$ and $x \in F$.

Some of the theory for vector spaces carries over nicely.

Proposition 3.96. Given a free R-module F over a commutative ring, we have a canonical injection $F \hookrightarrow F^{**}$. If F is of finite rank, then $F \cong F^{**}$.

Proof. This is the same as for vector spaces. For $m \in F$ and $\varphi \in F*$, the mapping is

$$\psi_\bullet : m \mapsto (\varphi \mapsto \varphi m).$$

We see $\psi_\bullet$ is linear because, given $r_1, r_2 \in R$ and $m_1, m_2 \in F$, we have

$$\psi_{r_1 m_1 + r_2 m_2}(\varphi) = \varphi(r_1 m_1 + r_2 m_2) = r_1 \varphi(m_1) + r_2 \varphi(m_2) = r_1 \psi_{m_1}(\varphi) + r_2 \psi_{m_2}(\varphi)$$

by plugging into the various module actions.

We can also check that this is injective: suppose that $m \in \ker \psi_\bullet$ so that $\psi_m : F^* \to R$ is the zero mapping. The key point is that F being free promises it is freely generated by some set $\{m_\alpha\}_{\alpha \in \lambda}$, and we note that

$$\pi_\alpha : \sum_{\alpha \in \lambda} r_\alpha m_\alpha \mapsto r_\alpha$$

is a linear transformation, well-defined because the $\{m_\alpha\}_{\alpha \in \lambda}$ are a basis. Then we note that $\psi_m(\pi_\alpha) = \pi_\alpha = m$, so each component of m under the basis will vanish. Thus, $m = 0$.

To show that $F \cong F^{**}$ when F is of finite rank, we actually show that $F \cong F^*$, non-canonically. Indeed, letting our basis be $\{m_k\}_{k=1}^n$ where $d = \operatorname{rank} F$, we see we have an isomorphism $R^n \cong F^*$ by

$$(a_1, \dots, a_n) \mapsto \left(\sum_{k=1}^n a_k m_k \mapsto \sum_{k=1}^n a_k r_k \right),$$

where $(a_1, \dots, a_n) \in R^n$ and $\sum_{k=1}^n a_k m_k$ is an arbitrary element of F. We won't bother showing that this is an isomorphism. ∎

The following also holds more generally.

Lemma 3.97. Given R-modules A, B, X we have that

$$\mathrm{Hom}_R(A \oplus B, X) \cong \mathrm{Hom}_R(A, X) \oplus \mathrm{Hom}_R(B, X).$$

Proof. This is essentially the universal property of $A \oplus B$: maps $A \oplus B \to X$ are in bijection with maps $A \to X$ and $B \to X$. We won't check that this is an R-module homomorphism and so on. ∎

The point of Lemma 3.97 is to give the following.

Proposition 3.98. Given a projective R-module P over a commutative ring, we have a canonical, injective morphism $P \hookrightarrow P^{**}$. If P is finitely generated, this is an isomorphism.

Proof. The point is that there is an R-module Q such that $F \coloneqq P \oplus Q$ is free, and if P is finitely generated, we can force F to be finitely generated. To start, our canonical injective homomorphism is $\psi_\bullet^P : P \to P^{**}$ defined by

$$\psi_\bullet^P : p \mapsto (\varphi \mapsto \varphi p),$$

and we define $\psi_\bullet^Q$; as usual, we won't check that this is an R-module homomorphism and so on.

We start by showing $\psi_\bullet^P$ is injective and that $\psi_\bullet^Q$ is injective will be similar. Indeed, if $p \in \ker \psi_\bullet^P$, then ψ_p^P is the zero map so that $\varphi(p) = 0$ for each $\varphi : P \to R$. But now, for each $\overline{\varphi} : F \to R$, we see $\overline{\varphi} = \varphi^P + \varphi^Q$ for $\varphi^P : P \to R$ and $\varphi^Q \to R$ by universal property, so

$$\overline{\varphi}(p, 0) = \varphi^P(p) + \varphi^Q(0) = 0 + 0 = 0$$

for each $\overline{\varphi} : F \to R$. But we know that the only element of F which vanishes under all morphisms $F \to R$ is $(0, 0)$, so we must have $p = 0$. This finishes.

To show that $\psi_\bullet^P : P \hookrightarrow P^{**}$ is an isomorphism when P is finitely generated, we bound the size of P^{**}. Indeed, we note that we have the isomorphisms

$$\begin{aligned}\operatorname{Hom}_R(\operatorname{Hom}_R(P \oplus Q, R), R) &\cong \operatorname{Hom}_R(\operatorname{Hom}_R(P, R) \oplus \operatorname{Hom}_R(Q, R), R) \\ &\cong \operatorname{Hom}_R(\operatorname{Hom}_R(P, R), R) \oplus \operatorname{Hom}_R(\operatorname{Hom}_R(Q, R), R)\end{aligned}$$

by repeatedly applying Lemma 3.97. But now we have maps

$$F = P \oplus Q \hookrightarrow P^{**} \oplus Q^{**} \cong F^{**} \cong F,$$

where all maps are injective, and in fact the composition is id_F if we track everything through.[5] It follows that our map $P \hookrightarrow P^{**}$ and $Q \hookrightarrow Q^{**}$ are actually isomorphisms. ∎

Remark 3.99. It is not necessarily true that $P \cong P^*$ when P is a finitely generated projective module; there is an example here.

Anyways, we should see some examples.

Example 3.100. For $R = \mathbb{Z}$ and $M = \mathbb{Z}/2\mathbb{Z}$, we have $M^* = \operatorname{Hom}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}) = 0$, which is not very interesting.

This is not interesting because we've immediately killed our module. So here is another definition which works better for abelian groups.

Definition 3.101 (Duality for abelian groups)**.** For M an abelian group, define

$$M^* \coloneqq \operatorname{Hom}_{\mathbb{Z}}(M, \mathbb{Q}/\mathbb{Z}).$$

Remark 3.102. We have chosen $\mathbb{Q}/\mathbb{Z}$ for our "dualizing object" instead of $\mathbb{Z}$ because we can do what we want. In specific cases, there might be good reasons to choose a different dualizing object than our original ring.

We can check the following, continuing the idea that double duals should behave well.

[5] Please don't ask me to actually track this through.

Proposition 3.103. For M a finitely generated abelian group. Then $M^{**} \cong M$.

Proof. The proof proceeds in steps.

1. We start by checking cyclic groups. We see
$$\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z}, \mathbb{Q}/\mathbb{Z}) \cong \mathbb{Z}/n\mathbb{Z}$$
because an element of $\mathbb{Z}/n\mathbb{Z}$ must map to an element of (additive) order n, so $1 \mapsto \frac{k}{n}$ for some $k \in \mathbb{Z}/n\mathbb{Z}$. So our maps are in bijection to $\mathbb{Z}/n\mathbb{Z}$, and it is not too hard to check that the map $\mathbb{Z}/n\mathbb{Z} \to \mathrm{Hom}(\mathbb{Z}/n\mathbb{Z}, \mathbb{Q}/\mathbb{Z})$ is in fact homomorphic.
Thus, $(\mathbb{Z}/n\mathbb{Z})^* \cong \mathbb{Z}/n\mathbb{Z}$.
2. We note that Lemma 3.97 implies that $(M \oplus N)^{**} \cong M^{**} \oplus N^{**}$ as before. In particular, if we know $M^{**} \cong M$ and $N^{**} \cong N$ already, then we get $(M \oplus N)^{**} \cong M^{**} \oplus N^{**}$.
3. To finish, any finitely generated abelian group M is the direct sum of some cyclic group, so we finish by applying 2 and then 1. ∎

Even though the above example worked so nicely, we lost things being canonical. Namely, the isomorphism $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z}, \mathbb{Q}/\mathbb{Z}) \cong \mathbb{Z}/n\mathbb{Z}$ was not canonical because we had to choose the generator $1 \in \mathbb{Z}/n\mathbb{Z}$. Regardless, we do still have the canonical isomorphism $M \cong M^{**}$, where the homomorphism is canonical and bijective for size reasons.

3.3.3 Fourier Analysis

Let's do some more examples. In what follows, we use S^1 as the dualizing object for our abelian groups instead of $\mathbb{Q}/\mathbb{Z}$; to make the distinction clear, we have the following definition.

Definition 3.104 (Character)**.** Given an abelian group M, χ is a *character* if and only if $\chi \in \mathrm{Hom}(M, S^1)$.

If the abelian group G was finite to begin with, then this is the same as G^* from earlier because each $g \in G$ must map into
$$e^{2\pi i k/\#G} \text{ for some } k \in \mathbb{Z},$$
which is in the image of $\mathbb{Q}/\mathbb{Z}$ in S^1. If our abelian group G was infinite to begin with, then it likely has some topology going on, so it still makes sense to use S^1 instead of $\mathbb{Q}/\mathbb{Z}$.

Example 3.105. For $(\mathbb{Z}/8\mathbb{Z})^\times$, we see that $(\mathbb{Z}/8\mathbb{Z})^\times = \langle 3, 5\rangle \cong (\mathbb{Z}/2\mathbb{Z})^2$. So we can write out our characters explicitly by tracking where 3 and 5 go.

	1	3	5	7
χ_0	1	1	1	1
χ_1	1	-1	1	-1
χ_2	1	-1	-1	1
χ_3	1	1	-1	-1

Remark 3.106. Dirichlet's original use of Dirichlet characters $\chi : (\mathbb{Z}/m\mathbb{Z})^\times \to S^1$ was to work with L-series of the form
$$L(\chi, s) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^2},$$
where $\chi(n) = 0$ when $\gcd(m, n) > 1$. If you are interested in why he cared about such things, read up on analytic number theory.

These ideas give us some notion of a Fourier transform. The following is our main result here.

Theorem 3.107. Fix M some finite abelian group. Then we can define the (Hermitian) inner product on functions $f, g : M \to \mathbb{C}$ by

$$\langle f, g \rangle \coloneqq \sum_{x \in M} f(x)\overline{g(x)}.$$

Then the characters in $\mathrm{Hom}_{\mathbb{Z}}(M, S^1)$ form an orthogonal basis of these functions in $\mathrm{Mor}(M, \mathbb{C})$.

The idea here is to generalize Fourier series, where we have as our specific case the abelian group $\mathbb{R}/\mathbb{Z}$, and we have

$$f(x) = \sum_{k \in \mathbb{Z}} c_k e^{2\pi i k x}$$

where

$$c_k \coloneqq \int f(x) e^{-2\pi i n x}\, dx.$$

Proof. We show the claims in reverse order.

- We start by showing that distinct characters are orthogonal. Indeed, pick up $\chi_1, \chi_2 : \mathrm{Hom}(M, S^1)$. Then we have
$$\sum_{x \in M} \chi_1(x)\overline{\chi_2}(x) = \sum_{x \in M} (\chi_1 \chi_2^{-1})(x).$$
If $\chi_1 = \chi_2$, then all entries are the trivial character $\chi_0 \equiv 1$, so we get out $\#M$ from the sum. On the other hand, we claim that
$$\sum_{x \in M} \chi(x) = 0$$
where χ is not the trivial character. Explicitly, take some $y \in M$ such that $\chi(y) \neq 1$. Then
$$\chi(y) \sum_{x \in M} \chi(x) = \sum_{x \in M} \chi(xy) = \sum_{xy \in M} \chi(xy).$$
So we are forced to conclude that
$$(1 - \chi(y)) \sum_{x \in M} \chi(x) = 0,$$
so $\sum_{x \in M} \chi(x) = 0$.
So in total, we find that
$$\langle \chi_1, \chi_2 \rangle = \sum_{x \in M} \chi_1(x)\overline{\chi_2(x)} = \begin{cases} \#M & \chi_1 = \chi_2, \\ 0 & \chi_1 \neq \chi_2. \end{cases}$$
In particular, distinct characters are indeed orthogonal.
- To show that characters span $\mathrm{Mor}(M, \mathbb{C})$, fix any $f \in \mathrm{Mor}(M, \mathbb{C})$, and we claim
$$f \overset{?}{=} \frac{1}{\#M} \sum_{\chi \in \mathrm{Hom}(M, S^1)} \langle f, \chi \rangle \chi.$$
Expand this out, we are interested in evaluating, for some $x \in M$,
$$\sum_{\chi \in \mathrm{Hom}(M, S^1)} \langle f, \chi \rangle \chi(x) = \sum_{\chi} \left(\sum_{y \in M} f(y)\overline{\chi}(y) \right) \chi(x) = \sum_{y \in M} f(y) \sum_{\chi} \chi(xy^{-1}).$$
We claim that all terms except $x = y$ vanish over the first sum, where the sum reads $\sum_{\chi} \chi(e) = \#M$. Well, if $x \neq y$, fix $z \coloneqq xy^{-1} \neq e$ so that we want to evaluate
$$\sum_{\chi} \chi(z).$$

The main point is that there is some character χ_1 such that $\chi(z) \neq 1$ because $z \neq e$.[6] Then we see that

$$\chi_1(x) \sum_\chi \chi(z) = \sum_\chi (\chi_1 \chi)(z) = \sum_\chi \chi(z),$$

so $\chi_1(x) \neq 1$ forces $\sum_\chi \chi(z) = 0$.

To finish, we see that

$$\sum_{\chi \in \mathrm{Hom}(M, S^1)} \langle f, \chi \rangle \chi(x) = \sum_{y \in M} f(y) \sum_\chi \chi(xy^{-1}) = f(x) \# M,$$

which is what we wanted. ∎

The above also holds in some form in more generality for locally compact abelian groups.

Example 3.108. Consider the following two character duals. (We won't prove these in detail because they would take us too far afield.)

- For $G = \mathbb{Z}$, we have $\mathrm{Hom}(\mathbb{Z}, S^1) \cong S^1$ by tracking where 1 goes.
- For $G = S^1$, we have $\mathrm{Hom}(S^1, S^1) \cong \mathbb{Z}$ because all such homomorphism take the form $z \mapsto z^n$ for $n \in \mathbb{N}$.

These together give us the theory of Fourier series.

Example 3.109. For $G = \mathbb{R}$, we have $\mathrm{Hom}(\mathbb{R}, S^1) \cong \mathbb{R}$ by $y \mapsto (x \mapsto e^{2\pi i x y})$. The above is the theory for the Fourier transform.

Example 3.110. For $G = \mathbb{Q}_p$, we still have $\mathrm{Hom}(\mathbb{Q}_p, S^1) \cong \mathbb{Q}_p$.

Remark 3.111 (Nir)**.** It is a remarkable fact that $\mathrm{Hom}(\mathbb{Q}_\nu, S^1) \cong \mathbb{Q}_\nu$, even though non-canonically. The correct theory here turns out to be the fact that the $\mathbb{Q}_\nu$ are "local fields."

Remark 3.112. A lot of number theory has to do with Fourier analysis on things like $\mathbb{Q}_p$ or $\mathbb{A}_\mathbb{Q}$.

3.3.4 Injective Modules for Abelian Groups

Injective modules are roughly dual to projective modules, where duality does not mean what we have been talking about so far. So we take the definition of projective and reverse the arrows. Here is the definition of projective.

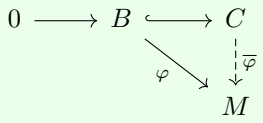
Definition 3.113 (Projective)**.** A module M is *projective* if and only if each surjection $B \twoheadrightarrow C$ with a map $\varphi : M \to C$, then we have a lifting map $\overline{\varphi} : M \to B$ making the diagram commute.

$$\begin{array}{ccccc} M & & & & \\ {\scriptstyle \overline{\varphi}}\downarrow & \searrow^{\varphi} & & & \\ B & \twoheadrightarrow & C & \longrightarrow & 0 \end{array}$$

And now we reverse the arrows.

[6] This is surprisingly technical. One way to do this is to decompose $M \cong \bigoplus_{k=1}^N \mathbb{Z}/n_k\mathbb{Z}$, find some coordinate $\mathbb{Z}/n_\bullet\mathbb{Z}$ where z is nonzero, and then send $1 \in \mathbb{Z}/n_\bullet\mathbb{Z} \mapsto e^{2\pi i/n_\bullet}$ while the other coordinates are sent to 1.

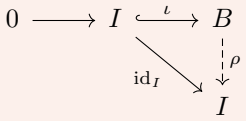
Definition 3.114 (Injective)**.** A module M is *injective* if and only if each injection $B \hookrightarrow C$ with a map $M \to B \to M$, then we have a lifting map $C \to M$ making the diagram commute.



Remark 3.115 (Nir)**.** Here is one reason why injective modules are nice; fix I an injective module. Dual to projective modules, any short exact sequence

$$0 \to I \xrightarrow{\iota} B \xrightarrow{\pi} C \to 0$$

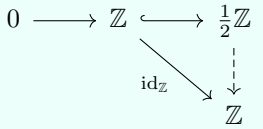
will split. The way to see this is that the diagram



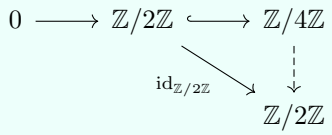
gives us some $\rho : B \to I$ such that $\rho \circ \iota = \mathrm{id}_I$. This ρ can be used to induce an isomorphism $B \cong I \oplus C$ by $b \mapsto (\rho b, \pi b)$; we won't actually check that this is an isomorphism here.

At a high level, being injective means that each homomorphism from a submodule to an injective module extends to a homomorphism from the full module. Let's try to find some injective modules.

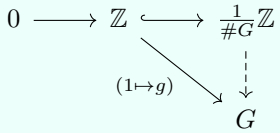
Non-Example 3.116. We have that $\mathbb{Z}$ is not injective because, for $\mathbb{Z} \subseteq \frac{1}{2}\mathbb{Z}$, we cannot extend $\mathrm{id}_{\mathbb{Z}} : \mathbb{Z} \to \mathbb{Z}$ to a full map $\frac{1}{2}\mathbb{Z} \to \mathbb{Z}$ because $\mathbb{Z}$ has no element which squares to 1. Here is the diagram.



Non-Example 3.117. We have that $\mathbb{Z}/2\mathbb{Z}$ is not injective because, for $\mathbb{Z}/2\mathbb{Z} \hookrightarrow \mathbb{Z}/4\mathbb{Z}$, we cannot extend $\mathrm{id}_{\mathbb{Z}/2\mathbb{Z}} : \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$. The problem is that no map $\mathbb{Z}/4\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ sends $2 \to 1$. Here is the diagram.



Non-Example 3.118. More generally, no nonzero finite abelian group G is injective. For example, we can use any $g \in G \setminus \{0\}$ to fix a map $\mathbb{Z} \to G$ by $1 \mapsto g$, but then this cannot be extended to $\frac{1}{\#G}\mathbb{Z}$ because every element $h \in G$ has $\#G \cdot h = e \neq g$.



Example 3.119. The group $\mathbb{Q}$ is injective. This will be true because $\mathbb{Q}$ is "divisible."

Roughly speaking, the problem with $\mathbb{Z}$ being injective is that we could not "divide by 2," but $\mathbb{Q}$ has no such problems. So we have the following property.

Definition 3.120. Fix G an abelian group. Then G is *divisible* if and only if the map $x \mapsto nx$ for any $n \in \mathbb{Z}^+$ is surjective.

Proposition 3.121. We have that M an injective abelian group if and only if M is divisible.

Proof. We show this in two parts.

- Take M injective and $n \in \mathbb{Z}^+$. Fixing any $m \in M$, we need to show that there is $x \in M$ with $m = nx$. The key point is the following diagram, well-defined because $n > 0$. Set $\varphi : \mathbb{Z} \to M$ by $1 \mapsto m$.

$$\begin{array}{ccccc} 0 & \longrightarrow & \mathbb{Z} & \hookrightarrow & \frac{1}{n}\mathbb{Z} \\ & & & \searrow^{\varphi} & \downarrow{\scriptstyle \overline{\varphi}} \\ & & & & M \end{array}$$

  Now injectivity induces $\overline{\varphi} : \frac{1}{n}\mathbb{Z} \to M$ such that $1 \mapsto m$. But then $n \cdot \overline{\varphi}\left(\frac{1}{n}\right) = \overline{\varphi}(1) = m$, so $\overline{\varphi}\left(\frac{1}{n}\right)$ is the desired element of M.

- Take M divisible, and we want to show M is injective. We are given an injection $C \hookrightarrow B$ with a map $\varphi : C \to M$ which we want to extend to a map $\overline{\varphi} : B \to M$. Well, take any $b \in B$. We have two cases.

  (a) If $nb \notin \mathbb{C}$ for all $n \in \mathbb{Z} \setminus \{0\}$, then $\langle C, b\rangle \cong C \oplus \mathbb{Z}$, so we can send $f : b \mapsto 0$.

  (b) Otherwise, suppose $nb \in C$ where n is the least such positive integer. But now M is divisible, so $f(nb) = ny$ for some $y \in M$, so we can send $f : b \mapsto y$.

  Now we invoke the Axiom of choice (specifically Zorn's lemma) to extend this all the way up to B. ∎

It turns out that divisible implies injective even for principal ideal domains, but we won't show this.

Remark 3.122. Roughly speaking, it is harder to find injective modules than projective modules. Namely, we needed the axiom of choice for the above proof.

Anyways, let's see some more examples.

Example 3.123. We have that $\mathbb{Q}/\mathbb{Z}$ is injective because it is divisible: for any $q \in \mathbb{Q}$ and $n \in \mathbb{Z}^+$, we have $\frac{q}{n} \overset{\times n}{\mapsto} q$. In fact, we can split up

$$\mathbb{Q}/\mathbb{Z} \cong \bigoplus_p \underbrace{\left\{x \in \mathbb{Q}/\mathbb{Z} : p^k x = 0 \text{ for some } k \in \mathbb{Z}\right\}}_{M_p :=},$$

which is roughly the Chinese remainder theorem. (We noted this many lectures ago.) It happens that each M_p is also injective, again because they are divisible. Namely, for any $M_p \in \mathbb{Q}$ we want to hit and $n \in \mathbb{Z}^+$, we can decompose $n = p^\nu m$ where $p \nmid m$. Then there is m' so that $mm' \equiv 1 \pmod{p^\nu}$ so that $n \cdot \frac{m'}{p^\nu} = 1$ and $n \cdot \left(\frac{m'}{p^\nu} \cdot q\right) = q$.

3.3.5 Injective Modules for General Rings

So let's find injective modules for more general rings. We claim the following.

Proposition 3.124. Given a ring R, we have that $\mathbb{Z}$-module $R^* \coloneqq \mathrm{Hom}_{\mathbb{Z}}(R, I)$ is an injective R-module, where I is a divisible abelian group.

Proof. The point is that being injective requires control of $\mathrm{Hom}_R(M, R^*)$. This is somewhat confusing because R^* is itself a Hom-set.

Regardless, R^* is in fact an R-module with R-action defined by

$$(r\varphi)(q) \coloneqq \varphi(qr).$$

Here multiplication is on the right because we want $\big((r_1r_2)\varphi\big)(q) = \varphi(qr_1r_2)$ to be equal to $\big(r_1(r_2\varphi)\big)(q) = (r_2\varphi)(qr_1) = \varphi(qr_1r_2)$. We will repeat this because it is confusing.

Warning 3.125. If M is a right R-module and G is an abelian group, then $\mathrm{Hom}_{\mathbb{Z}}(M, G)$ is a left R-module, and conversely.

The main claim is that, for any fixed R-module M,

$$\mathrm{Hom}_R(M, \mathrm{Hom}_{\mathbb{Z}}(R, I)) \cong \mathrm{Hom}_{\mathbb{Z}}(M, I).$$

Indeed, we take $\psi_\bullet \in \mathrm{Hom}_R(M, \mathrm{Hom}_{\mathbb{Z}}(R, I))$ to $\psi_\bullet(1) \in \mathrm{Hom}_{\mathbb{Z}}(M, I)$. We have the following checks.

- Well-defined: note that $\psi_{m_1+m_2}(1) = \psi_{m_1}(1) + \psi_{m_2}(1)$, so $\psi_\bullet \in \mathrm{Hom}_{\mathbb{Z}}(M, I)$.
- Homomorphic: given $\psi^1_\bullet, \psi^2_\bullet \in \mathrm{Hom}_R(M, \mathrm{Hom}_{\mathbb{Z}}(I))$, we see that each $m \in M$ has

  $$(\psi^1_\bullet + \psi^2_\bullet)(m)(1) = \psi^1_m(1) + \psi^2_m(1) = (\psi^1_\bullet(1) + \psi^2_\bullet(1))(m),$$

  where the left-hand side has addition in $\mathrm{Hom}_R(M, \mathrm{Hom}_{\mathbb{Z}}(R, I))$, and the right-hand side has addition in $\mathrm{Hom}_{\mathbb{Z}}(M, I)$.
- Injective: we show trivial kernel. Suppose that $\psi_\bullet \in \mathrm{Hom}_R(M, \mathrm{Hom}_{\mathbb{Z}}(R, I))$ has $\psi_m(1) = 0$ for each $m \in M$. Then, for each $r \in R$, the fact that ψ_m is an R-module homomorphism forces

  $$\psi_m(r) = r \cdot \psi_m(1) = r \cdot 0 = 0,$$

  so in fact $\psi_\bullet$ always return the zero map, so it is the zero object in $\mathrm{Hom}_R(M, \mathrm{Hom}_{\mathbb{Z}}(R, I))$.
- Surjective: fix $\varphi \in \mathrm{Hom}_{\mathbb{Z}}(M, I)$, and we define

  $$\psi_m(r) \coloneqq \varphi(rm).$$

  Note that this does have $\psi_m(1) = \varphi(m)$ for each $m \in M$, so $\psi_\bullet$ will go to φ, upon checking that $\psi_\bullet \in \mathrm{Hom}_R(M, \mathrm{Hom}_{\mathbb{Z}}(R, I))$. Indeed, fixing any $m \in M$ and $r_1, r_2 \in R$, we see

  $$\psi_m(r_1 + r_2) = \varphi((r_1 + r_2)m) = \varphi(r_1m + r_2m) + \varphi(r_1m) + \varphi(r_2m) = \psi_m(r_1) + \psi_m(r_2),$$

  so indeed, $\psi_m \in \mathrm{Hom}_{\mathbb{Z}}(R, I)$. Then for $r, r_1, r_2 \in R$ and $m_1, m_2 \in M$, we see

  $$\psi_{r_1m_1+r_2m_2}(r) = \varphi(rr_1m_1 + rr_2m_2) = \varphi(rr_1m_1) + \varphi(rr_2m_2) = \psi_{m_1}(rr_1) + \psi_{m_2}(rr_2).$$

  Now the key point is that the definition of the R-action on $\mathrm{Hom}_{\mathbb{Z}}(R, I)$ makes this equal to $(r_1\psi_{m_1} + r_2\psi_{m_2})(r)$, which is what we wanted.

We now show that $R^* = \mathrm{Hom}_{\mathbb{Z}}(R, I)$ is actually injective. Fix some injection $B \hookrightarrow C$ and map $\varphi_\bullet : B \to R^*$ so that we want to get a map $\overline{\varphi}_\bullet : C \to R^*$. Well, we can take $\varphi \in \mathrm{Hom}_R(B, \mathrm{Hom}_{\mathbb{Z}}(R, I))$ so that $\varphi_\bullet(1) \in \mathrm{Hom}(B, I)$ as above.

Now, the injectivity of I promises some map $\overline{\varphi} : C \to I$ such that $\overline{\varphi}(b) = \varphi_b(1)$ for each $b \in B$. Then we can take $\overline{\varphi}$ up to the map

$$\overline{\varphi}_\bullet \in \mathrm{Hom}(C, R^*)$$

satisfying $\overline{\varphi}_c(1) = \overline{\varphi}(c)$. We claim this $\overline{\varphi}_\bullet$ is the map we want. Indeed, for each $b \in B$ and $r \in R$, we see that

$$\overline{\varphi}_b(r) = (r\overline{\varphi}_b)(1) = \overline{\varphi}_{rb}(1) = \varphi_{rb}(1) = (r\varphi_b)(1) = \varphi_b(r),$$

so indeed, $\overline{\varphi}$ does extend φ. ∎

Here is another source of injective modules, if we already have some.

Proposition 3.126. Suppose $\{I_\alpha\}_{\alpha\in\lambda}$ are injective R-modules. Then $\prod_{\alpha\in\lambda} I_\alpha$ is an injective R-module.

Proof. Fix an inclusion of R-modules $B \hookrightarrow C$ with a map $\varphi : B \to \prod_\alpha I_\alpha$. Then we have the composite maps

$$\varphi_\beta : B \to \prod_{\alpha\in\lambda} I_\alpha \overset{\pi_\beta}{\twoheadrightarrow} I_\beta$$

for some fixed β. But because I_β is injective, we have the extension $\overline{\varphi_\beta} : C \to I_\beta$. So to finish, we define

$$\overline{\varphi}(c) \coloneqq (\overline{\varphi_\alpha} c)_{\alpha\in\lambda} \in \prod_{\alpha\in\lambda} I_\alpha,$$

for each $c \in C$. To see that this works, we note that any $b \in B$ has, for any $\beta \in \lambda$,

$$\pi_\beta(\overline{\varphi}(b)) = \pi_\beta\left((\overline{\varphi_\alpha} b)_{\alpha\in\lambda}\right) = \overline{\varphi_\beta}(b) = \varphi_\beta(b),$$

so we conclude that $\overline{\varphi}(b) = \varphi(b)$. This is what we wanted. ∎

So we have a reasonable supply of injective modules. Namely, we can show the following.

Proposition 3.127. Fix R a commutative ring. Then we have "enough injectives" in the category of R-modules. Explicitly, for any module M, we can find an injective R-module N for which there is an injection $M \hookrightarrow N$.

Remark 3.128. This is dual to saying that all R-modules M are projected onto from some projective module, which is much easier because all R-modules M are projected onto by some free module (e.g., $\bigoplus_{m\in M} Rm$).

Proof. We show this for abelian groups, so fix M an abelian group. The key step is that, given an $m \in M\setminus\{0\}$, we want an injective module N so that $f : M \to N$ has $f(m) \neq 0$. Take $N \coloneqq \mathbb{Q}/\mathbb{Z}$. We have two cases.

- If m has infinite order, then take $f(m)$ to be anything in $N\setminus\{0\}$, and extend this map to M by injectivity. This uses that $\mathbb{Q}/\mathbb{Z}$ is injective.
- If m has finite order n, then take $f(m) = \frac{1}{n}$, and extend this map to M by injectivity. This also uses that $\mathbb{Q}/\mathbb{Z}$ has elements of all finite orders.

Now, for each $m \in M$, we let the above map be $f_m : M \to \mathbb{Q}/\mathbb{Z}$. So we can glue these maps together to get

$$f : M \to \prod_{m \in M} \mathbb{Q}/\mathbb{Z},$$

where the map f is injective because it has trivial kernel,[7] and the product is injective because products of injective modules is injective. So we have embedded M into an injective module. ■

The proof for general rings is similar, so we will not show it here. We will say that the main obstacle is again, for each $m \in M$, finding some $\psi_\bullet \in \mathrm{Hom}_R(M, \mathrm{Hom}_{\mathbb{Z}}(R, \mathbb{Q}/\mathbb{Z}))$ such that ψ_m is nonzero.

To overcome this obstacle, the discussion above tells us how to find some $\psi \in \mathrm{Hom}_{\mathbb{Z}}(M, \mathbb{Q}/\mathbb{Z})$ with $\psi(m) \neq 0$, and the discussion in the proof of Proposition 3.124 shows us how to lift ψ into some $\psi_\bullet \in \mathrm{Hom}(M, \mathrm{Hom}_{\mathbb{Z}}(R, \mathbb{Q}/\mathbb{Z}))$ such that $\psi_m(1) = \psi(m) \neq 0$.

The injective module we made is frankly huge, but usually we can find a smaller one, and it turns out there is a "best" such injective module. Observe that the same is not true for projective modules.

Example 3.129. For $M = \mathbb{Z}/5\mathbb{Z}$, the free module $\mathbb{Z}$ can map $1 \mapsto 1$ or $1 \mapsto 2$, and neither of these appears to be "best projective module."

But here is what we have for injective modules.

Definition 3.130 (Injective evenlope)**.** The smallest injective module containing some R-module M is called the *injective envelope*.

We'll give examples in abelian groups.

Example 3.131. For $\mathbb{Z}$, the injective envelope is $\mathbb{Q}$. Note this is not finitely generated, sadly.

Example 3.132. For $\mathbb{Z}/p^n\mathbb{Z}$, the injective envelope is

$$M_p = \left\{x \in \mathbb{Q}/\mathbb{Z} : p^k x = 0 \text{ for some } k \in \mathbb{Z}\right\}$$

from earlier.

3.3.6 Modules over Euclidean Domains

We're running out of time, so let's do something else. The main thing we have to say here is the following.

Theorem 3.133. Any finitely generated module M over a principal ideal domain R is a direct sum of cyclic modules R/aR for $a \in R$.

We are not going to show this because it is somewhat technical. Instead, we provide a quick proof of the following corollary to Theorem 3.133.

Corollary 3.134. Any finitely generated module M over a Euclidean domain R is a direct sum of cyclic modules R/aR for $a \in R$.

Proof. The exact same proof as for $R = \mathbb{Z}$ will work here. Namely, all that proof required was the ability to use the division algorithm in $\mathbb{Z}$, so the proof extends to Euclidean domains R. ■

To see that we don't need the power of all principal ideal domains for applications, we have the following application of Corollary 3.134.

[7] For each $m \in M \setminus \{0\}$, we see that $f_m(m) \neq 0$ by construction, so $m \notin \ker f$.

Theorem 3.135 (Jordan normal form)**.** Fix V a finite-dimensional k-vector space, where k is algebraically closed. Then all linear transformations $T \in \operatorname{End}(V)$ are a direct sum of linear transformations which under a suitable basis look like

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

where $\alpha \in k$.

Proof. The key trick is to throw Corollary 3.134 at $R \coloneqq k[x]$. We see that R is Euclidean by using $\deg$. Now, by Corollary 3.134, we have that any $k[x]$-module named M will take the form

$$\bigoplus_{n=1}^{N} \frac{k[x]}{(p_n)},$$

where $p_\bullet \in k[x]$ for each $p_\bullet$. Because k is algebraically closed, we may take $p_\bullet = (x_\bullet - \lambda_\bullet)^{d_\bullet}$.

Now fix some $T \in \operatorname{End}(V)$. To get the desired statement, the idea is to view V itself as a $k[x]$-module, where our action is given by

$$p(x) \cdot v = p(T)v,$$

where $p(x) \in k[x]$ and $v \in V$. Essentially, we are taking the typical k-action on V and adding in a "transcendental linear operator T" to get out a $k[x]$-module. Anyways, the point is that we can write

$$V \cong \bigoplus_{n=1}^{N} \frac{k[x]}{(x-\lambda_n)^{d_n}}$$

for some $\lambda_\bullet, d_\bullet \in k$. For concreteness, we note that we can pull back each $k[x]/(x_n - \lambda_n)^{d_n}$ so that we can decompose

$$V = \bigoplus_{k=1}^{N} V_n \qquad \text{such that} \qquad V_n \cong \frac{k[x]}{(x-\lambda_n)^{d_n}}.$$

So now we see that the action of T on V will decompose nicely into the direct sum of the action of T on the V_n.

In particular, we claim that we can find a basis for which T restricted to V_n looks like

$$\begin{bmatrix} \lambda_n & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_n & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_n & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

Indeed, using $V_n \cong k[x]/(x-\lambda_n)^{d_n}$, we may pull the basis $\left\{(x-\lambda_n)^{d_n-1-e}\right\}_{e=0}^{d_n-1}$ back to $\{b_e\}_{e=0}^{d_n-1} \subseteq V_n$ so that

$$x \cdot (x-\lambda_n)^e = (x-\lambda_n)^{e+1} + \lambda_n (x-\lambda_n)^e$$

corresponds to

$$T \cdot b_e = \lambda_n b_e + b_{e-1}$$

for $0 \le e \le d_n - 1$, where b_{-1} is the pull-back of $(x-\lambda_n)^{d_n} = 0$, which is 0. These equations exactly describe the matrix we need, so we are done here. ∎

We leave with the exercise to describe the Jordan normal form over $\mathbb{R}$ by using the above proof, where we have to add in the possible irreducible quadratics.

3.4 October 14

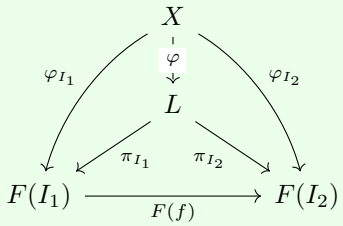
There's always more show.

3.4.1 Limits and Colimits

We're talking limits and colimits today. We have the following definitions. Here is the limit.

Definition 3.136 (Limit)**.** Fix $\mathcal{I}$ an index category and $F : \mathcal{I} \to \mathcal{C}$ a functor. Then the *limit* is an object $L \coloneqq \varprojlim_{\mathcal{I}} F(I)$ with maps $\pi_I : L \to F(I)$ for each $I \in \mathcal{I}$, which commute such that, for any $f : I_1 \to I_2$ in $\mathcal{I}$, we have $\pi_{I_2} = F(f) \circ \pi_{I_1}$.

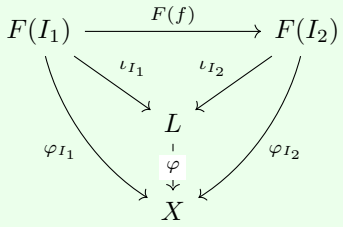
Further, L is universal with respect to this property: for any object X with maps $\varphi_I : X \to F(I)$ for each $I \in \mathcal{I}$ (which commute in the same way), then there is a unique induced map $\varphi : X \to L$ making the following diagram commute.



The dual notion of a limit is the colimit.

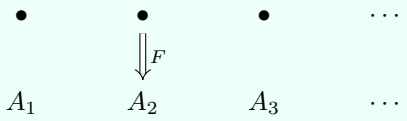
Definition 3.137 (Colimit)**.** Fix $\mathcal{I}$ an index category and $F : \mathcal{I} \to \mathcal{C}$ a functor. Then the *colimit* is an object $L \coloneqq \varinjlim_{\mathcal{I}} F(I)$ with maps $\iota_I : F(I) \to L$ for each $i \in \mathcal{I}$, which commute such that, for any $f : I_1 \to I_2$ in $\mathcal{I}$, we have $\iota_{I_2} \circ F(f) = \iota_{I_1}$.

Further, L is universal with respect to this property: for any object X with maps $\varphi_I : F(I) \to X$ for each $i \in \mathcal{I}$ (which commute in the same way), then there is a unique induced map $\varphi : L \to X$ making the following diagram commute.



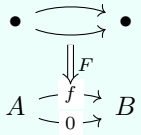
And here are the standard examples.

Example 3.138. Fix the discrete category $\mathcal{I}$ and functor $F : \mathcal{I} \to \mathcal{A}$ as follows.



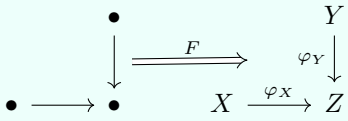
Then the limit is the direct product (the universal object projecting down into each individual), and the colimit is the direct sum (the universal objecting including each individual).

Example 3.139. A kernel of a morphism $f : A \to B$ is the limit of the following index category and functor.



Indeed, the kernel is the universal object $\operatorname{Ker} f$ with a map $\iota : \operatorname{Ker} f \to A$ such that $f \circ \iota = 0 \circ \iota = 0$. The cokernel/quotient object is the colimit of this diagram: it is the universal object $\operatorname{Coker} f$ with a map $\pi : B \to \operatorname{Coker} f$ such that $\pi \circ f = \pi \circ 0 = 0$.

Example 3.140. Pull-backs/fiber products are the limit of the following index category and functor.

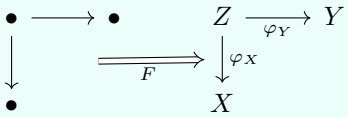


We were asked in the homework to show that, in the category of abelian groups,

$$X \times_Z Y = \{(x, y) \in X \times Y : \varphi_X x = \varphi_Y y\}.$$

This also holds in R-modules, but we will not show it explicitly; roughly speaking, $X \times_Z Y$ should consist of pairs of $X \times Y$ which are the "same" under φ_X and φ_Y.

Example 3.141. Push-outs/fiber coproducts are the colimit of the following index category and functor.



In commutative rings, this is the tensor product $X \otimes_Z Y$, where X and Y have Z-action given by $z \cdot x := \varphi_X(z)x$ and $z \cdot y := \varphi_Y(z)y$. We showed this a few days ago in the case where $Z = \mathbb{Z}$ (so that we are looking at the coproduct), and I am too lazy to do it again. (If someone wants me to, yell at me.)

3.4.2 Direct Limits

Warning 3.142. For this section, I work in slightly more generality than Borcherds did in lecture. Namely, all directed systems should be thought of as $\mathbb{N}$ under the usual ordering, and all inverse systems should be thought of as $\mathbb{N}$ under the reverse ordering.

A special example of a colimit is the "direct limit." We have the following definitions.

Definition 3.143 (Directed system)**.** Fix $\mathcal{I}$ a partially ordered set/category where every finite set has an upper bound. Then an *directed system* is a covariant functor $F : \mathcal{I} \to \mathcal{A}$ satisfying the commutativity requirements of a functor. Explicitly,

- $I \overset{f}{\preceq} I$ goes to $\mathrm{id}_{F(I)} : F(I) \to F(I)$,
- $I \overset{f}{\preceq} J \overset{g}{\preceq} K$ implies $F(g \circ f) = F(g) \circ F(f)$.

Definition 3.144 (Direct limit)**.** A *direct limit* is the colimit of a directed system.

Warning 3.145. The condition that $\mathcal{I}$ gives every finite set an upper bound will help us much later.

The most common example of such a directed system we might run into is $\mathbb{N}$ under the usual ordering, which is the category presented as follows.

$$1 \to 2 \to 3 \to \cdots.$$

Intuitively, if these maps are injective after applying the functor, then we are doing a kind of union along a chain of objects. If they aren't injective, we have to be more careful.

We have already seen an example before.

Exercise 3.146. We have that M_p is the direct limit of

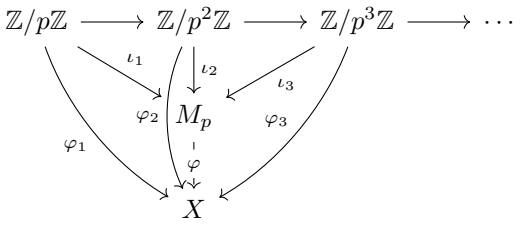
$$\mathbb{Z}/p\mathbb{Z} \hookrightarrow \mathbb{Z}/p^2\mathbb{Z} \hookrightarrow \mathbb{Z}/p^3\mathbb{Z} \hookrightarrow \cdots$$

Proof. Indeed, our inclusions are given by $\mathbb{Z}/p^k\mathbb{Z} \cong \frac{1}{p^k}\mathbb{Z}/\mathbb{Z} \hookrightarrow M_p$, or in other words, $1 \mapsto \frac{1}{p^k}$. We can check these commute: we need to check the map $f_{k\ell} : \mathbb{Z}/p^k\mathbb{Z} \to \mathbb{Z}/p^\ell\mathbb{Z}$ (for $k \le \ell$) does indeed satisfy $\iota_k = \iota_\ell \circ f_{k\ell}$, which is simply

$$\iota_k(n) = \frac{n}{p^k} = \frac{np^{\ell-k}}{p^\ell} = \iota_\ell\left(np^{\ell-k}\right) = \iota_\ell(f_{k\ell}n),$$

as needed.

We now show the universal property. Fix an object X with maps $\varphi_k : \mathbb{Z}/p^k\mathbb{Z} \to X$ such that $\varphi_k = \varphi_\ell \circ f_{k\ell}$. We need to exhibit a unique induced map $\varphi : M_p \to X$ making the following diagram commute.



We show uniqueness and existence one at a time.

- We show that φ is unique. Indeed, fix any $\frac{n}{p^k} \in M_p$. Then we have $\frac{n}{p^k} = \iota_k(n)$, so if the diagram is to commute, we must have

  $$\varphi\left(\frac{n}{p^k}\right) = (\varphi \circ \iota_k)(n) = \varphi_k(n),$$

  so φ is indeed forced.

- We now show that φ exists. As we worked out, we need to define φ by

  $$\varphi\left(\frac{n}{p^k}\right) \coloneqq \varphi_k(n).$$

  We see φ is well-defined as a function because, even though we might $\frac{n}{p^k} = \frac{np^{\ell-k}}{p^\ell}$ (where $k \le \ell$ without loss of generality), it is still true that

  $$\varphi_k(n) = \varphi_\ell\left(np^{\ell-k}\right) = \left(\varphi_\ell \circ f_{k\ell}\right)(n),$$

  where $\varphi_k = \varphi_\ell \circ f_{k\ell}$ by hypothesis on the $\varphi_\bullet$.

We technically have to check that φ is a group homomorphism. Well,

$$\varphi\left(\frac{n}{p^k}+\frac{m}{p^\ell}\right)=\varphi\left(\frac{np^\ell+mp^k}{p^{k+\ell}}\right)=\varphi_{k+\ell}\left(np^\ell+mp^k\right),$$

but now $\varphi_{k+\ell}$ is a group homomorphism, so this reads

$$\varphi_{k+\ell}\left(np^\ell\right)+\varphi_{k+\ell}\left(mp^k\right)=\varphi_k(n)+\varphi_\ell(m)=\varphi\left(\frac{n}{p^k}\right)+\varphi\left(\frac{m}{p^\ell}\right),$$

which is what we wanted. ■

3.4.3 Inverse Limits

If we wanted to compute the dual as $\mathrm{Hom}_{\mathbb{Z}}(M_p,\mathbb{Q}/\mathbb{Z})$, we see that $\mathrm{Hom}_{\mathbb{Z}}\left(\mathbb{Z}/p^n\mathbb{Z},\mathbb{Q}/\mathbb{Z}\right)\cong\mathbb{Z}/p^n\mathbb{Z}$ (tracking where 1 goes shows $\mathrm{Hom}\left(\mathbb{Z}/p^n\mathbb{Z},\mathbb{Q}/\mathbb{Z}\right)\cong\frac{1}{p^n}\mathbb{Z}$), but now the arrows are reversed, so we end up with the following system.

$$\mathbb{Z}/p\mathbb{Z}\longleftarrow\mathbb{Z}/p^2\mathbb{Z}\longleftarrow\mathbb{Z}/p^3\mathbb{Z}\longleftarrow\cdots$$

Explicitly, our map $g_{\ell k}:\mathbb{Z}/p^\ell\mathbb{Z}\to\mathbb{Z}/p^k\mathbb{Z}$ for $k\le\ell$ is really referring to the map $\mathrm{Hom}_{\mathbb{Z}}\left(\mathbb{Z}/p^\ell\mathbb{Z},\mathbb{Q}/\mathbb{Z}\right)\to\mathrm{Hom}_{\mathbb{Z}}\left(\mathbb{Z}/p^k\mathbb{Z},\mathbb{Q}/\mathbb{Z}\right)$ induced by $-\circ f_{k\ell}$. So we see

$$\left(1\mapsto\frac{n}{p^\ell}\right)\longmapsto\left(1\overset{f_{k\ell}}{\mapsto}p^{\ell-k}\mapsto\frac{n}{p^k}\right).$$

Thus, our map $g_{\ell k}:\mathbb{Z}/p^\ell\mathbb{Z}\to\mathbb{Z}/p^k\mathbb{Z}$ is just the projection $n\mapsto n$. We can check the commutativity laws $g_{\ell k}\circ g_{m\ell}=g_{mk}$ for $k\le\ell\le m$ because both sides are $n\mapsto n$

Now, when taking the limit (!), we are roughly asking for a "compatible" system of elements from each of the $\mathbb{Z}/p^\bullet\mathbb{Z}$. This sort of limit is called an "inverse limit." We have the following definitions.

Definition 3.147 (Invese system)**.** Fix $\mathcal{I}$ a partially ordered set/category where every finite set has an upper bound. Then an *inverse system* is a contravariant functor $F:\mathcal{I}\to\mathcal{A}$ satisfying the commutativity requirements of a functor. Explicitly,

- $I\overset{f}{\preceq}I$ goes to $\mathrm{id}_{F(I)}:F(I)\to F(I)$,
- $I\overset{f}{\preceq}J\overset{g}{\preceq}K$ implies $F(g\circ f)=F(f)\circ F(g)$.

Definition 3.148 (Inverse limit)**.** An *inverse limit* is the limit of an inverse system.

And now let's work out our example because I should do this at least once in my life.

Exercise 3.149. The p-adic integers $\mathbb{Z}_p$ is the inverse limit of the following diagram.

$$\mathbb{Z}/p\mathbb{Z}\longleftarrow\mathbb{Z}/p^2\mathbb{Z}\longleftarrow\mathbb{Z}/p^3\mathbb{Z}\longleftarrow\cdots$$

Proof. This is essentially the definition of $\mathbb{Z}_p$. Because I should say something here, I will show the following to describe $\mathbb{Z}_p$.

Lemma 3.150. Fix $\mathcal{I}$ an index category, and fix $\mathcal{C}$ any of the category of sets, groups, rings, or modules. Then, for any functor $F:\mathcal{I}\to\mathcal{C}$, we can write

$$\varprojlim_{\mathcal{I}}F(I)\cong\left\{(a_I)_{I\in\mathcal{I}}\in\prod_{I\in\mathcal{I}}F(I):F(f)(a_I)=a_J\text{ for each }f:I\to J\right\}.$$

Proof. We will in all of those categories at once as much as possible.[8] For brevity, let L be the given construction. If we are in the category of sets, L is allowed to empty; in the other categories, L is nonempty because it contains the identity. We can also check that the condition

$$F(f)(a_I) = a_J$$

for each $f : I \to J$ in $\mathcal{I}$ preserves group operation, ring multiplication, and linear combination, so L is closed under each of these under the respective categories. Then because

$$L \subseteq \prod_{I \in \mathcal{I}} F(I)$$

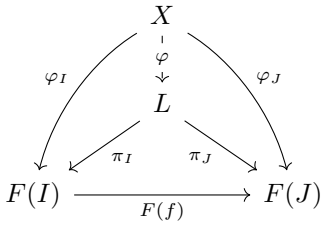
as constructed, we really only needed to test L as a subobject.

Continuing, we have projection maps $\pi_J : L \to F(J)$ for each $J \in \mathcal{I}$ by taking $(a_I)_{I\in\mathcal{I}}$ to a_J. This map preserves the (pointwise) operations on L, so it is a morphism in any of the given categories. We see that π_J commute as needed because, for any $f : J \to K$, we have

$$(F(f) \circ \pi_J)\left((a_I)_{I\in\mathcal{I}}\right) = F(f)(a_J) = a_K = \pi_K\left((a_I)_{I\in\mathcal{I}}\right),$$

for any $(a_I)_{I\in\mathcal{I}} \in L$, by hypothesis on the $(a_I)_{I\in\mathcal{I}}$.

It remains to show the universal property. Fix X any object with maps $\varphi_I : X \to F(I)$ for each $I \in \mathcal{I}$ such that $F(f) \circ \varphi_I = \varphi_J$ for each $f : I \to J$. Then we need to induce a unique map $\varphi : X \to L$ making the following diagram commute.



As usual, we show uniqueness and existence of φ one at a time.

- We show that φ is unique. Indeed, for any $x \in X$, if $\varphi(x) = (a_I)_{I\in\mathcal{I}}$, then the commutativity of the diagram forces
$$a_I = \pi_I(\varphi(x)) = \varphi_I(x)$$
for each $I \in \mathcal{I}$, so we are forced to have $\varphi(x) = (\varphi_I x)_{i\in\mathcal{I}}$.
- We show that φ exists. As above, we are forced to define
$$\varphi(x) \coloneqq (\varphi_I x)_{I\in\mathcal{I}}$$
for each $x \in X$. This is indeed an element of L because, for each $f : I \to J$, we see $F(f)(\varphi_I x) = \varphi_J x$ by assumption on the $\varphi_\bullet$.

  Technically, we do have to show that φ is also a morphism. Well, we note that φ is actually the induced map
$$X \to \prod_{I\in\mathcal{I}} F(I)$$
where we have restricted the output to live in L. So because the product exists as constructed in each of the given categories, we see $X \to L$ is a morphism. ∎

The point is that we can realize $\mathbb{Z}_p$ as

$$\varprojlim \mathbb{Z}/p^\bullet\mathbb{Z} \cong \left\{ (a_k)_{k\geq 1} \in \prod_{k=1}^{\infty} \mathbb{Z}/p^k\mathbb{Z} : a_\ell \equiv a_k \pmod{p^k} \text{ for each } \ell \geq k \right\}$$

using the above construction. In words, $\mathbb{Z}_p$ consists of infinite sequences of elements of $\mathbb{Z}?p^\bullet\mathbb{Z}$ where the elements are "compatible" with each other. ∎

[8] Regardless, I will be somewhat vague in the checks that functions are morphisms because I don't want to each check four times.

3.4.4 Duals of Direct Limits

Our story of $\mathbb{Z}_p$ was about dualizing the diagram for M_p, so it is a reasonable to hope that the dual of M_p is $\mathbb{Z}_p$. This is indeed true.

Exercise 3.151. We have that $\operatorname{Hom}(M_p, \mathbb{Q}/\mathbb{Z}) \cong \mathbb{Z}_p$.

Proof. We use the explicit construction of $\mathbb{Z}_p$ given by Lemma 3.150. We map $\varphi : \mathbb{Z}_p \to \operatorname{Hom}(M_p, \mathbb{Q}/\mathbb{Z})$ explicitly by taking $(a_k)_{k\ge1} \in \mathbb{Z}_p$ to the map

$$\varphi\left((a_k)_{k\ge1}\right) : \frac{n}{p^k} \mapsto \frac{na_k}{p^k}.$$

The rest of the proof is book-keeping; we check that φ is indeed an isomorphism.

- This map is well-defined because, even though we might have $\frac{n}{p^k} = \frac{m}{p^\ell}$ where $k \le \ell$ without loss of generality, we see that $np^{\ell-k} \equiv m \pmod{p^\ell}$, which implies that
$$a_k np^{\ell-k} \equiv a_\ell m \pmod{p^\ell}$$
because $a_k \equiv a_\ell \pmod{p^k}$. So indeed, $\frac{na_k}{p^k} \equiv \frac{a_\ell m}{p^\ell} \pmod 1$.
- We see $\varphi\left((a_k)_{k\ge1}\right)$ is indeed a homomorphism. Indeed, given $\frac{n}{p^k}, \frac{m}{p^\ell} \in M_p$, we have
$$\varphi(a)\left(\frac{n}{p^k} + \frac{m}{p^\ell}\right) = \varphi(a)\left(\frac{np^\ell + mp^k}{p^{k+\ell}}\right) = \frac{a_{k+\ell}np^\ell}{p^{k+\ell}} + \frac{a_{k+\ell}mp^k}{p^{k+\ell}} = \varphi\left(\frac{np^\ell}{p^{k+\ell}}\right) + \varphi\left(\frac{mp^k}{p^{k+\ell}}\right),$$
which collapses to what we want.
- We see φ is itself a homomorphism. Fix $a, b \in \mathbb{Z}_p$. Then for any $\frac{n}{p^k} \in M_p$, we have
$$\varphi(a+b)\left(\frac{n}{p^k}\right) = \frac{(a_k + b_k)n}{p^k} = \frac{a_k n}{p^k} + \frac{b_k n}{p^k} = \left(\varphi(a) + \varphi(b)\right)\left(\frac{n}{p^k}\right).$$
- We see φ is injective. Indeed, it suffices to show that φ has trivial kernel. So suppose $a \in \mathbb{Z}_p$ has $\varphi(a)$ the zero map. Well, for any $k \ge 1$, we see
$$\frac{a_k}{p^k} = \varphi(a)\left(\frac{1}{p^k}\right) = 0,$$
so $a_k \equiv 0 \pmod{p^k}$. So indeed, a is the zero element.
- We show φ is surjective. Fix $f \in \operatorname{Hom}(M_p, \mathbb{Q}/\mathbb{Z})$ some homomorphism. Then, for any $k \ge 1$, we note $f\left(1/p^k\right) = a_k/p^k$ for some $a_k \in \mathbb{Z}$ because $p^k \cdot f\left(1/p^k\right) = f(1) = f(0) = 0$. We claim that
$$a \coloneqq (a_k)_{k\ge1} \in \mathbb{Z}_p.$$
Indeed, for any $k \le \ell$, we see that
$$\frac{a_k}{p^k} = f\left(\frac{1}{p^k}\right) = f\left(\frac{p^{\ell-k}}{p^\ell}\right) = \frac{a_\ell p^{\ell-k}}{p^\ell} = \frac{a_\ell}{p^k},$$
so $a_k \equiv a_\ell \pmod{p^k}$.
So we claim that $f = \varphi(a)$. Indeed, for any $\frac{n}{p^k} \in M_p$, we see that
$$f\left(\frac{n}{p^k}\right) = n \cdot f\left(\frac{1}{p^k}\right) = n \cdot \frac{a_k}{p^k} = \frac{a_k n}{p^k} = \varphi(a)\left(\frac{n}{p^k}\right),$$
which is what we needed. ∎

Putting everything together, we saw that

$$\mathrm{Hom}\left(\varinjlim \mathbb{Z}/p^\bullet\mathbb{Z}, \mathbb{Q}/\mathbb{Z}\right) \cong \varprojlim \mathrm{Hom}\left(\mathbb{Z}/p^\bullet\mathbb{Z}, \mathbb{Q}/\mathbb{Z}\right).$$

In fact, this holds more generally.

Proposition 3.152. Fix $\mathcal{I}$ an index category and $F : \mathcal{I} \to \mathcal{C}$ a functor, where $\mathcal{C}$ is the category of sets, groups, rings, or modules. Then, for any object $X \in \mathcal{C}$,

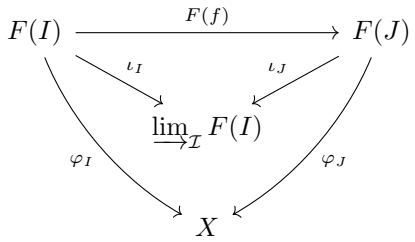
$$\mathrm{Hom}_R\left(\varinjlim_{\mathcal{I}} F(I), X\right) \cong \varprojlim_{\mathcal{I}} \mathrm{Hom}_R\left(F(I), X\right).$$

Proof. We essentially imitate the example. We use Lemma 3.150, which tells us that

$$\varprojlim_{\mathcal{I}} \mathrm{Hom}_R\left(F(I), X\right) \cong \left\{(\varphi_I) \in \prod_{I\in\mathcal{I}} \mathrm{Hom}_R(F(I), X) : \varphi_I = \varphi_J \circ F(f) \text{ for each } f : I \to J\right\} =: L,$$

where the commutativity laws come from the fact that $f : I \to J$ induces the morphism $\mathrm{Hom}(F(J), X) \to \mathrm{Hom}(F(I), X)$ by $- \circ F(f)$, so the condition $F(f)(\varphi_J) = \varphi_I$ (where $F(f)$ here is heavy abuse of notation) reads as $\varphi_J \circ F(f) = \varphi_I$.

Staring harder, we see L is tuples of maps $\varphi_I : F(I) \to X$ which make the following diagram commute. (Here, we are naming the maps of $\varinjlim_{\mathcal{I}} F(I)$ by $\iota_I : F(I) \to \varinjlim_{\mathcal{I}} F(I)$.)



But by the universal property of $\varinjlim_{\mathcal{I}} F(I)$, we can take tuples (φ_I) which commute with the $F(f)$ to unique morphisms $\varphi : \varinjlim_{\mathcal{I}} F(I) \to X$ such that $\varphi_I = \varphi \circ \iota_I pi$. Call this map

$$\psi : L \to \mathrm{Hom}_R\left(\varinjlim_{\mathcal{I}} F(I), X\right).$$

We can check by hand that ψ is an R-module homomorphism. This is more or less book-keeping.

- We see that ψ is well-defined because the morphism $\varinjlim_{\mathcal{I}} F(I) \to X$ induced by the universal property is unique.
- We show that ψ is an R-module homomorphism. Indeed, fix $r_1, r_2 \in R$ and $(\varphi_I^1), (\varphi_I^2) \in L$ so that $\varphi^1 := \psi\left((\varphi_I^1)\right)$ and $\varphi^2 := \psi\left((\varphi_I^2)\right)$. Then we see that $\varphi := r_1\varphi^1 + r_2\varphi^2$ satisfies

  $$(\varphi \circ \iota_I)(a_I) = r_1\varphi^1(\iota_I a_I) + r_2\varphi^2(\iota_I a_I) = (r_1\varphi_I^1 + r_2\varphi_I^2)(a_I),$$

  for any $a_I \in F(I)$. The point of this computation is that $r_1\varphi^1 + r_2\varphi^2$ commutes with the same diagram that $\psi\left((r_1\varphi_I^1 + r_2\varphi_I^2)_I\right)$ commutes with, so they are equal by uniqueness.
- We show that ψ is injective, for which it suffices to check that ψ has trivial kernel. Well, suppose $\psi\left((\varphi_I)\right) = 0$. Then, by the commuting in the universal property, we see that $\varphi_I = 0 \circ \iota_I = 0$ for each $I \in \mathcal{I}$, so indeed, (φ_I) is the zero element.

- We show that ψ is surjective. Indeed, given a morphism $\varphi : \varinjlim_{\mathcal{I}} F(I) \to X$, we set $\varphi_I \coloneqq \varphi \circ \iota_I$, which has $\varphi_I : F(I) \to X$. We can check that $(\varphi_I) \in L$ because, for each $f : I \to J$,

$$\varphi_J \circ F(f) = \varphi \circ \iota_J \circ F(f) = \varphi \circ \iota_I = \varphi_I$$

by hypothesis on the $\iota_\bullet$.

And to finish, we see that $\psi((\varphi_I)) = \varphi$ because $\varphi \circ \iota_I = \varphi_I$ by construction. ∎

As a remark, the dual of Proposition 3.152 is not generally true for size reasons: limits are big, and hom-sets tend to be bigger, so

$$\operatorname{Hom}_R\left(X, \varprojlim_{\mathcal{I}} F(I)\right)$$

is frankly huge. However, the direct limit tends to have a topology, and when this is taken to account, things tend to be better behaved. But this is more analysis than algebra.

3.4.5 Profinite Groups

Let's have another example.

Example 3.153. We have the "profinite" completion $\widehat{\mathbb{Z}}$ of $\mathbb{Z}$ by the inverse limit of the system $\mathbb{Z}/n\mathbb{Z}$, where we have maps $\mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ by projection whenever $m \mid n$. Of course, we do have

$$\mathbb{Z} \hookrightarrow \varprojlim \mathbb{Z}/n\mathbb{Z},$$

and in fact this is a compact ring because it is the product of lots of compact $\mathbb{Z}_p$ rings.

Explicitly, we can show the following.

Exercise 3.154. We have that

$$\widehat{\mathbb{Z}} \cong \prod_p \mathbb{Z}_p,$$

where the product is taken over primes p.

Proof. There is some Yoneda stuff that we can do because both are inverse limits[9], but we can just exhibit the isomorphism by hand.

Indeed, we use Lemma 3.150 to write

$$\widehat{\mathbb{Z}} \cong \left\{(a_n) \in \prod_{n\in\mathbb{N}} \mathbb{Z}/n\mathbb{Z} : a_n \equiv a_m \pmod m \text{ for } m \mid n\right\}.$$

In particular, we note that we have a map $\varphi_p : \widehat{\mathbb{Z}} \to \mathbb{Z}_p$ for each prime p by taking

$$\varphi_p((a_n)) \coloneqq \left(a_{p^k}\right)_{k\geq 1} \in \mathbb{Z}_p.$$

We do indeed get out an element of $\mathbb{Z}_p$ because, for any $k \geq \ell$, we need $a_{p^\ell} \equiv a_{p^k} \pmod{p^k}$, which is true because $p^k \mid p^\ell$. We can also check that this is a group homomorphism: given $(a_n) \in \widehat{\mathbb{Z}}$ and $b_n \in \widehat{\mathbb{Z}}$, we see that

$$\varphi_p((a_n) + (b_n)) = \varphi_p((a_n + b_n)) = \left(a_{p^k} + b_{p^k}\right)_{k\geq 1} = \left(a_{p^k}\right)_{k\geq 1} + \left(b_{p^k}\right)_{k\geq 1} = \varphi_p((a_n)) + \varphi_p(n).$$

[9] Maps into $\prod_p \mathbb{Z}_p$ are essentially maps into each of the $\mathbb{Z}/p^k\mathbb{Z}$ for each prime power p^k which commute at each prime. These maps can be uniquely assembled into a map into each $\mathbb{Z}/n\mathbb{Z}$ for each $n \in \mathbb{N}$, which are in bijection with maps into $\widehat{\mathbb{Z}}$.

Anyways, the morphisms $\varphi_p : \widehat{\mathbb{Z}} \to \mathbb{Z}_p$ for each p can be used to assemble a morphism

$$\varphi : \widehat{\mathbb{Z}} \to \prod_p \mathbb{Z}_p$$

by the universal property of the product. We claim that φ is an isomorphism.

- We show that φ is injective. We already know that φ is a morphism, so it suffices to show that φ has trivial kernel. Well, fix some $a = (a_n)_{n\in\mathbb{N}}$ which goes to 0 under φ. Then, for any $n \in \mathbb{N}$, we show $a_n = 0$, which will be enough to conclude $(a_n) = 0$. Indeed, fix the prime factorization

  $$n = \prod_{p|n} p^{\alpha_p}.$$

  Then $p^{\alpha_p} \mid n$ for each $p \mid n$, so

  $$a_n \equiv a_{p^{\alpha_p}} \pmod{p^{\alpha_p}},$$

  but $a_{p_p^\alpha} = \varphi_p(a) = 0$ because $a \in \ker\varphi$, so we see

  $$a_n \equiv 0 \pmod{p^{\alpha_p}}$$

  for each $p \mid n$. Using the Chinese remainder theorem to assemble this (finite) system of congruences, we see that

  $$a_n \equiv 0 \pmod{n},$$

  which is exactly what we wanted.

- We show that φ is surjective. Fix some tuple

  $$a = (a_p)_p = ((a_k)_{k\geq 1})_p \in \prod_p \mathbb{Z}_p$$

  that we want to hit by φ. Well, for each $n \in \mathbb{N}$, we note that we have the prime factorization

  $$n = \prod_p p^{\nu_p(n)},$$

  so we conjure a_n by using the Chinese remainder theorem to solve the (finite) system of congruences

  $$a_n \coloneqq (a_{\nu_p(n)})_p$$

  for each prime p. (This system is finite because we can ignore all the primes p where $\nu_p(n) = 0$, and only finitely many primes divide p.) We check that $(a_n)_{n\in\mathbb{N}}$ is a well-defined element of $\widehat{\mathbb{Z}}$: if $m \mid n$, then $\nu_p(m) \leq \nu_p(n)$ for each prime p, so

  $$a_n = (a_{\nu_p(n)})_p \equiv (a_{\nu_p(n)})_p = a_m \pmod{p^{\nu_p(m)}}$$

  because $(a_k)_p \in \mathbb{Z}_p$. So the Chinese remainder theorem now promises that $a_n \equiv a_m \pmod{m}$, as needed.

  It remains to check that $\varphi\left((a_n)_{n\in\mathbb{N}}\right) = a$. Well, by construction, we see that

  $$\varphi\left((a_n)_{n\in\mathbb{N}}\right) = \left(\left(a_{p^k}\right)_{k\geq 1}\right)_p = ((a_k)_{k\geq 1})_p = a,$$

  as needed. This finishes. ∎

Anyways, here is the definition of profinite.

Definition 3.155 (Profinite)**.** A *profinite group* is a group which is the inverse limit of some finite groups.

Remark 3.156. "Profinite" is short of "projective limit of finite things." These are compact, topologically, which is nice. (In fact, we can define profinite groups topologically as Hausdorff, compact, totally disconnected topological groups.)

As a warning, sometimes taking the profinite completion just gives 0, which is indeed compact, though not very useful.

Number theorists tend to like $\widehat{\mathbb{Z}}$ because it almost looks like the ring of adeles.

Example 3.157. The ring of adeles is $(\mathbb{R}\times\widehat{\mathbb{Z}})\otimes_{\mathbb{Z}}\mathbb{Q}$, so we do have some reason to care about the profinite completion. Roughly speaking, this is because we are really looking at

$$\left(\mathbb{R}\times\prod_p \mathbb{Z}_p\right)\otimes_{\mathbb{Z}}\mathbb{Q},$$

and a specific tensor with $\mathbb{Q}$ can only "introduce" finitely primes into the denominator. At a high level, the finite places are coming from the $\mathbb{Z}_p$, and the infinite places are coming from $\mathbb{R}$.

3.4.6 Colimits and Exactness

Let's see if limits and colimits preserve exactness. Namely, fix our index category $\mathcal{I}$ with functors $A_\bullet, B_\bullet, C_\bullet : \mathcal{I}\to\mathcal{C}$, with prescribed exact sequences

$$0\to A_I \xrightarrow{\iota_I} B_I \xrightarrow{\pi_I} C_I\to 0$$

for each $I\in\mathcal{I}$. We will also require each square induced by $f: I\to J$

$$\begin{array}{ccc} I & & A_I & \xrightarrow{\iota_I} & B_I & \xrightarrow{\pi_I} & C_I \\ {\scriptstyle f}\downarrow & & {\scriptstyle A_f}\downarrow & & {\scriptstyle B_f}\downarrow & & {\scriptstyle C_f}\downarrow \\ J & & A_J & \xrightarrow{\iota_J} & B_J & \xrightarrow{\pi_J} & C_J \end{array}$$

to commute. Then we can ask if

$$0\to \varprojlim_{\mathcal{I}} A_I \to \varprojlim_{\mathcal{I}} B_I \to \varprojlim_{\mathcal{I}} C_I \to 0$$

is exact, as well as the same question for $\varinjlim$. To be explicit, the composite maps

$$\varprojlim_{\mathcal{I}} A_I \to A_I \xrightarrow{\iota_I} B_I$$

induce a (unique commuting) map $\varprojlim_{\mathcal{I}} A_I \to \varprojlim_{\mathcal{I}} B_I$ because we can see that $\varprojlim_{\mathcal{I}} \to B_I \to B_J$ and $\varprojlim_{\mathcal{I}} A_I \to B_J$ are the same by the commutativity hypothesis on the $\iota_\bullet$. In other words, the following diagram commutes.

$$\begin{array}{ccc} I & & \varprojlim_{\mathcal{I}} A_I & \longrightarrow & A_I & \xrightarrow{\iota_I} & B_I \\ {\scriptstyle f}\downarrow & & & \searrow & {\scriptstyle A_f}\downarrow & & {\scriptstyle B_f}\downarrow \\ J & & & & A_J & \xrightarrow{\iota_J} & B_J \end{array}$$

Then we can induce $\varprojlim_{\mathcal{I}} B_I \to \varprojlim_{\mathcal{I}} C_I$ in the same way.

Similarly, the composite maps

$$A_I \to B_I \to \varinjlim_{\mathcal{I}} B_I$$

induce a (unique commuting) map $\varinjlim_{\mathcal{I}} A_I \to \varinjlim_{\mathcal{I}} B_I$, and we can induce $\varinjlim_{\mathcal{I}} B_I \to \varinjlim_{\mathcal{I}} C_I$ in the same way.

As a more concrete example, taking $\mathcal{I}$ to be a category with no morphisms, $\varprojlim$ is asking if products preserve exactness, and $\varinjlim$ is asking if coproducts preserve exactness.

Example 3.158. Taking direct sums if two short exact sequences

$$0 \to A_1 \to B_1 \to C_1 \to 0$$

and

$$0 \to A_2 \to B_2 \to C_2 \to 0$$

we get

$$0 \to A_1 \oplus A_2 \to B_1 \oplus B_2 \to C_1 \oplus C_2 \to 0,$$

which is still exact.

In general, however, the best we can say is that limits will preserve left exactness, giving the exact sequence

$$0 \to \varprojlim A_\alpha \to \varprojlim B_\alpha \to \varprojlim C_\alpha,$$

and colimits will preserve right exactness, giving the short exactness

$$\varinjlim A_\alpha \to \varinjlim B_\alpha \to \varinjlim C_\alpha \to 0.$$

Essentially this is because limits preserve kernels, which is equivalent to the left exactness; more generally, limits preserve limits. (And dually, on the other hand, colimits will preserve colimits.)

Proposition 3.159. Limits preserve limits. Namely, if we have index categories $\mathcal{I}$ and $\mathcal{J}$ with the system $\{F(I,J)\}_{i\in I, j\in J}$, then

$$\varprojlim_I \varprojlim_J F(I,J) \cong \varprojlim_J \varprojlim_I F(I,J).$$

Proof. The idea is to show that both sides are

$$\varprojlim_{I\times J} F(I,J).$$

We leave the details as an exercise. ∎

By considering duals, we have the following.

Corollary 3.160. Colimits preserve colimits.

Proof. Push everything into the opposite category, where the statement is that limits preserve limits. ∎

In particular, because cokernels are colimits, we find that colimits preserve cokernels, so colimits are right exact.

And here is the corresponding counterexample: colimits do not always preserve left exactness.

Exercise 3.161. Colimits do not always preserve kernels.

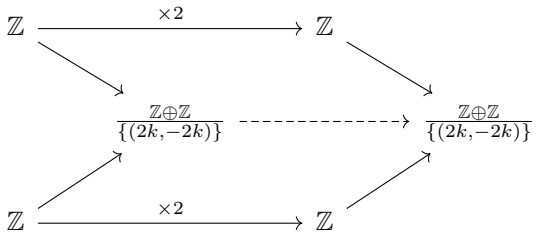
Proof. Pick up our favorite counterexample

$$0 \to \mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0,$$

and because Professor Borcherds is in an incredibly unimaginative mood this morning, we simply take the fiber products of this sequence three times.

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \mathbb{Z} & \xrightarrow{\times 2} & \mathbb{Z} & \longrightarrow & \mathbb{Z}/2\mathbb{Z} & \longrightarrow & 0 \\
 & & \uparrow{\scriptstyle \times 2} & & \uparrow{\scriptstyle \times 2} & & \uparrow{\scriptstyle \times 2} & & \\
0 & \longrightarrow & \mathbb{Z} & \xrightarrow{\times 2} & \mathbb{Z} & \longrightarrow & \mathbb{Z}/2\mathbb{Z} & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \times 2} & & \downarrow{\scriptstyle \times 2} & & \downarrow{\scriptstyle \times 2} & & \\
0 & \longrightarrow & \mathbb{Z} & \xrightarrow{\times 2} & \mathbb{Z} & \longrightarrow & \mathbb{Z}/2\mathbb{Z} & \longrightarrow & 0
\end{array}$$

We really only have to pay attention to the map on the left. Indeed, we stare at the following diagram.



The induced map here will follow the $\times 2$ through, so our map is

$$\frac{\mathbb{Z}\oplus\mathbb{Z}}{\{(2k,-2k)\}} \xrightarrow{(\times 2,\times 2)} \frac{\mathbb{Z}\oplus\mathbb{Z}}{\{(2k,-2k)\}}.$$

We can check that this is not surjective because any representative in the output will vanish under projecting into $(\mathbb{Z}/2\mathbb{Z})^2$, but the projection $\frac{\mathbb{Z}\oplus\mathbb{Z}}{\{(2k,-2k)\}} \to \frac{\mathbb{Z}\oplus\mathbb{Z}}{2\mathbb{Z}\oplus 2\mathbb{Z}}$ still sends $(1,1)$ somewhere nontrivial. ■

Remark 3.162. In fact, we have that

$$\frac{\mathbb{Z}\oplus\mathbb{Z}}{\{(2k,-2k)\}} \cong \mathbb{Z}\oplus\mathbb{Z}/2\mathbb{Z},$$

say by taking $(x,y) \mapsto (x+y, y \pmod 2)$. We won't explicitly check that this is an isomorphism, but it can be checked.

Anyways, it turns out that many cases do have colimits preserving kernels. Namely, in the case of direct limits they do because we added the condition that every finite set has an upper bound (!).

Example 3.163. The integers with the usual ordering is a "directed set": any finite set does indeed have an upper bound by taking the maximum. More generally, any totally ordered set is directed.

Anyways, we have the following.

Proposition 3.164. Colimits over directed sets do preserve exactness.

Proof. We show this for index category given by the partially ordered set $\mathbb{N}$, for ease of notation. Namely,

we have a list of short exact sequences as follows.

$$\begin{array}{ccccccccc}
 & & \vdots & & \vdots & & \vdots & & \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \longrightarrow & A_2 & \longrightarrow & B_2 & \longrightarrow & C_2 & \longrightarrow & 0 \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \longrightarrow & A_1 & \longrightarrow & B_1 & \longrightarrow & C_1 & \longrightarrow & 0 \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \longrightarrow & A_0 & \longrightarrow & B_0 & \longrightarrow & C_0 & \longrightarrow & 0
\end{array}$$

Recall that because colimits preserve colimits and hence cokernels (i.e., quotients), we know that

$$\varinjlim_{k\in\mathbb{N}} A_k \to \varinjlim_{k\in\mathbb{N}} B_k \to \varinjlim_{k\in\mathbb{N}} C_k \to 0$$

is exact. We want to get injectivity of the map $\varinjlim A_k \to \varinjlim B_k$. Well, pick $a \in \varinjlim A_k$ which is in the kernel of this map, and because the set is directed, we may choose a particular $a_k \in A_k$ representing A. Essentially, what is happening here is that

$$\varinjlim A_k = \bigcup_{k\in\mathbb{N}} A_k \Big/ \text{some equivalence relation.}$$

But now $a_k \to 0$ for some B_ℓ where $\ell \geq k$, so $a = 0$ by exactness of the original sequence. ■

Importantly, the above argument fails for the diagram

$$\begin{array}{ccc}
\bullet & \longrightarrow & \bullet \\
\downarrow & & \\
\bullet & &
\end{array}$$

because elements of the fiber product do not need to come from either A or B, for they might come from a pair of both.

Remark 3.165. Colimits do still preserve exactness over "filtered" categories, which are categories $\mathcal{C}$ for which any objects A and B have a third object C with $A \to C$ and $B \to C$, as well as the condition that any time we have two maps $A \to B$, there is a map $B \to C$ for which the composites are equal.

We also have the following.

Proposition 3.166. Colimits preserve exactness over discrete categories.

Proof. Colimits over discrete categories are the direct sum, so we are saying that a set of short exact sequences

$$0 \to A_\alpha \xrightarrow{\iota_\alpha} B_\alpha \xrightarrow{\pi_\alpha} C_\alpha \to 0$$

for $\alpha \in \lambda$ induces a short exact sequence

$$0 \to \bigoplus_{\alpha\in\lambda} A_\alpha \xrightarrow{\iota} \bigoplus_{\alpha\in\lambda} B_\alpha \xrightarrow{\pi} \bigoplus_{\alpha\in\lambda} C_\alpha \to 0.$$

To show this, we again note that we are only interested in the injectivity of left-hand map. Well, suppose that an element $(a_\alpha)_{\alpha\in\lambda} \in \bigoplus_\alpha A_\alpha$ is in the kernel; tracking through the inclusions, we see that we must have

$$\pi\left((a_\alpha)_{\alpha\in\lambda}\right) = \left(\pi_\alpha a_\alpha\right)_{\alpha\in\lambda},$$

so being in the kernel forces that $a_\alpha \in \ker \pi_\alpha$ for each $\alpha \in \lambda$. But by exactness, we must have $a_\alpha = 0$ for each $\alpha \in \lambda$, so indeed, $(a_\alpha)_{\alpha\in\lambda}$ is the zero element. ■

It is very sad that colimits preserve exactness of discrete categories and directed sets, but not in general. These are perhaps our "trial-run" categories, so something very weird is happening to cause this to fail in general.

3.4.7 The Mittag-Leffler Condition

As in our general set-up with limits, we have $\mathcal{I}$ our index category and a given short exact sequence

$$0 \to A_I \xrightarrow{\iota_I} B_I \xrightarrow{\pi_I} C_I \to 0$$

for each $I \in \mathcal{I}$, such that, for each $f : I \to J$ in $\mathcal{I}$, we have $B_f \circ \iota_I = \iota_J \circ A_f$ and $C_f \circ \pi_I = \pi_J \circ B_f$. Because limits commute with limits and hence with kernels, we at least know that

$$0 \to \varprojlim A_k \to \varprojlim B_k \to \varprojlim C_k$$

is exact, so we are worried about the surjectivity of the last map.

For example, do limits preserve exactness over "co-directed" categories? The answer is no.

Example 3.167. Of course, we start with our favorite sequence

$$0 \to \mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0$$

and place all these sequences into a vertical sequence, where the vertical maps are multiplication by 3.

$$\begin{array}{ccccccccc}
 & & \vdots & & \vdots & & \vdots & & \\
 & & \downarrow{\scriptstyle \times 3} & & \downarrow{\scriptstyle \times 3} & & \downarrow{\scriptstyle \times 3} & & \\
0 & \longrightarrow & \mathbb{Z} & \xrightarrow{\times 2} & \mathbb{Z} & \longrightarrow & \mathbb{Z}/2\mathbb{Z} & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \times 3} & & \downarrow{\scriptstyle \times 3} & & \downarrow{\scriptstyle \times 3} & & \\
0 & \longrightarrow & \mathbb{Z} & \xrightarrow{\times 2} & \mathbb{Z} & \longrightarrow & \mathbb{Z}/2\mathbb{Z} & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \times 3} & & \downarrow{\scriptstyle \times 3} & & \downarrow{\scriptstyle \times 3} & & \\
0 & \longrightarrow & \mathbb{Z} & \xrightarrow{\times 2} & \mathbb{Z} & \longrightarrow & \mathbb{Z}/2\mathbb{Z} & \longrightarrow & 0
\end{array}$$

The limit of the $\mathbb{Z}$s must be 0 because no element can be tripled to itself indefinitely, but the limit of the $\mathbb{Z}/2\mathbb{Z}$s will be nontrivial because multiplication by three is an isomorphism. So the resulting sequence is

$$0 \to 0 \to 0 \to \text{nontrivial} \to 0,$$

which is sadly not exact.

So when do limits preserve exactness? The answer is the Mittag-Leffler condition. It turns out that here we only care about the $A_\bullet$ instead of trying to care about the $B_\bullet$ or the map $B_\bullet \to C_\bullet$, which is a testament to short exact sequences caring about all the terms.

We slowly build towards the Mittag-Leffler condition. The set-up is that $\mathcal{I} = \mathbb{N}$ and A_k, B_k, and C_k functors giving a commuting sequence of short exact sequences in that, for $k \le \ell$, the following diagram commutes with exact rows.

$$\begin{array}{ccccccccc}
0 & \longrightarrow & A_\ell & \xrightarrow{\iota_\ell} & B_\ell & \xrightarrow{\pi_\ell} & C_\ell & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle f_k^\ell} & & \downarrow{\scriptstyle g_k^\ell} & & \downarrow{\scriptstyle h_k^\ell} & & \\
0 & \longrightarrow & A_k & \xrightarrow{\iota_k} & B_k & \xrightarrow{\pi_k} & C_k & \longrightarrow & 0
\end{array}$$

Now here is our starting case.

Exercise 3.168. Fix everything as above. If the maps $f_k^{k+1} : A_{k+1} \to A_k$ are surjective, then $\varprojlim B_k \to \varprojlim C_k$ is surjective.

Proof. The idea is to diagram-chase with Lemma 3.150. Fix any $c \coloneqq \{c_k\}_{k\in\mathbb{N}} \varprojlim C_k$ which we want to hit. We recursively[10] construct a sequence $b = \{b_k\}_{k\in\mathbb{N}} \in \varprojlim B_k$ which hits c.

Our base case is to pick up any b_0 which maps to c_0, which exists by exactness of

$$0 \to A_0 \to B_0 \to C_0 \to 0.$$

For the inductive step, we start with $\{b_k\}_{k=1}^n$ and want to find b_{n+1} which maps down to b_n and across to c_{n+1}. Well, we start by simply picking up any b'_{n+1} which goes to c_{n+1}. Here is our diagram so far.

$$\begin{array}{ccc} b'_{n+1} & \xmapsto{\pi_{n+1}} & c_{n+1} \\ & & \downarrow{\scriptstyle h_n^{n+1}} \\ b_n & \xmapsto{\pi_n} & c_n \\ \downarrow & & \downarrow \\ \vdots & & \vdots \end{array}$$

Now, the point is that $b_{n+1} \overset{\pi_{n+1]}}{\mapsto} c_{n+1} \overset{h_n^{n+1}}{\mapsto} c_n$ along one side of the diagram, so along the other side of the diagram,

$$\pi_n(g_n^{n+1} b_{n+1} - b_n) = c_n - c_n = 0.$$

So $g_n^{n+1} b_{n+1} - b_n \in \ker \pi_n = \operatorname{im} \iota_n$, so there is some a_n such that $\iota_n a_n = g_n^{n+1} b_{n+1} - b_n$.

But now the surjectivity of $A_{n+1} \to A_n$ lets us lift a_n to some $a_{n+1} \in A_{n+1}$ with $f_n^{n+1} a_{n+1} = a_n$. And lastly, we push a_{n+1} forwards to define

$$b_{n+1} \coloneqq b'_{n+1} - \iota_{n+1} a_{n+1}.$$

The point is that we still have $\pi_{n+1} b_{n+1} = \pi_{n+1} b'_{n+1} - (\pi_{n+1} \circ \iota_{n+1}) a_{n+1} = c_{n+1}$, but now

$$g_n^{n+1} b_{n+1} = g_n^{n+1} b'_{n+1} (g_n^{n+1} \circ \iota_{n+1}) a_{n+1} = g_n^{n+1} b'_{n+1} 0 (\iota_n \circ f_{n+1}) a_{n+1} = g_n^{n+1} b'_{n+1} - \iota_n a_n,$$

which collapses into what we want after plugging in for $\iota_n a_n$. This finishes the inductive step. ∎

And here is our next case.

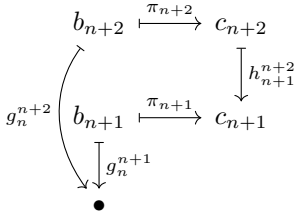
Exercise 3.169. Fix everything as in our set-up from earlier. Further suppose that $A_{k+1} \to A_k$ is the zero map for each $k \in \mathbb{N}$. Then $\varprojlim B_k \to \varprojlim C_k$ is surjective.

Proof. Fix any $\{c_n\}_{n\in\mathbb{N}} \in \varprojlim C_k$ which we want to hit. For each $n \in \mathbb{N}$, find any $b_{n+2} \in B_{n+2}$ such that $\pi_{n+2} b_{n+2} = c_{n+2}$. Similarly, find any $b_{n+1} \in B_{n+1}$ such that $\pi_{n+1} b_{n+1} = c_{n+1}$. Then the main claim is that

$$g_n^{n+2} b_{n+2} \overset{?}{=} g_n^{n+1} b_{n+1}.$$

[10] It is possible to rigorize the following argument with Zorn's lemma. In short, the partially ordered set is over countable sequences $\{b_k\}_{k=0}^N$ with $N \in \mathbb{N} \cup \{\infty\}$ such that $\pi_k b_k = c_k$ for each $0 \le k \le N$. The ordering is given by restriction: $\{b_k\}_{k=0}^N \preceq \{b'_k\}_{k=0}^{N'}$ if and only if $N \le N'$ and $b_k = b'_k$ for each $0 \le k \le N$. The inductive step shows that maximal elements have $N = \infty$.

Here is the diagram.



The claim is equivalent to evaluating $g_n^{n+1}\left(g_{n+1}^{n+2}b_{n+2} - b_{n+1}\right)$. Well, we see that

$$\pi_{n+1}\left(g_{n+1}^{n+2}b_{n+2} - b_{n+1}\right) = \left(\pi_{n+2} \circ h_{n+1}^{n+2}\right)(b_{n+2}) - \pi_{n+1}b_{n+1} = c_{n+1} - c_{n+1} = 0.$$

It follows that $g_{n+1}^{n+2}b_{n+2} - b_{n+1} \in \operatorname{im} \iota_{n+1}$ by exactness. But now we see that

$$g_n^{n+1}(\operatorname{im} \iota_{n+1}) = \operatorname{im} g_n^{n+1} \circ \iota_{n+1} = \operatorname{im} \iota_n \circ f_n^{n+1} = \iota_n(\operatorname{im} f_n^{n+1}) = \iota_n(\{0\}) = \{0\},$$

so indeed, we have that

$$g_n^{n+1}\left(g_{n+1}^{n+2}b_{n+2} - b_{n+1}\right) = 0,$$

which is what we needed.

With the claim finished, we are primed to give the proof. the point is that, we can pick any $\{b_k'\}_{k\in\mathbb{N}}$ such that $\pi_k b_k' = c_k$ for each $k \in \mathbb{N}$. Now the final trick is to set

$$b_n \coloneqq g_n^{n+1}b_{n+1}',$$

for each $n \in \mathbb{N}$. Even though the original $b_\bullet'$ sequence need not be compatible to live in $\varprojlim B_k$, we now see that each $n \in \mathbb{N}$ has

$$g_n^{n+1}b_{n+1} = g_n^{n+1}\left(g_{n+1}^{n+2}b_{n+2}'\right) = g_n^{n+2}b_{n+2}' = g_n^{n+1}b_{n+1}' = b_n$$

by the claim from earlier. So indeed, $\{b_k\}_{k\in\mathbb{N}} \in B$ while

$$\pi_n b_n = \pi_n g_n^{n+1}b_{n+1}' = h_n^{n+1}\pi_{n+1}b_{n+1}' = \pi_{n+1}c_{n+1} = c_n,$$

which is exactly what we needed. ∎

Remark 3.170 (Nir). We can in fact extend this to merely require $A_\ell \to A_k$ to be the zero map for sufficiently large ℓ, given k. This was the way it was presented in lecture, but I did not do this for psychological reasons.

We would like to unify the above two examples. Even though the examples are essentially opposite (trivial cokernel vs. trivial kernel). Regardless, the way to do this is the following somewhat odd condition.

Definition 3.171 (Mittag-Leffler condition). Suppose we have a sequence of (say) modules $\{A_k\}_{k\in\mathbb{N}}$ with morphisms $f_k^\ell : A_\ell \to A_k$ for each $\ell \geq k$, which commute in that $f_\ell^m \circ f_k^\ell = f_k^m$ for each $k \leq \ell \leq m$.

Now, for each k, we check the sequence

$$\operatorname{im}(A_k \to A_k), \qquad \operatorname{im}(A_{k+1} \to A_k), \qquad \operatorname{im}(A_{k+2} \to A_k), \qquad \cdots$$

If the images here stabilize, then we satisfy the *Mittag-Leffler condition*.

Briefly, we can check Definition 3.171 is satisfied in the given examples: when the $A_{k+1} \to A_k$ are surjective, then the images stabilize to A_k; and when the maps are equal, then the images stabilize to 0.

We remark that because $A_{k+n} \to A_k$ is equal to the composite

$$A_{k+n} \to A_{k+n-1} \to \cdots \to A_{k+1} \to A_k,$$

the sequence

$$\operatorname{im}(A_k \to A_k), \qquad \operatorname{im}(A_{k+1} \to A_k), \qquad \operatorname{im}(A_{k+2} \to A_k), \qquad \cdots$$

is in fact decreasing.

Before doing anything formal, we outline "where" this condition is coming from. For each k, set

$$\overline{A_k} = \bigcap_{\ell \in \mathbb{N}} \operatorname{im}(A_{k+\ell} \to A_k)$$

to be the stable image of our sequence. Now, the point is that we have a system of short exact sequences as follows.

$$\begin{array}{ccccccccc}
 & & \vdots & & \vdots & & \vdots & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \overline{A_2} & \longrightarrow & A_2 & \longrightarrow & A_2/\overline{A_2} & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \overline{A_1} & \longrightarrow & A_1 & \longrightarrow & A_1/\overline{A_1} & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \overline{A_0} & \longrightarrow & A_0 & \longrightarrow & A_0/\overline{A_0} & \longrightarrow & 0
\end{array}$$

Indeed, the maps $A_\ell \to A_k$ for $\ell \geq k$ induce the maps on the left and right, where $\overline{A_\ell}$ maps into $\overline{A_k}$ by the stability.

In fact, the image of $\overline{A_\ell} \to A_k$ is the stabilized image of $A_m \to A_\ell \to A_k$ for $m \geq k$, which is the stabilized image of $A_m \to A_k$, which is $\overline{A_k}$. Thus, the maps on the left of our diagram are all surjective! So our work from Exercise 3.168 emerges.

On the other hand, given some k, there is some $\ell \geq k$ so that the image of $A_\ell \to A_k$ is $\overline{A_k}$ by the Mittag-Leffler condition (!), in which case $A_\ell/\overline{A_k} \to A_k/\overline{A_k}$ is the zero map. So the maps on the right of our diagram are all (eventually) zero! Again, again our work from Exercise 3.169 will come into play.

It turns out that there is a way to meld the given arguments for the left column and right column together to get ths surjectivity of $\varprojlim B_k \to \varprojlim C_k$ from the Mittag-Leffler condition on the middle. Here is the main result, as promised.

Theorem 3.172. Fix everything as in the set-up from earlier, and suppose that the $\{A_k\}_{k \in \mathbb{N}}$ satisfy the Mittag-Leffler condition. Then $\varprojlim B_k \to \varprojlim C_k$ is surjective.

Proof. Omitted; see Lang. ∎

To finish off, here's a useful case of the Mittag-Leffler condition at work.

Example 3.173. If all the A_k are finite, then we satisfy the Mittag-Leffler condition, and here we do indeed need the full Mittag-Leffler condition. Essentially this is because

$$A_k \supseteq \operatorname{im}(A_{k+1} \to A_k) \supseteq \operatorname{im}(A_{k+1} \to A_k) \supseteq \operatorname{im}(A_{k+3} \to A_k) \supseteq \cdots$$

is a decreasing sequence of finite groups and hence must stabilize.

3.4.8 Combining Limits and Colimits

Let's do some more abstract category theory.

Remark 3.174. Professor Borcherds is quite aware how much everyone loves category theory.

Recall that limits preserve limits and colimits preserve colimits. However, limits do not necessarily preserve colimits. For example, limits did not preserve right exactness.

However, there is something present.

Proposition 3.175. Given index categories $\mathcal{I}$ and $\mathcal{J}$ with a functor $F : \mathcal{I} \times \mathcal{J} \to \mathcal{C}$, then there is a natural map

$$\varinjlim_{\mathcal{I}} \varprojlim_{\mathcal{J}} F(I,J) \to \varprojlim_{\mathcal{J}} \varinjlim_{\mathcal{I}} F(I,J).$$

The direction of the arrows here matters significantly.

Proof. Fix some objects $i \in \mathcal{I}$ and $j \in \mathcal{J}$. We start with our maps promised by the colimit, which are

$$F(i,j) \to \varinjlim_{I \in \mathcal{I}} F(I,j).$$

Taking the limit over $\mathcal{J}$, we see that these maps induce[11] a map

$$F(i,j) \to \varprojlim_{J \in \mathcal{J}} \varinjlim_{I \in \mathcal{I}} F(I,J).$$

Additionally, we have a map $\varprojlim_{J \in \mathcal{J}} F(i,J) \to F(i,j)$ promised by the limit, so we have the composites

$$\varprojlim_{J \in \mathcal{J}} F(i,J) \to F(i,j) \to \varprojlim_{J \in \mathcal{J}} \varinjlim_{I \in \mathcal{I}} F(I,J).$$

So now we have these maps into an object for each $i \in \mathcal{I}$, so we may assemble these into a map

$$\varinjlim_{I \in \mathcal{I}} \varprojlim_{J \in \mathcal{J}} F(I,J) \to \varprojlim_{J \in \mathcal{J}} \varinjlim_{I \in \mathcal{I}} F(I,J),$$

which is what we wanted. ∎

Let's have an explicit example.

Example 3.176. We work in the category of sets, where $\mathcal{I} = \{0,1\}$ is a category with no maps, where colimits are the coproduct $\sqcup$. We take $\mathcal{J} = \{a,b\}$ as the same category so that limits are the product is $\times$. Now given four sets $S_{0a}, S_{0b}, S_{1a}, S_{1b}$. Now

$$\varprojlim_{\mathcal{J}} \varinjlim_{\mathcal{I}} S_{ij} \cong \varprojlim_{\mathcal{J}} (S_{0j} \sqcup S_{1j}) \cong (S_{0a} \sqcup S_{1a}) \times (S_{0b} \sqcup S_{1b})$$

while

$$\varinjlim_{\mathcal{I}} \varprojlim_{\mathcal{J}} S_{ij} \cong \varinjlim_{\mathcal{I}} (S_{ia} \times S_{ib}) \cong (S_{0a} \times S_{0b}) \sqcup (S_{1a} \times S_{1b}).$$

These are not equal most of the time for size reasons (e.g., make all sets have size 4, and then $4 \cdot 4 + 4 \cdot 4 < (4+4)(4+4)$), though there is an inclusion map upwards by assembling $S_{0a} \times S_{0b} \hookrightarrow (S_{0a} \sqcup S_{1a}) \times (S_{0b} \sqcup S_{1b})$ and $S_{1a} \times S_{1b} \hookrightarrow (S_{0a} \sqcup S_{1a}) \times (S_{0b} \sqcup S_{1b})$.

[11] Technically we have to show that maps we provided commute with the internal maps of the system $\varinjlim_{I \in \mathcal{I}} F(I,j)$. I am going to ignore these sorts of checks for this proof.

Theme 4
Polynomial Pages

The shortest path between two truths in the real domain passes through the complex domain.

—Jacques Hadamard

4.1 October 19

It knows no fear, possibly because it has no brain.

4.1.1 Polynomials Over Fields

So we're talking about polynomials today. Let's review polynomials over a field.

Theorem 4.1. Fix k a field. Then $k[x]$ is Euclidean by using degree for the Euclidean metric.

This has some nice consequences, as usual.

Proposition 4.2. Fix k a field. Then $k[x]$ is a principal ideal domain and hence a unique factorization domain.

Proof. All Euclidean domains are principal ideal domains. And we showed that principal ideal domains are unique factorization domains. ■

We also saw the following directly.

Proposition 4.3. Fix k a field. Any finitely generated $k[x]$-module is a direct sum of cyclic modules.

Proof. We showed that this holds because $k[x]$ is a Euclidean domain, though we technically only need to know that $k[x]$ is a principal ideal domain. ■

We saw that the above proposition implied the Jordan normal form for k algebraically closed.

Example 4.4. Let's look for irreducible polynomials over $\mathbb{F}_2$. Finding irreducible polynomials over $\mathbb{F}_2$ is somewhat similar as finding primes for $\mathbb{Z}$; for example, we can imitate the Sieve of Eratosthenes. We start by writing out all polynomials over $\mathbb{F}_2[x]$, writing them in order of degree.

$$0, \quad 1. \quad x, \quad x+1, \quad x^2, \quad x^2+1, \quad x^2+x, \quad x^2+x+1, \quad x^3, \quad \dots.$$

We ignore 0 and 1, and then we have $\boxed{x}$ and cross out multiples of x (which are elements of zero constant term) giving

$$\boxed{x}, \quad x+1, \quad x^2+1, \quad x^2+x+1, \quad x^3+1, \quad x^3+x+1, \quad \dots.$$

Now we find that $x+1$ is irreducible, and so we can cross our multiples of $x+1$, which are polynomials whose coefficients sum to 0.

$$\boxed{x}, \quad \boxed{x+1}, \quad x^2+x+1, \quad x^3+x+1, \quad x^3+x^2+1, \quad \dots.$$

Now we see that $\boxed{x^2+x+1}$ is irreducible, and it is the only irreducible of degree 2. So the primes $\{x, x+1, x^2+x+1\}$ are enough to determine if any given polynomial of degree at most 4 is irreducible.

We also recall the following statement.

Lemma 4.5. Fix R a commutative ring. If $f \in R[x]$ has $a \in R$ with $f(a) = 0$, then $(x-a) \mid f(x)$.

Proof. For R a field, we can do Euclidean division to write

$$f(x) = (x-a)g(x) + r(x)$$

where $r(x) \in R$ because $r = 0$ or $0 \le \deg r < \deg(x-a) = 1$. But plugging in $x = a$ forces $r = 0$, so conclude $f(x) = (x-a)g(x)$.

However, technically this holds for general rings. Indeed, write

$$f(x) = \sum_{k=0}^{\deg f} a_k x^k$$

so that we have

$$f(x) - f(a) = \sum_{k=0}^{\deg f} a_k \left(x^k - a^k\right) = (x-a) \sum_{k=0}^{\deg f} \left(a_k \sum_{\ell=0}^{k-1} x^\ell a^{k-1-\ell} \right),$$

where the factorization of $x^k - a^k$ is purely formal and hence holds in any commutative ring. It follows that $f(a) = 0$ implies $(x-a) \mid f(x)$. ∎

When we restrict to a field, we get the following.

Proposition 4.6 (Lagrange)**.** A nonzero polynomial f over a field k of degree n has at most n roots.

Proof. This is by induction on n. For $n = 0$, we note that nonzero polynomials of degree 0 are constant and nonzero and hence have no roots.

Then for the inductive step, take $n \coloneqq \deg f > 0$, and we note that if f has no roots, then we are done. Otherwise, f has a root $x_0 \in k$ so that we can write

$$f(x) = (x - x_0)g(x).$$

We note that $f \neq 0$ implies that $g \neq 0$ as well because $k[x]$ has no zero-divisors.

Now the key observation is that k has no zero-divisors, so if a is a root of $f(x)$, then $f(a) = 0$ so that $a - x_0 = 0$ or a is a root of $g(x)$. But $\deg g = \deg f - 1 = n - 1$, so g has at most $n - 1$ roots by induction, so we see that

$$\#\{a \in k : f(a) = 0\} \leq \#\{a \in k : a = x_0\} + \#\{a \in k : g(a) = 0\} \leq 1 + (n-1) = n,$$

which is what we wanted. ■

So fields are nice, but we are obligated to note that things which are not fields are not nice.

Warning 4.7. Lagrange's theorem on polynomials may fail over rings with zero-divisors or over rings which are non-commutative.

Let's see some examples of the above warning.

Example 4.8. In $R = \mathbb{Z}/8\mathbb{Z}$, the polynomial $x^2 - 1$ has the four roots $1, 3, 5, 7$ despite having degree 2.

Example 4.9. In $R = \mathbb{H}$ the ring of quaternions as an $\mathbb{R}$-algebra, the polynomial x^2+1 has an uncountable number of roots. At the very least, $\pm i, \pm j, \pm k$ are roots, but in fact

$$\{bi + cj + dk : b^2 + c^2 + d^2 = 1\}$$

are also roots. Indeed, we can expand and rearrange

$$(bi + cj + dk)^2 = \left(-b^2 - c^2 - d^2\right) + bc(ij + ji) + cd(jk + kj) + db(ki + ik),$$

which evaluates to $-1 + 0 + 0 + 0 = -1$.

Here's a nice application of Lagrange's theorem.

Theorem 4.10. The group of units $\mathbb{F}_p^\times$ is cyclic.

Example 4.11. In $\mathbb{F}_7$, we have that 3 is a generator of $\mathbb{F}_7^\times$. Its powers are $1, 3, 2, 6, 4, 5$, which covers everything. In general, it is hard to explicitly find a generator.

In fact, we can show the following.

Proposition 4.12. Fix k a field and G a finite subgroup of $k^\times$. Then G is cyclic.

Essentially this is saying that the roots of unity are our only candidate finite multiplicative groups in a field. Anyways, let's see this.

Proof. For any $n \in \mathbb{N}$, note that the equation $x^n - 1$ has at most n roots over k. In particular, this implies that G has at most n elements of multiplicative order dividing n. Now there are a few ways to finish from this condition on G.

- Take G a group with at most n elements of multiplicative order dividing n, and we show G is cyclic. There is a clever way to do this by carefully counting the number of elements of a particular order n.

  Let $\varphi_G(n)$ be the number of elements of multiplicative order exactly n in G. It is possible that $\varphi_G(n) = 0$; but if $\varphi_G(n) > 0$ so that there is an element g of order n, then we see that

  $$\langle g \rangle \subseteq \{x \in G : x^n = e\}$$

while $n = \#\langle g\rangle \le \#\{x \in G : x^n = e\} \le n$. Thus, $\langle g\rangle = \{x \in G : x^n = e\}$, so all elements of order dividing in n are in $\langle g\rangle$. But we can count that the number of elements of order n in $\langle g\rangle \cong \mathbb{Z}/n\mathbb{Z}$ is $\varphi(n)$.

The point is that $\varphi_G(n) \in \{0, \varphi(n)\}$. Now, all elements have order dividing into n, so we see that

$$\#G = \sum_{d|\#G} \varphi_G(n) \le \sum_{d|\#G} \varphi(d) = \#G,$$

where the last equality is by Möbius inversion. (Alternatively, count elements of prescribed additive order in $\mathbb{Z}/\#\mathbb{Z}$.) So we need equalities everywhere, so $\varphi_G(\#G) = \varphi(\#G) \neq 0$, meaning that there is an element of order $\#G$, so G is indeed cyclic.

- Here take G a non-cyclic abelian group, and we show that it has an n with more than n elements of multiplicative order dividing n. We use the structure theorem for abelian groups. If G is non-cyclic, then its factorization

$$G \cong \bigoplus_{k=1}^{r} \mathbb{Z}/p_k^{\nu_k}\mathbb{Z}$$

(with $\nu_\bullet \ge 1$) must have the same prime repeated somewhere, lest the factors all be coprime and may be combined by the Chinese remainder theorem.

So without loss of generality take $p_1 = p_2$. Then we see that, for any $a, b \in \mathbb{Z}/p\mathbb{Z}$, we have unique elements

$$p \cdot \left(ap_1^{\nu_1 - 1}, bp_2^{\nu_2 - 1}, 0, 0, \ldots\right) = 0 \in G,$$

but now this gives p^2 elements of multiplicative order dividing p, which finishes. (These elements are unique because $ap_1^{\nu_1-1}$ lives in $\mathbb{Z}/p^{\nu_1}\mathbb{Z}$ and similar for $bp_2^{\nu_2-1}$.) ∎

Remark 4.13. Professor Borcherds does not care what happens for nonabelian groups.

Proof of Theorem 4.10. Because $\mathbb{F}_p$ is finite, $\mathbb{F}_p^\times$ is a finite cyclic group of a field, so it is cyclic. ∎

Anyways, we do care for fields which are not finite sometimes.

Example 4.14. In $\mathbb{C}$, we now see that any finite multiplicative subgroup $G \subseteq \mathbb{C}^\times$ must be cyclic and hence essentially roots of unity. Of course, we can see this somewhat directly because all $g \in G$ must have $g^{\#G} = 1$ and hence be roots of unity, and then we can check for the smallest n for which $g^n - 1 = 0$ for each $g \in G$.

And naturally, this fails when k is not a field.

Non-Example 4.15. In $\mathbb{Z}/8\mathbb{Z}$, the group of units $\{1, 3, 5, 7\}$ is non-cyclic because it has no generator, or no "primitive root."

Non-Example 4.16. The quaternions have many finite non-cyclic subgroups. For example, the subgroup $\{\pm 1, \pm i, \pm j, \pm k\}$ is non-cyclic. Additionally, the binary permutation groups we found earlier in this course work as well.

Remark 4.17. If the field is finite, then we can have a polynomial vanish at all points without being 0. For example, in $\mathbb{F}_p$, everything is a root of $x^p - x$. The point here is that being zero as a function or a polynomial are different here. (However, this is "essentially" the only counterexample: if $f(x) = g(x)$ on $\mathbb{F}_p$, then $x^p - x = x(x-1)(x-2)\cdots(x-p+1) \mid f(x) - g(x)$.)

Regardless, if the field is infinite, then $f(a) = g(a)$ on each $a \in k$ does imply $f = g$ as polynomials because $f - g$ would have infinitely many roots and hence must be the zero polynomial.

4.1.2 Polynomials Over Unique Factorization Domains

Now let's move to polynomials over $\mathbb{Z}[x]$. Note that $\mathbb{Z}[x]$ is not a principal ideal domain and hence not Euclidean.

Example 4.18. The ideal $(2, x) \subseteq \mathbb{Z}[x]$, which consists of the polynomials of even constant term, is not principal.

However, $\mathbb{Z}[x]$ does have unique prime factorization!

Theorem 4.19. The ring $\mathbb{Z}[x]$ has unique factorization.

Remark 4.20 (Nir)**.** Those familiar with this proof already are encouraged to think of $\mathbb{Z}$ in the following proof as a general unique factorization domain. I will not write this proof out explicitly in this generality for psychological reasons, but we will not use anything of $\mathbb{Z}$ beyond that it is a unique factorization domain anyways.

Proof. The idea here is to reduce to $\mathbb{Q}[x]$, which we know has unique factorization. What is annoying here is that

$$3x^2 + 6$$

is irreducible over $\mathbb{Q}[x]$ because 3 is a unit in $\mathbb{Q}$, but 3 is a prime in $\mathbb{Z}[x]$, so $3 \cdot (x^2 + 2)$ is a nontrivial factorization in $\mathbb{Z}[x]$.

To deal with this, we have the following definition.

Definition 4.21 (Content)**.** Given nonzero $f \in \mathbb{Z}[x]$, we define the *content* $c(f)$ to be the greatest common divisor of the coefficients of f. (In general, for R a unique factorization domain, we may set $c(f)$ to be the ideal generated by the greatest common divisor, to avoid unit problems.)

Example 4.22. The content of $3x^2 + 6$ is 3.

It follows that, for any $f \in \mathbb{Z}[x]$, we have that $f/c(f)$ is a polynomial in $\mathbb{Z}[x]$, where the coefficients are coprime.

We also note that an integer n divides into f if and only if $n \mid c(f)$.[1] Indeed, setting

$$f(x) = \sum_{k=0}^{\deg f} a_k x^k,$$

we have that $n \mid c(f)$ implies $n \mid a_k$ for each k implies that $\frac{1}{n} f(x) = \sum_{k=0}^{\deg f} \frac{a_K}{n} x^k \in \mathbb{Z}[x]$. Conversely, if $n \mid f$, then $f = gn$ for some $g \in \mathbb{Z}[x]$, but writing out the coefficients of g shows that $a_k = nb_k$ for some b_k, for each k. This finishes.

The main result is as follows.

Lemma 4.23 (Gauss's)**.** Fix $f, g \in \mathbb{Z}[x]$ nonzero. Then $c(f)c(g) = c(fg)$.

Proof. The fact that $c(f)c(g) \mid c(fg)$ is not hard: it suffices to show that $c(f)c(g) \mid fg$, but this is true because $c(f) \mid f$ and $c(g) \mid g$. So the problem is showing equality. Because the content preserves multiplication by a constant ($n \mid f$ if and only if $n/a \mid f/a$ for some a), we see that we are interested in showing

$$c\left(\frac{f}{c(f)} \cdot \frac{g}{c(g)}\right) = \frac{c(fg)}{c(f)c(g)} \stackrel{?}{=} 1.$$

[1] This is more or less why we care about the content: it is extracting out the "non-field" part of an irreducible.

So setting $f \leftarrow f/c(f)$ and $g \leftarrow g/c(g)$, we see that it suffices to show $c(fg) = 1$ given that $c(f) = c(g) = 1$.

For this we show that each irreducible $p \in \mathbb{Z}$ does not divide $c(fg)$, which will be good enough by, say, ring theory: this implies that the content is not contained in any maximal ideal and hence must be a unit. For concreteness, we set

$$f(x) = \sum_{k=0}^{\deg f} a_k x^k \qquad \text{and} \qquad g(x) = \sum_{k=0}^{\deg g} b_k x^k.$$

Because $c(f) = 1$, we know that there is some m for which $p \nmid a_m$, so there is a least m for which $p \nmid a_m$; similarly, there is a least n for which $p \nmid b_n$. Multiplying, we find that

$$(fg)(x) = \sum_{r=0}^{\deg f + \deg g} \left(\sum_{k+\ell=r} a_k b_\ell \right) x^r.$$

The point is that the coefficient with degree $r = m + n$ looks like

$$\sum_{k+\ell=m+n} a_k b_\ell = \left(\sum_{k=0}^{m-1} a_k b_{m+n-k} \right) + a_m b_n + \left(\sum_{\ell=0}^{n-1} a_{m+n-\ell} b_\ell \right).$$

Here, each term for the left sum is divisible by p because each of the $a_\bullet$ are. Similarly, each term for the right sum is divisible by p because each of the $b_\bullet$ are. But $p \nmid a_m b_n$, so we see that the coefficient has

$$\sum_{k+\ell=m+n}^{m+n} a_k b_\ell \equiv a_m b_n \not\equiv 0 \pmod{p}.$$

Thus, there is a coefficient of fg not divisible by p, so we conclude that $p \nmid c(fg)$. This finishes the proof, as described. ∎

The main use of Gauss's lemma is to classify the irreducibles over $\mathbb{Z}[x]$. Here is a technical lemma that will come up a couple of times.

Lemma 4.24. Suppose that $f \in \mathbb{Z}[x]$ has content $c(f) = 1$, and $q \in \mathbb{Q}$ gives $qf \in \mathbb{Z}[x]$ while $c(qf) = 1$. Then $q \in \mathbb{Z}^\times$ is a unit in $\mathbb{Z}$.

Proof. It suffices to show that $\nu_p(q) = 0$ for each prime p of $\mathbb{Z}$. For concreteness, we set

$$f(x) = \sum_{k=0}^{\deg f} a_k x^k$$

so that $c(f) = \gcd_k(a_k) = 1$. Taking the greatest common denominator in $\mathbb{Z}$ as a unique factorization domain, we find that

$$0 = \nu_p(c(qf)) = \nu_p(\gcd_k(qa_k)) = \min_k \left(\nu_p(q) + \nu_p(a_k) \right) = \nu_p(q) + \min_k \nu_p(a_k) = \nu_p(q),$$

where we have used that $c(f) = 1$ in the last equality. This is what we wanted. ∎

And here is our classification of irreducibles.

Lemma 4.25. The irreducibles in $\mathbb{Z}[x]$ are either irreducible elements in $\mathbb{Z}$ or irreducible in $\mathbb{Q}[x]$ with content a unit.

Proof. Fix $\pi \in \mathbb{Z}[x]$ an irreducible. We remark that π is either a unit or irreducible in $\mathbb{Q}[x]$: if π is constant, than it is a unit. Otherwise, suppose that we have a nontrivial factorization $\pi = \alpha\beta$ for $\alpha, \beta \in \mathbb{Q}[x]$, and we have to show that one of the factors is a unit.

By clearing denominators, there exists a' such that $a'\alpha \in \mathbb{Z}[x]$, and then we define $a \coloneqq a'/c(a'\alpha)$ so that $a\alpha \in \mathbb{Z}[x]$ with content 1. We define b similarly so that $b\beta \in \mathbb{Z}[x]$ with content 1. The point is that

$$ab\pi = (a\alpha)(b\beta) \in \mathbb{Z}[x],$$

and by Gauss's lemma, $c(ab\pi) = 1$. So by Lemma 4.24, we conclude that $ab \in \mathbb{Z}^\times$ is a unit, so $ab\pi$ is irreducible in $\mathbb{Z}[x]$. Thus, $a\alpha$ or $b\beta$ is a unit in $\mathbb{Z}[x]$ and hence a unit in $\mathbb{Q}[x]$, which finishes the check that π is irreducible.

We turn directly to our classification. The point is that we can factor

$$\pi = c(\pi) \cdot \frac{\pi}{c(\pi)}.$$

Because π is irreducible in $\mathbb{Z}[x]$, one of these factors is a unit in $\mathbb{Z}[x]$. We have two cases.

- If $c(\pi)$ is a unit, then we note that π must be non-constant, lest it divide into its constant term and hence divide into $c(\pi) = 1$ and be a unit. Thus, π is not a unit in $\mathbb{Q}[x]$, so π is irreducible in $\mathbb{Q}[x]$ with content 1.
- If $\pi/c(\pi)$ is a unit, then we note that our equation implies that π is a unit multiplied by $c(\pi) \in \mathbb{Z}$, so $\pi \in \mathbb{Z}$. It remains to show that π is irreducible in $\mathbb{Z}$. Well, π is not a unit in $\mathbb{Z}$ because it is not a unit in $\mathbb{Z}[x]$, and if $\pi = ab$ where $a, b \in \mathbb{Z}$, then this factorization lifts to $\mathbb{Z}[x]$, so one of a or b is a unit in $\mathbb{Z}[x]$ and hence a unit in $\mathbb{Z}$.

This finishes the classification of irreducibles in $\mathbb{Z}[x]$.

It remains to verify that these are all in fact irreducible. We have two cases.

- If π is irreducible in $\mathbb{Z}$, then $\pi = ab$ in $\mathbb{Z}[x]$ must have $a, b \in \mathbb{Z}$ by degree arguments, but then one of a, b is a unit in $\mathbb{Z}$ and hence a unit in $\mathbb{Z}[x]$.
- If π is irreducible in $\mathbb{Q}[x]$ with content 1, then take $\pi = \alpha\beta$. By taking this into $\mathbb{Q}$, we see that one of α or β must be a unit in $\mathbb{Q}[x]$, which means that one of α or β is constant. But α and β must have content 1 by Gauss's lemma, so we conclude one of α or β is a unit in $\mathbb{Z}[x]$. ∎

We now attack unique factorization. Showing that every element has a factorization comes down to $\mathbb{Z}[x]$ being Noetherian, roughly speaking. We proceed along the same outline as when we showed that principal ideal domains were unique factorization domains.

Lemma 4.26. Fix $f \in \mathbb{Z}[x]$ not zero and not a unit. Then f is divisible by some irreducible element of $\mathbb{Z}[x]$.

Proof. If f is constant, then this reduces to the situation in $\mathbb{Z}$. Similarly, if f has content not a unit, then f has an prime factor in $\mathbb{Z}$, which we know to be irreducible. Otherwise, f is non-constant, so embedding f into $\mathbb{Q}[x]$, it has some irreducible factor $\alpha \in \mathbb{Q}[x]$ so that $f = \alpha\beta$ for some β.

However, we might have $\alpha \notin \mathbb{Z}[x]$, so there is still work to be done. Clear denominators to find some $a' \in \mathbb{Z}$ with $a'\alpha \in \mathbb{Z}[x]$, and then we define $a \coloneqq a'/c(a'\alpha)$ so that $a\alpha \in \mathbb{Z}[x]$ with content 1. We can do similar for β to get $b \in \mathbb{Z}$ so that $b\beta \in \mathbb{Z}[x]$ with content 1. Then

$$abf = (a\alpha)(b\beta) \in \mathbb{Z}[x]$$

while the right-hand side has content 1. We conclude from Lemma 4.24 that ab is a unit in $\mathbb{Z}[x]$. It follows $a\alpha \mid abf \mid f$, so $a\alpha$—which is irreducible in $\mathbb{Q}[x]$ with content 1—is the irreducible we are looking for. ∎

Lemma 4.27. Every nonzero $f \in \mathbb{Z}[x]$ has a factorization into irreducibles.

Proof. The morally correct thing to do here would be to show that $\mathbb{Z}[x]$ is Noetherian using the Hilbert basis theorem, but the correct machinery is annoying to build. So we cheat.

We proceed by induction on $\deg f$. If $\deg f = 0$, then $f \in \mathbb{Z}$, so it has a factorization into irreducibles because all elements of $\mathbb{Z}$ do, and $\mathbb{Z}$-irreducibles are $\mathbb{Z}[x]$-irreducibles.

Now take f of positive degree. We proceed by induction on the number of (not necessarily distinct) irreducible factors of $c(f)$. If $c(\pi)$ has no irreducible factors, then we note that f is still nonzero and not a unit, so it has some irreducible factor π.

But now π must be of positive degree because π being constant would divide into the content. So we see that f/π has degree smaller than f, so f/π has a factorization into irreducibles, so f has a factorization into irreducibles.

So suppose that $c(f)$ has a positive number of irreducible factors. Let one such $\mathbb{Z}$-irreducible factor be p. But then f/p has the same degree as f while having one fewer factor in $c(f/p) = c(f)/p$, so we can induct downwards here. ■

Remark 4.28 (Nir)**.** I am fairly sure that the above proof still works in general unique factorization domains (namely, not assuming Noetherian), but this requires some care. I think one should do induction on $\deg f + \sum_{\pi \text{ irred.}} \nu_\pi(c(f))$, where the sum is finite because R is a unique factorization domain.

We now turn to showing uniqueness of factorizations. The key is the following lemma.

Lemma 4.29. An element $\pi \in \mathbb{Z}[x]$ is irreducible if and only if it is a nonzero prime.

Proof. If π is prime, it is not too hard to check that π is irreducible. Note π is not a unit because π is prime. For the hard check, write $\pi = \alpha\beta$ for some $\alpha, \beta \in \mathbb{Z}[x]$ implies that $\pi \mid \alpha\beta$, so $\pi \mid \alpha$ or $\pi \mid \beta$ because π is prime. Without loss of generality take $\pi \mid \alpha$, but then $\pi = \pi(\alpha/\pi)\beta$, so β is a unit, finishing.

The other direction is more difficult; fix π irreducible, and we simply run through our classification to check that it is prime.

- Take π a prime in $\mathbb{Z}$. The point is that $\pi = uc(\pi)$ for some unit $u \in \mathbb{Z}$. Now, $\pi \mid \alpha\beta$ in $\mathbb{Z}[x]$ implies that $c(\pi) \mid c(\alpha)c(\beta)$, where we are using Gauss's lemma quite liberally. But $c(\pi)$ is a prime in $\mathbb{Z}$, so π divides $c(\alpha)$ or $c(\beta)$, so π divides α or β.
- Take π an irreducible in $\mathbb{Q}[x]$ with content 1. Now take $\pi \mid \alpha\beta$ for some $\alpha, \beta \in \mathbb{Z}[x]$. Taking $\alpha \leftarrow \alpha/c(\alpha)$ and $\beta \leftarrow \beta/c(\beta)$, it suffices to take α and β with content 1.

  Now, the trick is to embed this into $\mathbb{Q}[x]$ so that $\pi \mid \alpha$ or $\pi \mid \beta$ in $\mathbb{Q}[x]$. Without loss of generality, $\pi \mid \alpha$, so take $\gamma \in \mathbb{Q}[x]$ such that
  $$\pi\gamma = \alpha$$
  in $\mathbb{Q}[x]$. It remains to show that $\gamma \in \mathbb{Z}[x]$. Well, as usual, we can clear denominators and then divide out by the content to get $g \in \mathbb{Q}$ such that $g\gamma \in \mathbb{Z}[x]$ with content 1. But now
  $$g\alpha = \pi(g\gamma) \in \mathbb{Z}[x],$$
  where the right-hand side has content 1. But α has content 1, so Lemma 4.24 shows that g is a unit in $\mathbb{Z}$ and hence in $\mathbb{Z}[x]$. So $g\gamma \in \mathbb{Z}[x]$ shows $\gamma \in \mathbb{Z}[x]$. ■

And now we can show uniqueness of factorizations

Lemma 4.30. Factorization into irreducibles in $\mathbb{Z}[x]$ is unique.

Proof. This follows from Proposition 2.69. Yes, exactly the same proof works now. ■

This finishes the proof of Theorem 4.19. ■

As we remarked earlier, the above argument can be pushed further to show the following.

Theorem 4.31. Fix R is a unique factorization domain, then $R[x]$ is also a unique factorization domain.

Proof. Copy the proof of Theorem 4.19, replacing each occurrence of $\mathbb{Z}$ with R and each occurrence of $\mathbb{Q}$ with $\mathrm{Frac}(R)$. ■

The main point is that Lemma 4.23 used that $\mathbb{Z}$ was a unique factorization domain, but that is the only thing of $\mathbb{Z}$ that we need to make this work.

Example 4.32. Inducting on the above theorem, we can say that $k[x_1, \dots, x_n]$ is a unique factorization domain for any $n \in \mathbb{N}$.

Example 4.33. In fact, $k[x_1, x_2, \dots]$ going on infinitely is also a unique factorization domain because polynomials are finite, so any polynomial here must live in some finite $k[x_1, \dots, x_n]$, which we know has unique factorization.

Example 4.34. Doing a similar induction shows that $\mathbb{Z}[x_1, \dots, x_n]$ is a unique factorization domain.

4.1.3 Effective Factorization for $\mathbb{Z}[x]$

We might be interested in a real factorization algorithm for $\mathbb{Z}[x]$. Of course, this is hard even for degree zero polynomials because factoring (large) integers is difficult, but what can we do?

Here is a slow algorithm, due to Kronecker. Fix $f \in \mathbb{Z}[x]$ of degree n.

1. Choose n integers $a_1, \dots, a_n$ and look at $f(a_1), \dots, f(a_n)$. If any are zero, then we have a linear factor and can induct downwards.
2. Otherwise, we look at all factorizations of $f(a_\bullet)$ down in $\mathbb{Z}$. The point is that $g(a_\bullet) \mid f(a_\bullet)$ always if $g \mid f$, so there are only finitely many possibilities of the $g(a_\bullet)$, and because we have n points $g(a_\bullet)$ here, our finitely many possibilities can be uniquely interpolated to g. So we can check all of these possibilities.

Of course, factoring the $f(a_\bullet)$ is difficult, and doing it n times is somewhat annoying. Additionally, the "finite check" at the end is potentially very large if the $f(a_\bullet)$ have lots of factors; at the very least we will have ± 1 to check, which gives 2^n possibilities for g.

Speeding up Kronecker's algorithm is hard. We remark that there is the Lenstra–Lenstra–Lovasz algorithm which can factor in $\mathbb{Q}[x]$ in polynomial time. So the point is that we can factor after getting rid of the content, so we have "reduced" fast factorization of polynomials in $\mathbb{Z}[x]$ to just factoring the content of the polynomial, which of course is somewhat hard.

Remark 4.35 (Nir)**.** Get used to seeing the L^3 algorithm around. It comes up everywhere in computational number theory.

Remark 4.36. Shor's algorithm can do fast factorization, if we have a large quantum computer (with on the order of millions of qubits).

Remark 4.37. Professor Borcherds seems somewhat bitter about all the quantum hype.

We also remark that even though we can factor in $\mathbb{Z}[x_1, \dots, x_n]$ (for example, iterate Kronecker's algorithm), there is literally no algorithm to check for $\mathbb{Z}$-roots. This is by the work down in resolving Hilbert's 10th problem by Matiyasevich and others.

4.1.4 Irreducibility Testing

If we cannot factor, the next best thing is to test if polynomials in $f \in \mathbb{Z}[x]$ are irreducible.

The Generic Test

Here is a probabilistic test.

1. Return that the problems is irreducible.

This is not terribly interesting, but generic polynomials in $\mathbb{Z}[x]$ do turn out to be irreducible.

Reduce $\pmod p$

Here is one possible test. Fix $f(x) \in \mathbb{Z}[x]$.

1. Fix p a prime not dividing the leading coefficient of $f(x)$, and we check if $\overline{f} \in \mathbb{F}_p[x]$ is irreducible. If irreducible, return that f is irreducible.
2. If $\overline{f}$ is not irreducible, then check another prime.

The point is that if $\overline{f}$ is irreducible in $\mathbb{F}_p[x]$, then we can lift this to irreducibility in $\mathbb{Z}[x]$.

Lemma 4.38. Fix p a prime. Suppose that $f(x) \in \mathbb{Z}[x]$ has leading coefficient not divisible by p. If $\overline{f} \in \mathbb{F}_p[x]$ is irreducible, then $f \in \mathbb{Z}[x]$ is irreducible.

Proof. We proceed by contraposition. Suppose we have a nontrivial factorization $f = gh$ in $\mathbb{Z}[x]$ and write it in $\mathbb{F}_p[x]$ as

$$\overline{f} = \overline{g} \cdot \overline{h}.$$

Because p does not divide the leading coefficient of f, it won't divide the leading coefficients of either g or h, so $\deg g, \deg h \geq 1$ implies that $\deg \overline{g}, \deg \overline{h} \geq 1$. Thus, $\overline{f} = \overline{g} \cdot \overline{h}$ is indeed a nontrivial factorization. ∎

Anyways, let's see some examples of this algorithm.

Example 4.39. The polynomial $x^4 + x + 1$ is irreducible in $\mathbb{F}_2[x]$ as we showed earlier: it has constant term 1, so it is not divisible by x; it's coefficient sum is 1, so it is not divisible by $x + 1$; lastly, we see

$$(x+1)\left(x^2 + x + 1\right) = x^3 - 1,$$

so $x^4 + x + 1 \equiv x + x + 1 \equiv 1 \pmod{x^2 + x + 1}$, so we are not divisible by $x^2 + x + 1$ either, which is enough. The point is that it follows $x^4 + x + 1$ is irreducible in $\mathbb{Z}[x]$.

Non-Example 4.40. The polynomial $3x^3 + x^2 + 3x + 1$ reduces to $x^2 + 1 \pmod 3$, which is irreducible in $\mathbb{F}_3[x]$ (it has no roots), but we can still factor

$$3x^3 + x^2 + 3x + 1 = (3x+1)\left(x^2+1\right).$$

The issue is that reducing $\pmod 3$ will view $3x + 1$ as a unit even though it is not a unit in $\mathbb{Z}[x]$.

Remark 4.41 (Nir)**.** This algorithm is not effective: there are irreducible polynomials which factor non-trivially modulo every prime. For example,

$$x^4 + 1$$

factors modulo every prime even though it is irreducible in $\mathbb{Z}$. We can show the factorization by hand (do casework on which of $\left(\frac{-1}{p}\right), \left(\frac{2}{p}\right), \left(\frac{-2}{p}\right)$ is equal to 1).

Eisenstein's Criterion

Here is Eisenstein's criterion.

Proposition 4.42 (Eisenstein)**.** Fix $f \in \mathbb{Z}[x]$ given by

$$f(x) = \sum_{k=0}^{n} a_k x^k.$$

If prime p divides all $a_\bullet$ except a_n, and $p^2 \nmid a_0$, then f is irreducible in $\mathbb{Q}[x]$.

Remark 4.43. Apparently everyone remembers this from undergrad.

Remark 4.44 (Nir)**.** Recalling from our previous work, we note that we need $c(f) = 1$ in order to be sure f is irreducible in $\mathbb{Z}[x]$. As a counterexample, $3x + 6$ is not irreducible in $\mathbb{Z}[x]$ even though it satisfies Eisenstein's criterion for $p = 2$.

Proof. The idea is to take $f(x) = g(x)h(x)$ and eventually reach $p \mid a_n$. Indeed, we proceed by contraposition, supposing that $f = gh$ is a nontrivial factorization in $\mathbb{Q}[x]$ with $p \mid a_k$ with $0 \le k < n$ and $p^2 \nmid a_0$. Then we claim that $p \mid a_n$.

By using the typical content tricks, we can force $c(f) = 1$ so that any nontrivial factorization $f = gh$ can be forced to have $g, h \in \mathbb{Z}[x]$ by the typical content tricks. Now set

$$g(x) = \sum_{k=0}^{\deg g} b_k x^k \qquad \text{and} \qquad h(x) = \sum_{\ell=0}^{\deg h} c_\ell x^\ell.$$

By checking the constant term, we see that $p \mid b_0 c_0$, so $p \mid b_0$ or $p \mid c_0$. However, $p^2 \nmid b_0 c_0 = a_0$, so p cannot divide both. So without loss of generality Then without loss of generality $p \mid b_0$ and $p \nmid c_0$.

Now we claim that $p \mid b_m$ for each $m \le \deg g$ by induction, where we have already done our base case. Indeed, if $p \mid b_k$ for $k < m$, then we see look at

$$a_m = \sum_{k+\ell=m} b_k c_\ell,$$

which reduces to $b_m c_0 \equiv 0 \pmod{p}$; in particular, $p \mid a_m$ because $m \le \deg g < \deg f$. But now $p \nmid c_0$ shows $p \mid b_m$, as needed.

So to finish, we note that $p \mid g$ as a polynomial, so it follows $p \mid f$, implying that $p \mid a_n$. This finishes the proof. ∎

Here is the standard example of Eisenstein's criterion.

Example 4.45. We show that $\Phi_p(x) := \frac{x^p - 1}{x - 1} = 1 + \cdots + x^{p-1}$ is irreducible. The trick is to plug in $x \mapsto x + 1$, for nontrivial factorizations of $\Phi_p(x)$ an be turned into nontrivial factorizations of $\Phi_p(x + 1)$. Well, we can evaluate

$$\Phi_p(x+1) = \frac{(x+1)^p - 1}{(x+1) - 1} = \frac{1}{x}\left((x+1)^p - 1\right) = \sum_{k=1}^{p} \binom{p}{k} x^{k-1}$$

by the binomial theorem.

We now check Eisenstein's criterion using p as our prime. The leading term is x^{p-1}, which is indeed not divisible by p. The middle terms have coefficients of $\binom{p}{k} = \frac{p!}{k!(p-k)!}$ where $0 < k < p$, so they are all divisible by p because the numerator has p while the denominator does not. And lastly, the constant term is $\binom{p}{1} = p$, so it is not divisible by p.

Example 4.46. Fix n a positive integer. We can also check that $\Phi_{p^n}(x) := \left(x^{p^n}-1\right) / \left(x^{p^{n-1}}-1\right) = \sum_{k=0}^{p-1} x^{p^{n-1}k}$ is also irreducible using a similar trick. The idea is that

$$x^{p^n} - 1 \equiv \left(x^{p^{n-1}} - 1\right)^p \pmod{p}$$

by the binomial theorem, so it follows

$$\Phi_{p^n}(x) = \frac{x^{p^n}-1}{x^{p^{n-1}}-1} \equiv \left(x^{p^{n-1}} - 1\right)^{p-1} \pmod{p}.$$

Again using the binomial theorem, we have

$$\Phi_{p^n}(x) \equiv (x-1)^{p^{n-1}(p-1)} \pmod{p}.$$

It follows that $\Phi_{p^n}(x+1) \equiv x^{p^{n-1}(p-1)} \pmod{p}$, so all terms except for the leading term of $\Phi_{p^n}(x+1)$. are divisible by p. Further, the constant term of $\Phi_{p^n}(x+1)$ is $\Phi_{p^n}(1) = p$ and notably not divisible by p^2. So we are done by Eisenstein's criterion on p.

Remark 4.47. The reason why Eisenstein's criterion works is roughly speaking due to totally ramified primes of $\mathbb{Q}$. For example, this works for $\Phi_p(x)$ as above because (p) is totally ramified in $\mathbb{Q}(\zeta_p)$.

Intermission: Aurifeuillian Factorization

Before continuing, we remark that

$$x^4 + 4a^4$$

for some fixed $a \in \mathbb{Z}$ looks irreducible but isn't. Namely,

$$x^4 + 4a^4 = \left(x^4 + 4a^2x^2 + 4a^4\right) - (2ax)^2 = \left(x^2 + 2xa + 2a^2\right)\left(x^2 - 2xa + 2a^2\right)$$

by using the difference of two squares factorization. This is the Aurifeuillian factorization.

Remark 4.48. In general, sums of monomials are potentially tricky. There is a page on Wikipedia for other such factorizations.

Example 4.49. The number n^4+4^n is never prime for $n > 1$. If n is even, then n^4+4^n is even. Otherwise, we take $n = 2m+1$ so that we can write

$$n^4 + 4^n = n^4 + 4 \cdot (2^m)^4 = \left(n^2 + 2^{m+1}n + 2^{2m+1}\right)\left(n^2 - 2^{m+1}n + 2^{2m+1}\right).$$

It remains to show that each term is bigger than 1. Well,

$$n^2 \pm 2^{m+1}n + 2^{2m+1} = (n \pm 2^m)^2 + 2^{2m}$$

after some rearranging, and surely $2^{2m} > 1$ because $m \geq 1$ from $n \geq 3$.

Example 4.50. Factoring the number $2^{58}+1$ was hard, as done by Laundry in 1869. But in 1871, Aurifeuille showed that the factorization is trivial because this is

$$2^{58} + 1 = 1^4 + 4\left(2^{14}\right)^4.$$

Rational Root Theorem

Let's continue discussing our polynomial factorization. We can also check for rational roots in hopes of finding a linear factor; we have the following two statements.

Proposition 4.51. Fix

$$f(x) = \sum_{k=0}^{\deg f} a_k x^k \in \mathbb{Z}[x].$$

If $(ax+b) \mid f(x)$, then $a \mid a_{\deg f}$ and $b \mid a_0$.

Proof. Write

$$f(x) = (ax+b) \sum_{k=0}^{\deg f - 1} b_k x^k,$$

where the $\deg f - 1$ is by degree arguments. Then $a_0 = bb_0$ and $a_{\deg f} = ab_{\deg f - 1}$, which is what we wanted. ∎

Proposition 4.52. Fix $f(x) \in \mathbb{Z}[x]$. Then, given $a, b \in \mathbb{Z}$ with $\gcd(a,b) = 1$, we have $(ax+b) \mid f(x)$ if and only if $f(-b/a) = 0$.

Proof. In one direction, if $(ax+b) \mid f(x)$, then write $f(x) = (ax+b)g(x)$ for $g \in \mathbb{Z}[x]$. Then $f\left(-\frac{b}{a}\right) = 0 \cdot g\left(-\frac{b}{a}\right) = 0$, which is what we wanted.

The other direction is harder. The point is that $c(ax+b) = 1$. Without loss of generality, take f of content 1, for this does not change $f(-b/a) = 0$, and $(ax+b) \mid f(x)/c(f)$ implies $(ax+b) \mid f(x)$. Certainly $x + \frac{b}{a} \mid f(x)$ in $\mathbb{Q}[x]$ because $-\frac{b}{a}$ is a root of f in $\mathbb{Q}[x]$, so it follows $ax+b \mid f(x)$ in $\mathbb{Q}[x]$. So we get some $g(x) \in \mathbb{Q}[x]$ such that

$$f(x) = (ax+b)g(x).$$

As usual, we can find some $q \in \mathbb{Q}$ such that $qg \in \mathbb{Z}[x]$ with content 1, but then

$$(qf)(x) = (ax+b)(qg)(x) \in \mathbb{Z}[x].$$

But now Lemma 4.24 lets us conclude that $q \in \mathbb{Z}^\times$, so $g = q^{-1} \cdot qg \in \mathbb{Z}[x]$, which finishes. ∎

So the combination of these two give us a viable way to check for linear factors: create candidates by using Proposition 4.51, and then test the candidates using Proposition 4.52.

This can actually be used to test irreducibility of polynomials of degree at most three can be because degree-three polynomials must factor with some linear term (by degree arguments) if they factor nontrivially at all. However, things become worse with higher degrees because we must take into account quadratic factors and so on.

Example 4.53. The polynomial $x^3 - 3x + 1$ is irreducible. Namely, if it were to factor, it would have a linear factor, so it would have a rational root, but the only candidates are ± 1, which are not roots because 1 gives -1, and -1 gives 1.

Remark 4.54. The roots of $f(x) := x^3 - 3x + 1$ are in fact

$$2\cos\left(\frac{2\pi}{9}\right), \qquad 2\cos\left(\frac{4\pi}{9}\right), \qquad 2\cos\left(\frac{8\pi}{9}\right).$$

Indeed, the point is that $f(2x)/2 = 4x^3 - 3x + \frac{1}{2}$, which resembles the triple angle formula: we have

$$f(2\cos\theta) = 4\cos^3\theta - 3\cos\theta + \frac{1}{2} = \cos 3\theta + \frac{1}{2},$$

so we want θ with $\cos 3\theta = -\frac{1}{2}$; these exactly give the roots. Anyways, the roots of $f(x)$ come up when showing that the 60° angle cannot be trisected using ruler and compass, for then we could construct a root of $x^3 - 3x + 1$.

Remark 4.55. Ruler and compass constructions Professor Borcherds might mention again for at most two seconds. It is much easier to trisect an angle using a protractor.

Berlekamp's Algorithm

Lastly, let's outline the ideas for Berlekamp's algorithm, which works at reasonable speed for factoring in $\mathbb{F}_p[x]$. Fix $f \in \mathbb{F}_p[x]$, and we note that we compute $\gcd(f, g)$ somewhat efficiently by Euclidean division.

The key point is that

$$\prod_{\substack{\pi \text{ monic, irred.} \\ \deg \pi = 1}} \pi = \prod_{k \in \mathbb{F}_p} (x - k) = x^p - x,$$

so $\gcd(f, x^p - x)$ will quickly check if f has any linear factors. More generally, it is a result from the theory of finite fields that

$$\prod_{\substack{\pi \text{ monic, irred.} \\ \deg \pi \mid d}} \pi = x^{p^d} - x.$$

For example, the factors on the left-hand side are coprime, and each divides into $x^{p^d} - x$ because the roots of any polynomial on the left-hand side will be inside of $\mathbb{F}_{p^d}$, where all elements satisfy $x^{p^d} - x = 0$.

Anyways, the point is that we can check f for having any irreducible factor of degree dividing into d by computing $\gcd\left(f, x^{p^d} - x\right)$. By looping over the possible d, this is able to quickly check if f is irreducible, provided we can compute these $\gcd$s efficiently.

But how do we compute $\gcd\left(f, x^{p^d} - x\right)$ quickly? For example, large primes p might make $x^{p^d} - x$ quite large. Well, the idea is to work in $\mathbb{Z}[x]/(f)$, and then we are able to evaluate

$$x^{p^d} \pmod{f(x)}$$

via modular exponentiation by repeated squaring! So this reduces the computation of $\gcd\left(f, x^{p^d} - x\right)$ down to a $\gcd$ where both terms have degree at most $\deg f$, which is about the best we could hope for. From here, Euclidean division is fast enough for our purposes.

Remark 4.56 (Nir)**.** Berlekamp's algorithm is actually for factoring polynomials in $\mathbb{F}_p[x]$. In short, I am under the impression that careful choice of g is able to not just tell us what degree the irreducibles dividing into f are but also closer information about the irreducible.

4.2 October 21

I want to be a frog because of no schoolwork, no stress, no problems.

4.2.1 Noetherian Rings

Today we're talking about Noetherian rings. Here is our motivation.

Example 4.57. Fix k a field. Then $k[x]$ is a principal ideal domain because it is Euclidean. However, $k[x,y]$, is not principally generated: for example, (x,y) is not generated by one element. Similarly, $k[x,y,z]$ has (x,y,z) which needs three elements, and so on.

We might hope that $k[x_1,\dots,x_n]$ requires n elements to generate, but this is not true.

Exercise 4.58. Fix k a field. For any $n \in \mathbb{N}$, there exist ideals of $k[x,y]$ not generated by n elements.

Proof. We claim that

$$I \coloneqq \left(x^n, x^{n-1}y, \dots, xy^{n-1}, y^n\right) \subseteq k[x,y]$$

is not generated by n elements.

This is surprisingly annoying because one could imagine that some kind of massive cancellation among specially chosen polynomials might be able to do this. Anyways, we modify the proof from here. The trick is to move everything into a vector space, where we have better control. Set $\mathfrak{m} \coloneqq (x,y)$, which consists of all polynomials with vanishing constant term. As such, we see that

$$\frac{k[x,y]}{\mathfrak{m}} \cong k$$

by sending $x \mapsto 0$ and $y \mapsto 0$; indeed, the morphism $k[x,y] \to k$ is simply evaluating at $(0,0)$, giving out the constant term, so the kernel consists of $\mathfrak{m}$. The point is that $\mathfrak{m}$ is a maximal ideal because it its quotient gives a field.

Thus, we can assign $I/\mathfrak{m}I$ a $k[x,y]$-action as the quotient module, but this action vanishes on $\mathfrak{m}$ by definition on $\mathfrak{m}I$, so in fact we have an action by $k[x,y]/\mathfrak{m} \cong k$, so $I/\mathfrak{m}I$ is a k-vector space. The key claim is that

$$\dim_k I/\mathfrak{m}I \stackrel{?}{=} n+1.$$

Indeed, we see that the residue classes of $x^k y^{n-k}$ certainly span I and hence span $I/\mathfrak{m}I$, so $\dim_k I/\mathfrak{I} \le n+1$. To finish, we claim that the residue classes for $x^k y^{n-k}$ are in fact k-linearly independent. Well, suppose we have $\{a_k\}_{k=0}^n$ such that

$$f \coloneqq \sum_{k=0}^{n} a_k x^k y^{n-k} \in \mathfrak{m}I.$$

If f is nonzero, then it has degree n because each monomial has degree n. However, each nonzero element of $\mathfrak{m}I$ has degree at least $n+1$ because nonzero elements of I have degree at least n, and nonzero elements of $\mathfrak{m}$ have degree at least 1. So $f \neq 0$ would imply that $\deg f = n$ and at least $n+1$, which makes no sense.

So we have that $f = 0$ (as a polynomial in $k[x,y]$) so it follows that the $a_\bullet = 0$ identically, giving us our linear independence. Thus, the residue classes for $x^k y^{n-k}$ form a basis, so we see that

$$\dim_k I/\mathfrak{m}I = n+1.$$

To convert the result, we note that if I were generated by m elements, then we can take the residue classes of these elements in $I/\mathfrak{m}I$ to span $I/\mathfrak{m}I$ with m elements. But using the dimension here, we see that $m \ge n+1$, so I cannot be generated by fewer than $n+1$ elements. ∎

The point is that there is no absolute finite bound on ideals for $k[x,y]$, though there is the following result.

Theorem 4.59 (Hilbert basis)**.** Every ideal in some polynomial ring $k[x_1,\dots,x_n]$ is finitely generated.

Remark 4.60. Hilbert's proof of this theorem was somewhat complicated. Noether went back and provided a simpler proof.

The above sorts of rings have a name.

Definition 4.61 (Noetherian)**.** We say that a ring R is *Noetherian* if and only if all ideals are finitely generated.

This turns out to be a really nice and reasonable smallness property for R. In practice, most rings in number theory or algebraic geometry are Noetherian, so it's good enough for our purposes.

Non-Example 4.62. The ring $k[x_1, x_2, \ldots]$ with infinitely many transcendental elements, then the ideal

$$I = (x_1, x_2, \ldots)$$

is not finitely generated. Indeed, any finite set of polynomials $\{f_k\}_{k=1}^n$ must have each $f_\bullet$ have only finitely many monomials and hence only use finally many $x_\bullet$. So any linear combination of the $\{f_k\}_{k=1}^n$ will only use finitely many of the $x_\bullet$ and hence cannot fully cover I.

As a warning, we note that being finitely generated as an ideal and finitely generated as an algebra without identity are different.

Example 4.63. In $k[x, y]$, we consider the following objects generated by y.

- The ideal generated by y includes all polynomials which are a multiple of y. To generate this as a k-algebra without identity, we would need all elements of the form $x^k y$ for $k \geq 0$. To generate this as a k-vector space, we would need all elements of the form $x^k y^\ell$ for $k \geq 0$ and $\ell \geq 1$.
- The k-algebra generated by y includes $k[y]$, notably including 1 even though the ideal does not. To generate $k[y]$ as a k-vector space, we need all powers $y^\bullet$.
- The k-vector space generated by y includes elements of the form cy for $c \in k$.

4.2.2 Noetherian Grab-Bag

Noether's version of the Hilbert basis theorem is as follows.

Theorem 4.64 (Hibert basis, II)**.** If R is Noetherian, then $R[x]$ is also Noetherian.

Example 4.65. The ring $\mathbb{Z}[x]$ is Noetherian because $\mathbb{Z}$ is Noetherian. In particular, $\mathbb{Z}$ is Noetherian because it is a principal ideal domain, so all ideals are generated by a single element.

We note that, inductively, we also have the following.

Corollary 4.66. Fix k a field. Then $k[x_1, \ldots, x_n]$ is Noetherian.

Proof. Induct on n using Theorem 4.64. For $n = 0$, we see that k only has two ideals, (0) and (1). For the inductive step, we have that $k[x_1, \ldots, x_n]$ is Noetherian and note that

$$k[x_1, \ldots, x_n][x_{n+1}]$$

is Noetherian by Theorem 4.64. This is what we wanted. ∎

Here are some equivalent conditions for a ring being Noetherian; there are a few important ones to keep track of.

Proposition 4.67. The following are equivalent.

(a) R is Noetherian.

(b) Every ideal of R is finitely generated.

(c) Every nonempty set of ideals has a maximal ideal.

(d) Any increasing chain of ideals notated

$$I_1 \subseteq I_2 \subseteq \cdots$$

of R must stabilize.

We remark that "not (d)" provides a increasing chain of ideals

$$I_1 \subseteq I_2 \subseteq \cdots$$

which does not stabilize. But this may be turned into an infinite strictly ascending chain of ideals: set $n_1 \coloneqq 1$, and for each n_k, the lack of stabilization implies there is $n_{k+1} > n_k$ such that $I_{n_k} \subsetneq I_{n_{k+1}}$, so we have the infinite strictly ascending chain

$$I_1 \subsetneq I_2 \subsetneq \cdots .$$

And of course, we conversely have that an infinite strictly ascending chain violates (d) immediately.

Proof. We have the following implications.

- To start, we note that (a) and (b) are equivalent by definition of Noetherian.
- The fact that (c) and (d) are equivalent holds because the set of ideals of a ring is partially ordered set.
  - We show not (d) implies not (c). If we have an infinite, non-stabilizing chain of ideals

    $$I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq \cdots ,$$

    then we note that this chain is a nonempty set of ideals with no maximal ideal. Indeed, each I_N is not maximal because $I_N \subsetneq I_{N+1}$.
  - We show not (c) implies not (d). Suppose S is a nonempty set of ideals with no maximal ideals. We start with $I_1 \in S$, which exists because S is nonempty.
    Now, we recursively note that for each $k \in \mathbb{Z}^+$, we note that I_k is not maximal in S, so there is an ideal $I_{k+1} \in S$ such that $I_K \subsetneq I_{k+1}$. So we get a strictly increasing chain

    $$I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq \cdots ,$$

    successfully violating (d). Technically this argument uses some Axiom of choice to construct all of these ideals at once.[2]
- We next show that (b) implies (d). Well, given an a chain of ideals

  $$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots ,$$

  and because this is a chain, we see that

  $$I \coloneqq \bigcup_{k \in \mathbb{Z}^+} I_k$$

  is itself an ideal, which we can check by hand: we see I contains $0 \in I_1$ and so is nonempty; then for any $a, b \in I$ and $r, s \in R$, there is N such that $a, b \in I_N$ because our ideals are in a chain, so $ra + sb \in I_N \subseteq I$. Thus, I is an R-submodule of R and hence an ideal.

[2] There are actually reasons to care about this use of choice: in algebraic geometry, there are structures called topoi we might want to work in, which don't have an Axiom of choice.

Now, I is finitely generated because R is Noetherian (!), so set

$$I \coloneqq (a_1, \dots, a_n).$$

Each a_k lives in some I_{n_k}, so setting $N \coloneqq \max_k\{n_k\}$, we see that $a_k \in I_{n_k} \subseteq I_N$ for each k. Thus, for each $n > N$, we have

$$I_n \subseteq \bigcup_{k\in\mathbb{Z}^+} I_k = I = (a_1, \dots, a_n) \subseteq I_N \subseteq I_n,$$

so $I_n = I_N$ follows. Thus, our chain does stabilize.

- We now show that not (d) implies not (b). Indeed, suppose that I is not finitely generated, and we construct an infinite strictly ascending chain of ideals.

  Start with $a_1 \coloneqq 0 \in I$. Now, we recursively note that any finite set $\{a_k\}_{k=1}^n \subseteq I$ cannot generate I, so we can always for $n \in \mathbb{Z}^+$ find some

  $$a_{n+1} \in I \setminus (a_1, \dots, a_n).$$

  Continuing in this manner, we get a strictly ascending chain

  $$(a_1) \subsetneq (a_1, a_2) \subsetneq (a_1, a_2, a_3) \subsetneq \cdots,$$

  which contradicts (d). Indeed, this chain is strictly ascending because, for each $n \in \mathbb{Z}^+$, we see that $a_{n+1} \notin (a_1, \dots, a_n)$ implies that $(a_1, \dots, a_n) \subsetneq (a_1, \dots, a_{n+1})$. ■

Let's see an example.

Example 4.68. Fix $R = k[x_1, x_2, \dots]$ to have infinitely many variables. Then

$$(x_1, x_2, \dots)$$

is not finitely generated as discussed earlier. From the above work, we see that this gives rise to the infinite strictly ascending chain

$$(x_1) \subsetneq (x_1, x_2) \subsetneq (x_1, x_2, x_3) \subsetneq \cdots.$$

And of course, there is no maximal element among the above chain using the logic described above.

4.2.3 Artinian Rings

As an aside, we note that we can flip around the condition for Noetherian and ask for decreasing chains to stabilize.

Definition 4.69 (Artinian)**.** A ring R is called *Artinian* if and only if all descending chains of ideals stabilize.

Example 4.70. Fields are Artinian because they only have two ideals.

This is a very strong condition; some of our favorite rings are not Artinian.

Non-Example 4.71. The integers $\mathbb{Z}$ has the infinite strictly decreasing chain

$$(2) \supsetneq (4) \supsetneq (8) \supsetneq \cdots.$$

Remark 4.72. The Artinian condition is so strong that it implies the Noetherian condition.

We won't be talking about Artinian rings any more for now, but it might come up in commutative algebra.

4.2.4 Hilbert Basis Theorem

Let's jump into the Hilbert basis theorem. Recall Noether's statement.

Theorem 4.64 (Hibert basis, II)**.** If R is Noetherian, then $R[x]$ is also Noetherian.

Proof. Given an ideal $I \subseteq R[x]$, our goal is to find a finite set of generators. Well, we set

$$I_0 \coloneqq \{0\} \cup \{\text{leading coefficient of } f : f \in R[x] \text{ and } \deg f = 0\}.$$

This is simply $I \cap R[x]$ and might appear silly, but more generally, we define

$$I_k \coloneqq \{0\} \cup \{\text{leading coefficient of } f : f \in R[x] \text{ and } \deg f = k\}.$$

We have the following two observations.

- The I_k are ideals for each $k \in \mathbb{Z}^+$. By construction, they contain 0. Then if we have $a, b \in I_k$ and $r, s \in R$, then we need $ar + bs \in I_k$. If $ar + bs = 0$, then we are done. Otherwise, we can find polynomials f and g of degree k with leading coefficients a and b respectively. Then the polynomial

  $$af + bg$$

  has leading term $(ar + bs)\, x^k$, so indeed, $ar + bs \in I_k$.
- We have that $I_k \subseteq I_{k+1}$ for each $k \in \mathbb{Z}^+$. Indeed, $0 \in I_{k+1}$, and for $r \in I_k \setminus \{0\}$, we can find $f(x)$ of degree k with leading coefficient r. Then $x \cdot f(x)$ has degree $k+1$ with leading coefficient r, so $r \in I_{k+1}$.

Thus, we have an ascending chain of ideals

$$I_0 \subseteq I_1 \subseteq I_3 \subseteq \cdots,$$

and we note that R Noetherian implies that this sequence must stabilize to some I_N.

This use of the chain condition more or less tells us that we only care about $\{I_k\}_{k=1}^N$. Each I_k is finitely generated, so we fix

$$I_k = (r_{k,1}, r_{k,2}, \ldots, r_{k,n_k}),$$

and then, for each ℓ, we find polynomials $f_{k,\ell}$ with leading coefficient $r_{k,\ell}$ of degree k. (If $r_{k,\bullet} = 0$, just take $f_{k,\bullet} = 0$, though this doesn't matter.) Now we claim that I is generated by

$$S \coloneqq \bigcup_{k=0}^{N} \{f_{k,1}, f_{k,2}, \ldots, f_{n_k}\} \subseteq I,$$

which will be good enough because S is finite.

Essentially, S generates I by induction. Fix $p \in I$. If $p = 0$, then of course $p \in (S)$. Otherwise, we induct on $\deg p$; if $\deg p = 0$, then $p \in I_0$, so we can write p as an R-linear combination of the $f_{0,\bullet}$, finishing. Otherwise, $\deg p > 0$, and we have two cases.

- If $d \coloneqq \deg p \leq N$, then name the leading coefficient $r \in I_d$. In particular, we have some R-linear combination

  $$r = \sum_{\ell=1}^{n_d} a_\ell r_{k,\ell},$$

  so

  $$f \coloneqq \sum_{\ell=1}^{n_d} a_\ell f_{k,\ell} \in (S)$$

  will have leading term rx^d, matching p. Thus, $p - f$ will thus have smaller degree by cancelling out the leading term, so $p - f \in (S)$ by induction, so $p \in (S)$.

- If $d \coloneqq \deg p > N$, then again name the leading coefficient $r \in I_d$. But by the stabilization, we see that $r \in I_N$, so we have an R-linear combination

$$r = \sum_{\ell=1}^{n_N} a_\ell r_{N,\ell}.$$

But now

$$f \coloneqq \sum_{\ell=1}^{n_d} a_\ell f_{k,\ell} x^{d-N}$$

will have leading term rx^d, matching p. So again, $p - f$ has smaller degree by cancelling the leading term, implying that $p - f \in (S)$ and hence $p \in (S)$. ■

Remark 4.73 (Nir)**.** The end of this proof is essentially doing Euclidean division with many polynomials.

4.2.5 Analytic Examples

Let's see some more examples.

Example 4.74. We have the following list.

- The ring $\mathbb{C}[x]$ of polynomials is Noetherian.
- The ring of holomorphic functions on $\mathbb{C}$ is not Noetherian.
- The ring of functions which are holomorphic on the closed unit disk is Noetherian.
- The ring of functions which are holomorphic on the open unit disk is not Noetherian.
- The ring of functions which are holomorphic in some neighborhood are 0 is Noetherian.
- The ring of functions smooth at 0 is not Noetherian.
- The ring of formal power series at 0 is Noetherian.

All but the last item are contained in the previous, but smooth functions at 0 are not all represented by formal power series.

All of the rings in the above are quite similar, but being Noetherian is switching on and off. Let's do some of the explanations.

- The ring $\mathbb{C}[x]$ is Noetherian because it is principal.
- Similarly, $\mathbb{C}[[x]]$ is Euclidean and hence principal and hence Noetherian. The trick is to reverse our definition of degree. Namely, given a nonzero power series

$$a(x) \coloneqq \sum_{k=0}^{\infty} a_k x^k \in \mathbb{C}[[x]],$$

we define $|a|$ to be the least k such that $a_k \neq 0$, which exists because $a \neq 0$. Then to divide some power series $a(x) = \sum_{k=0}^{\infty} a_k x^k$ by a nonzero power series $b(x) = \sum_{k=0}^{\infty} b_k x^k$, we note that $|a| \geq |b|$ implies we can write

$$r \coloneqq a - \frac{a_{|a|}}{b_{|b|}} x^{|a|-|b|} \cdot b.$$

Here, r has the $a_{|a|}x^{|a|}$ term vanish while adding no lower-degree terms, so $r = 0$ or $|r| < |a|$. In this way, we can inductively push the degree downwards until $r = 0$ or $|r| < |b|$, which is what we need for Euclidean division.

Remark 4.75. The ring $\mathbb{C}[[x]]$ is an example of a "local" ring.

- Holomorphic functions on $\mathbb{C}$ is not Noetherian. For example, define

$$I_n \coloneqq \{f \text{ holomorphic} : f(k) = 0 \text{ for each positive integer } k \geq n\}$$

for each $n \in \mathbb{N}$. Then we can check that we have an ascending chain

$$I_0 \subseteq I_1 \subseteq I_2 \subseteq \cdots.$$

One way to see that this containment is strict is to note that, for each $n \in \mathbb{N}$,

$$f_n(z) \coloneqq \frac{\sin \pi z}{\prod_{k<n}(z-k)}$$

vanishes at positive integers at least n but returns 1 for positive integers less than n. Thus, $f_n \in I_n \setminus I_{n-1}$.

More generally, we can replace $\mathbb{N}$ with any set of complex numbers with no limit point, though we have to do some complex analysis to see this.

- Holomorphic functions on the closed unit disk are Noetherian because it is a principal ideal domain. Indeed, the main point is that a holomorphic function on the closed unit disk must only have finitely many roots. Given an ideal I, we can define

$$S \coloneqq \{z \in \mathbb{C} : f(z) = 0 \text{ for each } f \in I\},$$

where zeroes of higher order are counted with multiplicity in S. Then we set

$$f_S(z) \coloneqq \prod_{a \in S}(z-a)$$

again counting with multiplicity. Then $J \coloneqq \frac{1}{f_S} I$ is still an ideal, and it has no points upon which all functions in J vanish; it follows by waving our hands a bit[3] we can construct a function in J with no roots, so $J = (1)$, so $I = (f_S)$.

- Holomorphic functions on the open unit disk is not Noetherian for the same reason that holomorphic functions on $\mathbb{C}$ is not Noetherian. Namely, we have the infinite isolated sequence

$$a_n \coloneqq 1 - \frac{1}{n}$$

for $n > 0$ in the open disk, which can be used to construct the strictly ascending chain

$$I_n \coloneqq \{f \text{ holomorphic} : f(a_k) = 0 \text{ for each positive integer } k \geq n\}$$

for each $n > 0$.

- Holomorphic functions at 0 can be identified with formal power series

$$f(z) = \sum_{k=0}^{\infty} a_k z^k$$

such that $\limsup_{n\to\infty} \sqrt[n]{|a_n|} < \infty$. We can show that these functions are closed under addition and multiplication, so they inherit the division algorithm of $\mathbb{C}[[x]]$.

[3] I suspect this claim follows from some random complex analytic result, but I do not know what it is.

- Smooth functions at 0 is quite odd. For example, we can we can set I equal to the functions vanishing with infinite order at 0. Namely, the function e^{-1/x^2} vanishes with infinite order at 0 along $\mathbb{R}$.

  There are other ways to see that this ring is not Noetherian. For example, Noetherian rings have that all elements are the product of irreducibles, which we showed for principal ideal domains, whose proof carries over nicely here. But $f \coloneqq e^{-1/x^2}$ has

  $$f = \left(\sqrt{f}\right)^2 = \left(\sqrt[4]{f}\right)^4 = \cdots,$$

  so f is not the product of irreducibles.

So being Noetherian can be a kind of fuzzy condition to work with.

4.2.6 Noetherian Philosophy

Let's have some informal ways of thinking about the Noetherian condition before we continue.

- Rings of finite-dimensional algebraic objects tend to be Noetherian. For example, polynomials over a finite number of variables are more or less actions on a finite-dimensional subspace. However, infinitely many variables would act on infinite-dimensional space.
- Rings in analysis tend to be non-Noetherian. This is sad for analysts.
- Noetherian rings are more or less associated with zeroes of functions being nice. For example, on the closed unit disk, holomorphic functions have a finite number of zeroes. But this is not the case for all of $\mathbb{C}$ or for the open disk.

4.2.7 Hilbert's Finiteness Theorem: Set-Up

We will spend the rest of lecture discussing an application of Noetherian rings.

Here is the set-up: suppose tha a group G acts on a finite-dimensional k-vector space V. One way to study V would be to look at the ring of polynomial functions out of V. Fixing a basis $\{v_1, \dots, v_n\}$, this ring is

$$k[V] \coloneqq k[v_1^*, \dots, v_n^*],$$

where $v_\bullet^* : V \to k$ is the coordinate function for $v_\bullet$.[4]

We are interested in studying the G-invariants of V, and we can do this by studying G-invariants of $k[V]$. Namely, Now, the G-action on V induces an action on $k[V]$ by

$$(\sigma \cdot f)(v) \coloneqq f(\sigma^{-1} v),$$

where $\sigma \in G$ and $f \in k[V]$ and $v \in V$. Here, we are inverting in order to make the associative law for the group action actually behave. So we find that the G-invariants of $k[V]$ are

$$k[V]^G \coloneqq \{f \in k[V] : \sigma \cdot f = f \text{ for each } \sigma \in G\}.$$

So what we can say about $k[V]^G$? It's not too hard to see that our G-action preserves addition and multiplication in $k[V]$, so we have that $k[V]^G$ is a subring of $k[V]$. Further, the G-action preserves scalar multiplication by k, so the most structure we can give to $k[V]^G$ is in fact as a full k-algebra!

Well, what does $k[V]^G$ look like as a k-algebra? For example, is it finitely generated? This turns out to be a difficult question; we will show that the answer is yes for G a finite group and $k = \mathbb{C}$, though the answer is yes in more general contexts.

> **Remark 4.76.** Hilbert invented the notion of Noetherian in order to talk about the above result.

[4] Sure, there are other functions $V \to \mathbb{C}$, but from an algebraic perspective, these polynomial ones are more or less the only ones we can guarantee to exist when working in full generality.

Exercise 4.77. Fix $G = S_n$ acting on $V = \mathbb{C}^n$ by permuting the coordinates. Then $\mathbb{C}[V]^G$ is finitely generated as a $\mathbb{C}$-algebra.

Proof. Our coordinate ring is

$$\mathbb{C}[V] = \mathbb{C}[x_1, \dots, x_n],$$

where $x_k : V \to \mathbb{C}$ projects onto the kth coordinate. Then we can see that $(\sigma \cdot x_k)(v) = x_k(\sigma^{-1}v) = x_{\sigma k}$ because the $\sigma(k)$th coordinate gets moved to the kth coordinate. Because we are working in a polynomial ring, this extends uniquely to a full G-action on $\mathbb{C}[V]$.

Namely, G acts on $\mathbb{C}[V]$, by permuting coordinates, so $\mathbb{C}[V]^G$ consists of the symmetric polynomials. For example, the polynomials

$$e_1 \coloneqq x_1 + x_2 + \cdots + x_n, \qquad e_2 \coloneqq \sum_{k<\ell} x_k x_\ell, \dots \qquad e_m \coloneqq \sum_{k_1<k_2<\dots<k_m} x_{k_1} x_{k_2} \cdots x_{k_m}$$

are the "elementary" symmetric polynomials. It turns out that all symmetric polynomials are a polynomial of the $e_\bullet$, so

$$\mathbb{C}[V]^G = \mathbb{C}[e_1, \dots, e_n]$$

is indeed finitely generated as a $\mathbb{C}$-algebra. For example, we can write

$$x_1^2 + \cdots + x_n^2 = (x_1 + \cdots + x_n)^2 - 2\sum_{k<\ell} x_k x_\ell = e_1^2 - 2e_2.$$

Some kind of process like this works for all symmetric polynomials. ∎

Exercise 4.78. Fix $G = \langle \sigma \rangle \cong \mathbb{Z}/4\mathbb{Z}$ acting on $V = \mathbb{C}^2$ by $\sigma v \coloneqq i^{-1}v$. Then $\mathbb{C}[V]^G$ is finitely generated as a $\mathbb{C}$-algebra.

Proof. Our coordinate ring is $\mathbb{C}[x, y]$, where x projects onto the first coordinate and y onto the second. Further, we can check that our action is by

$$(\sigma \cdot x)(v) = x(\sigma^{-1}v) = x(iv) = (ix)(v),$$

so $\sigma \cdot x = ix$. Similarly, $\sigma \cdot y = iy$, and this extends uniquely to a full G-action on $\mathbb{C}[V]$ because $\mathbb{C}[V]$ is a polynomial ring.

Looking at more general monomials, we see that

$$\sigma\left(x^k y^\ell\right) = i^{k+\ell} x^k y^\ell,$$

so the fixed monomials are the ones with $i^{k+\ell} = 1$, which is equivalent to having degree divisible by 4. Now, if a polynomial

$$f(x, y) \coloneqq \sum_{k,\ell \in \mathbb{N}} a_{k,\ell} x^k y^\ell$$

is fixed by the G-action, then we see that

$$(\sigma \cdot f)(x, y) = \sum_{k,\ell \in \mathbb{N}} i^{k+\ell} a_{k,\ell} x^k y^\ell.$$

In particular, G preserves the monomials themselves, so we need $i^{k+\ell} = 1$ for each $a_{k,\ell} = 0$. Thus, we have that $\mathbb{C}[V]^G$ consists of the polynomials all of whose monomials have degree divisible by 4. In other words,

$$\mathbb{C}[V]^G = \mathbb{C}\left[\left\{x^k y^\ell : k + \ell \equiv 0 \pmod 4\right\}\right].$$

For example, this is infinite-dimensional as a $\mathbb{C}$-vector space, spanned by the $x^k y^\ell$ with $k+\ell \equiv 0 \pmod 4$.

However, for each $x^k y^\ell$ with $k+\ell \equiv 0 \pmod 4$, we can write $k \equiv k' \pmod 4$ and $\ell \equiv \ell' \pmod 4$ with $0 \le k', \ell' < 4$, so $0 \le k'+\ell' < 8$. The point is that we can write

$$x^k y^\ell = \left(x^4\right)^{(k-k')/4} x^{k'} \cdot \left(y^4\right)^{(\ell-\ell')} y^{\ell'}.$$

But now because $k'+\ell' < 8$ while $k'+\ell' \equiv 0 \pmod 4$, we see that $x^{k'}y^{\ell'} \in \left\{x^0y^0, x^1y^3, x^2y^2, x^3y^1\right\}$. It follows that we can fit all monomials of $\mathbb{C}[V]^G$ into

$$\mathbb{C}\left[y^4, xy^3, x^2y^2, x^3y, x^4\right],$$

where we have thrown out $x^0y^0 = 1$ and added the needed x^4 and y^4. It follows that we can write $\mathbb{C}[V]^G = \mathbb{C}\left[x^4, x^3y, x^2y^2, xy^3, y^4\right]$, so $\mathbb{C}[V]^G$ is indeed finitely generated as a $\mathbb{C}$-algebra. ■

In general, rings of invariants are quite complicated, and even if finitely generated, they might require lots of generators.

Example 4.79. Let $G = \mathrm{SL}_2(\mathbb{C})$ act on the binary quantics, which are polynomials of the form

$$\sum_{k=0}^{n} a_k x^k y^{n-k}.$$

Namely, these are all two-variable polynomials which are homogeneous of degree n, and the G-action is by multiplication multiplication like

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax+by \\ cx+dy \end{bmatrix}.$$

It turns out that the ring of G-invariants are finitely generated, which is due to Gordon.

4.2.8 Hilbert's Finiteness Theorem: Proof

Proving that these invariants are finitely generated was an incredibly hard problem, but Hilbert presented a disturbingly simple proof of it.

Theorem 4.80 (Hilbert's finiteness)**.** Fix G a finite group and V a finite-dimensional $\mathbb{C}$-vector space with a (linear) G-action. Then $\mathbb{C}[V]^G$ is finitely generated as a $\mathbb{C}$-algebra.

Proof. The key trick is that, when G is a finite group, we have a "Reynolds" operator, which is essentially a "G-average." Namely, for some function $f \in \mathbb{C}[V]$, we define

$$\rho(f) \coloneqq \frac{1}{\#G} \sum_{\sigma \in G} \sigma(f) \in \mathbb{C}[V].$$

The division by $\#G$ is legal in $\mathbb{C}$ but not in all fields because we might end up dividing by the characteristic.

Remark 4.81. This is the same Reynolds who did fluid dynamics, who used the Reynolds operator to average fluid flow over time operators.

Now, starting with $\mathbb{C}[V]$, let $I = \mathbb{C}[V]^G$ be the ring of invariants, which we grade by degree. Explicitly, we have

$$I = I_0 \oplus I_1 \oplus I_2 \oplus I_3 \oplus \cdots,$$

where I_k has terms of degree k. (Indeed, any polynomial in $\mathbb{C}[V]$ can be uniquely decomposed into polynomials of various fixed degrees.) Then we set

$$J \coloneqq (I_1, I_2, \ldots) \subseteq \mathbb{C}[V]$$

to be the ideal generated by the homogeneous polynomials in I of positive degree, where we exclude constants to avoid the full ring.

But $\mathbb{C}[V]$ is simply a polynomial ring and hence Noetherian! So J is generated by some finite number of elements (as an ideal) in $\mathbb{C}[V]$. Further, each of these finitely many generating elements can be written as some finite $\mathbb{C}[V]$-linear combination of nonconstant homogeneous polynomials in I (by construction of J), so in fact we may write

$$J = (j_1, \ldots, j_m),$$

where the $j_\bullet$ are nonconstant homogeneous G-invariant polynomials. We would like to show that these $j_\bullet$ generate I as a $\mathbb{C}$-algebra; in other words, we claim that

$$I \stackrel{?}{=} \mathbb{C}[j_1, \ldots, j_m],$$

which will indeed finitely generate I as a $\mathbb{C}$-algebra.

Remark 4.82. Generating as an ideal and generating as an algebra are again, quite different, and this is where the main difficulty is in the proof. For instance, the ideal $(y) \subseteq \mathbb{C}[x, y]$ is not finitely generated as a $\mathbb{C}$-algebra.

So we need to use some property that I is a ring of invariants, and as promised, we will use the Reynolds operator. Due to the grading, it suffices to show that $I_n \subseteq \mathbb{C}[j_1, \ldots, j_m]$ for each $n \in \mathbb{N}$. We show this by induction; note that I_0 consists of constants in $\mathbb{C}$, which are in $\mathbb{C}[j_1, \ldots, j_m]$ automatically.

Otherwise, pick some $b \in I_n$ with $n > 0$, and we need to show $b \in \mathbb{C}[j_1, \ldots, j_m]$. Because $b \in J$, and $J = (j_1, \ldots, j_m)$, we may write

$$b = \sum_{k=1}^{m} c_k j_k.$$

Without loss of generality, we may take $\deg c_\bullet = \deg b - \deg j_\bullet$ because any terms of $c_\bullet$ which are not of this degree will have to cancel out somewhere else because b and $j_\bullet$ are homogeneous. We repeat again that the issue here is that the $c_\bullet$ live in $\mathbb{C}[V]$, not in $\mathbb{C}[V]^G$ the ring of invariants.

But the key trick is to apply the Reynolds operator! The Reynolds operator has the following magical properties.

- If $f \in \mathbb{C}[V]^G$ and $g \in \mathbb{C}[V]$, then $\rho(fg) = f\rho(g)$. Indeed,

$$\rho(fg) = \frac{1}{\#G} \sum_{\sigma \in G} \sigma(fg) = \frac{1}{\#G} \sum_{\sigma \in G} \sigma(f)\sigma(g) = f \cdot \frac{1}{\#G} \sum_{\sigma \in G} \sigma(f) = f\rho(g).$$

- In particular, if $f \in \mathbb{C}[V]^G$, then $\rho(f) = \rho(f \cdot 1) = f\rho(1) = f$.
- If $f \in \mathbb{C}[V]$, then $\rho(f) \in \mathbb{C}[V]^G$. Indeed, for any $\tau \in G$, we have that

$$\tau \cdot \rho(f) = \tau \cdot \frac{1}{\#G} \sum_{\sigma \in G} \sigma(f) = \frac{1}{\#G} \sum_{\tau\sigma \in G} (\tau\sigma)(f) = \rho(f).$$

Thus, we can write

$$b = \rho(b) = \sum_{k=1}^{m} j_k \rho(c_k),$$

which essentially finishes immediately. Indeed, the $\rho(c_\bullet)$ are elements of I of degree smaller than $\deg b$, so they live in $\mathbb{C}[j_1, \ldots, j_m]$, so we finish by induction. ∎

Remark 4.83. This proof is quite amazing. It is little more than applying the Reynolds operator, wiping out hundreds of pages unreadable invariant theory because explicitly writing out the invariants can be truly terrible.

We close with some more remarks.

Remark 4.84. Hilbert's proof is not constructive: we don't have an algorithm to actually find the generators from an ideal in the above proof. He would later provide a more explicit construction, which comes from making the Hilbert basis theorem more constructive.

Remark 4.85. We don't need the full force of G finite. For example, G compact lets us integrate for our Reynolds operator using a Haar measure, which is safe. We can also do $G = \mathrm{SL}_2(\mathbb{C})$ by the "unitarian trick," where we observe that $\mathrm{SL}_2(\mathbb{C})$ (which is not compact) contains the compact group $\mathrm{SU}_2(\mathbb{C})$, and then it turns out that the $\mathrm{SL}_2(\mathbb{C})$-invariants are the same as the $\mathrm{SU}_2(\mathbb{C})$-invariants.

Remark 4.86. Nagata found a group G where the invariants are not finitely generated, so this statement is not true for all groups G. (This provided a negative answer to Hilbert's 14th problem.) So we do need some smallness condition on G; the correct property for the above proof turns out to be "linearly reductive."

Remark 4.87. The fact we are working over $\mathbb{C}$ is also unnecessary as shown by Haboush, though it makes the proof more difficult.

4.3 October 26

I am okay with being stabbed.

4.3.1 Symmetric Polynomials

We're talking about symmetric polynomials and resultants today. Recall that variables $\{\alpha_k\}_{k=1}^n$ has an S_n-action by permuting the indices, and a symmetric function on these variables are the S_n-invariants.

Example 4.88. The function

$$e_1 \coloneqq \alpha_1 + \cdots + \alpha_n \qquad \text{and} \qquad e_2 \coloneqq \sum_{k<\ell} \alpha_k \alpha_\ell$$

are elementary symmetric polynomials. More generally, we have

$$e_m = \sum_{k_1 < \cdots < k_m} \alpha_{k_1} \cdots \alpha_{k_m}$$

We can combine these using Vieta's formulae to get

$$\prod (x - \alpha_k) = x^n - e_1 x^{n-1} + e_2 x^{n-2} + \cdots .$$

Our goal for today is the following.

Theorem 4.89. Fix k a field. Any symmetric function in $k[\alpha_1, \ldots, \alpha_n]$ lives in $k[e_1, \ldots, e_n]$.

Proof. The key idea is to order monomials and kill off the biggest, essentially forcing an induction to exist. Ordering can essentially be done in any reasonable way; Gröbner bases are essentially the standard way to induce an ordering, but we will use lexicographic ordering: we order two monomials by

$$\alpha_1^{p_1} \cdots \alpha_n^{p_n} > \alpha_1^{q_1} \cdots \alpha_n^{q_n}$$

if and only if the least i for which $a_i \neq b_i$ has $a_i > b_i$. For example, $\alpha_1^2 > \alpha_1\alpha_2^4$ and $\alpha_1\alpha_2^2\alpha_3^3 > \alpha_1\alpha_2^2\alpha_3$. We can check that this is a total ordering because the least k for which $p_i \neq q_i$ exists whenever the monomials are unequal, and at this point either $p_i > q_i$ or $p_i < q_i$.

Notation for multivariate polynomials is quite annoying, so we will write, for $v \coloneqq (v_1, \dots, v_n) \in \mathbb{N}^n$,

$$\alpha^v \coloneqq \alpha_1^{v_1} \cdots \alpha_n^{v_n}.$$

In particular, we can notate a particular $f \in k[\alpha_1, \dots, \alpha_n]$ by

$$f = \sum_{v \in \mathbb{N}^n} c_v \alpha^v,$$

which lets us avoid having to drown in indices.

Warning 4.90 (Nir)**.** This notation is not standard.

Under this notation, we note that the lexicographic ordering of monomials is the same as the lexicographic ordering of $\mathbb{N}^n$.

Before continuing, we note that the leading monomial is multiplicative.

Lemma 4.91. The leading monomial (using the lexicographic ordering) is multiplicative: if f has leading monomial α^p while g has leading monomial α^q, then fg has leading monomial the product α^{p+q}.

Proof. This statement comes from the fact the lexicographic is "additive." In particular, fix $p \geq v$ and $q \geq w$ for $v, w \in \mathbb{N}^n$, where we are using the lexicographic ordering on $\mathbb{N}^n$. Then we claim that

$$p + q \overset{?}{\geq} v + w,$$

with equality if and only if $p = v$ and $q = w$. Note that if $v = p$ or $w = q$, then we can cancel the equal term from both sides to get the statement we want.

Otherwise, $p > v$ and $q > w$, and we want to show that $p + q > v + w$. We know the least index i for which $p_i \neq v_i$ has $p_i > v_i$, and the least index j for which $q_j \neq w_j$ has $q_j > v_j$. But now we see that each index before $\min\{i, j\}$ has $p_\bullet = v_\bullet$ and $q_\bullet = w_\bullet$ so that $p_\bullet + v_\bullet = q_\bullet + w_\bullet$. But then

$$p_{\min\{i,j\}} + q_{\min\{i,j\}} > v_{\min\{i,j\}} + w_{\min\{i,j\}},$$

so we do find that $p + q > v + w$.

For concreteness, write

$$f = \sum_{v \in \mathbb{N}^n} c_v \alpha^v \qquad \text{and} \qquad g = \sum_{w \in \mathbb{N}^n} d_w \alpha^w.$$

Then, distributing, we see that

$$fg = \sum_{x \in \mathbb{N}^n} \left(\sum_{v+w=x} c_v d_w \right) \alpha^x.$$

We claim that the leading monomial of fg is from α^{p+q}. Indeed, we note that if $c_v d_w \neq 0$ so that $c_v \neq 0$ and $d_w \neq 0$, then $v \leq p$ and $w \leq q$. So by the above, it follows that

$$v + w \leq p + q$$

with equality if and only if $v = p$ and $w = q$.

To finish, we see that this implies the only x with any nonzero term $c_v d_w$ with $v + w = x$ will have to be $x \leq p + q$, and in fact $p + q$ exactly only has $v = p$ and $w = q$ so that the coefficient comes out to $c_v d_w \neq 0$. So it follows that α^{p+q} is the largest monomial with nonzero coefficient in fg, which finishes. ∎

We now attack the proof directly. Suppose that $f \in k[\alpha_1, \dots, \alpha_n]$ is a symmetric function, and use the lexicographic ordering to find its largest monomial $\alpha_1^{p_1} \cdots \alpha_n^{p_n}$ with nonzero coefficient. We claim that

$$p_1 \overset{?}{\geq} p_2 \overset{?}{\geq} \cdots \overset{?}{\geq} p_n.$$

Indeed, if $k < \ell$ has $p_k < p_\ell$, then we note that the monomial

$$(k, \ell) \cdot \alpha_1^{p_1} \cdots \alpha_{k-1}^{p_{k-1}} \alpha_k^{p_k} \alpha_{k+1}^{p_{k+1}} \cdots \alpha_{\ell-1}^{p_{\ell-1}} \alpha_\ell^{p_\ell} \alpha_{\ell+1}^{p_{\ell+1}} \cdots \alpha_n^{p_n} = \alpha_1^{p_1} \cdots \alpha_{k-1}^{p_{k-1}} \alpha_k^{p_\ell} \alpha_{k+1}^{p_{k+1}} \cdots \alpha_{\ell-1}^{p_{\ell-1}} \alpha_\ell^{p_k} \alpha_{\ell+1}^{p_{\ell+1}} \cdots x_n^{a_n}$$

will have nonzero coefficient in f because f is symmetric. But the above monomial has exponents equal to $\alpha_1^{p_1} \cdots \alpha_n^{p_n}$ up until $\alpha_k^{p_k}$, at which point we see that the above monomial is strictly greater than our hypothesized largest monomial $\alpha_1^{p_1} \cdots \alpha_n^{p_n}$ because $p_\ell > p_k$.

Now, the main idea in the proof is to repeatedly kill off the leading term of f. Indeed, we note that

$$s \coloneqq e_1^{p_1 - p_2} e_2^{p_2 - p_3} \cdots e_{n-1}^{p_{n-1} - p_n} e_n^{p_n}$$

is going to have the desired leading term: note that the leading term of e_i is $\alpha_1 \cdots \alpha_i$ by using the lexicographic ordering, but then multiplicativity of the leading coefficient tells us that the leading term of the s

$$\alpha_1^{p_1 - p_2} (\alpha_1 \alpha_2)^{p_2 - p_3} \cdots (\alpha_1 \cdots \alpha_{n-1})^{p_{n-1} - p_n} (\alpha_1 \cdots \alpha_n)^{p_n} = \alpha^p.$$

Now we see that $f - c_p s$ will kill the leading term of f, so we can more or less induct downwards to eventually kill all of f.

The details of the induction are actually quite annoying because we did not well-order the monomials: there are infinitely many monomials smaller than x_1^2, e.g. of the form $x_1 x_2^\bullet$. But an induction still works: the process above will give a sequence of strictly decreasing monomials (which are the leading terms of the polynomial we're trying to inductively kill), and we need to show that this sequence must eventually take f down to the least monomial $\alpha_1^0 \cdots \alpha_n^0$, which is immediately in $k[e_1, \dots, e_n]$.

So we need to show that there are no infinite strictly descending chains in the lexicographic ordering of $\mathbb{N}^n$.

Lemma 4.92. All descending chains in the lexicographic ordering of $\mathbb{N}^n$ must stabilize.

Proof. We proceed by induction on n. For $n = 1$, this is the assertion that $\mathbb{N}$ is well-ordered. Otherwise, suppose that there are no infinite strictly descending chains in $\mathbb{N}^n$ must stabilize and fix an infinite descending chain

$$v_1 \geq v_2 \geq v_3 \geq \cdots$$

in $\mathbb{N}^{n+1}$. Projecting onto the first n coordinates, we see that the first n coordinates must eventually stabilize. There is an N_1 for which $m > N_1$ has

$$(v_m)_k = (v_N)_k$$

for each $1 \leq k \leq n$. But now the last coordinate past N_1 is a descending sequence in $\mathbb{N}$ because the lexicographic ordering now has $v_m \geq v_{m+1}$ requires

$$(v_m)_{n+1} \geq (v_{m+1})_{n+1}.$$

As this is a descending sequence in $\mathbb{N}$, it must also eventually stabilize and hence must also stabilize past some N_2, so the full $v_\bullet$ stabilize past N_2. ∎

The above lemma finishes the proof. ∎

Remark 4.93. Professor Borcherds is doing a lot of killing in this proof.

Remark 4.94. We used the fact that f is symmetric when we write down $p_1 \geq p_2 \geq p_3 \geq \cdots \geq p_n$, which is necessary to get the symmetric polynomials to kill off the leading monomial. Amazingly, this is the only place we used the symmetric condition.

Remark 4.95 (Nir)**.** We can avoid the annoyance at the end of the proof about infinite descending chains by using a better-behaved ordering than the lexicographic one. For example, if p_i is the ith prime, then we can weight by

$$w\left(\alpha^{(v_1,\ldots,v_n)}\right) := \sum_{i=1}^{n} v_i \sqrt{p_i},$$

and it is not too hard to check that this ordering is multiplicative and again has the additive property we need. But now this ordering is in fact a well-order, so the induction is more free.

4.3.2 Newton's Sums

The algorithm suggested in Theorem 4.89 is constructive but somewhat annoying to use because its run-time is ineffective at the end. Using a better-behaved order as described in Remark 4.95 does make the run-time effective, but the run-time is still very bad because of the large numbers of symmetric polynomial computations.

As a specific case we might care about, consider the symmetric polynomial

$$p_m := \alpha_1^m + \cdots + \alpha_n^m \in k[\alpha_1, \ldots, \alpha_n].$$

Here are some small examples.

Example 4.96. We see that $p_0 = n$ and $p_1 = e_1$. Further,

$$p_2 = \sum_{i=1}^{n} \alpha_i^2 = \left(\sum_{i=1}^{n} \alpha_i\right)^2 - 2\sum_{i>j} \alpha_i \alpha_j = e_1^2 - 2e_1.$$

More generally, we have Newton's sums.

Exercise 4.97 (Newton's sums)**.** Fix k a field and work in $k[\alpha_1, \ldots, \alpha_n]$. Fix $p_m := \alpha_1^m + \cdots + \alpha_n^m$. We write p_m explicitly as in $k[e_1, \ldots, e_n]$.

Proof. By Vieta's formulae, we note that we can fully expand

$$f(x) := \prod_{i=1}^{n}(x - \alpha_i) = \sum_{d=0}^{n}\left(\sum_{\substack{S \subseteq \{1,\ldots,n\} \\ \#S = n-d}} \prod_{i \in S} -\alpha_i\right) x^d = \sum_{d=0}^{n}(-1)^{n-d} e_d x^d = x^n - e_1 x^{n-1} + \cdots$$

so that we want the mth powers of the roots of f. Here we have taken $e_0 = 1$ by convention.

The key trick is to take the logarithmic derivative of both sides of

$$f(x) = \prod_{i=1}^{n}(x - \alpha_i).$$

This might appear unmotivated, but it will greatly simplify things.

Remark 4.98. Logarithmic derivatives are good tools whenever we have products. Namely, logarithms turn bad products into awkward sums involving the logarithm, but then the logarithm gets rid of the logarithm.s.

On one hand, we see that

$$f'(x) = \sum_{d=1}^{n} d(-1)^{n-d} e_d x^{d-1} = nx^{n-1} - (n-1)e_1 x^{n-2} + \cdots,$$

and this derivative is a purely algebraic operation, definable over any field. In particular, we can check the following.

Proposition 4.99. Fix R a commutative ring. Given $f(x) \in R[x]$ represented by $f(x) = \sum_{k=0}^{n} a_k x^k$, we formally define

$$f'(x) = \frac{df}{dx} \coloneqq \sum_{k=1}^{n} k a_k x^{k-1}.$$

Then, for $f, g \in R[x]$ and $a, b \in R$, we have the following.

- $(af + bg)' = af' + bg'$.
- $(fg)' = fg' + f'g$.
- $(f \circ g)'(x) = f'(g(x))g'(x)$.

Proof. The first two are doable by direct force. For the chain rule, one should first show this for $f(x) = x^n$ by inducting on the multiplication rule and then extend linearly to all polynomials f. We will not show the details here because I am lazy. ■

Working in the quotient field $k(\alpha_1, \ldots, \alpha_n)$, we can verify by hand that we do have a "formal" logarithmic differentiation

$$\frac{(fg)'}{fg} = \frac{fg' + f'g}{fg} = \frac{f'}{f} + \frac{g'}{g}$$

by using the product rule in the numerator. So indeed, the product becomes a sum.

Remark 4.100. Importantly this logarithmic derivative rule does not require us to formally define a logarithm over arbitrary fields. Professor Borcherds in office hours said that, if we wanted to formally add a logarithm, one thing we could do was formally adjoin a function with derivative $\frac{1}{x}$ to $k[x]$. Apparently this is is somewhat standard practice in the field of differential Galois theory.

Using our logarithmic differentiation, we see that

$$\frac{f'(x)}{f(x)} = \sum_{i=1}^{n} \frac{1}{x - \alpha_i}.$$

But we notice that, now upgrading once more to $k((\alpha_1, \ldots, \alpha_n))$, we have that

$$\frac{1}{x - \alpha_i} = \frac{x^{-1}}{1 - x^{-1}\alpha_i} = \sum_{m=0}^{\infty} x^{-m-1} \alpha_i^m$$

by using the geometric series formula, so

$$\frac{f'(x)}{f(x)} = \sum_{i=1}^{n} \sum_{m=0}^{\infty} x^{-m-1} \alpha_i^m = \sum_{m=0}^{\infty} x^{-m-1} \left(\sum_{i=1}^{n} \alpha_i^m \right) = \sum_{m=0}^{\infty} p_m x^{-m-1}$$

after exchanging the order of summation. Now we can multiply both sides by $f(x)$ and equate coefficients. Written out in summation notation, this reads as

$$\sum_{d=1}^{n} d(-1)^{n-d} e_d x^{d-1} = \left(\sum_{d=0}^{n} e_d x^d\right)\left(\sum_{m=0}^{\infty} p_m x^{-m-1}\right),$$

but this is a bit easier to read as

$$nx^{n-1} - (n-1)e_1 x^{n-2} + (n-2)e_2 x^{n-3} - \cdots = \\ \left(x^n - e_1 x^{n-1} + e_2 x^{n-2} + \cdots\right)\left(p_0 x^{-1} + p_1 x^{-2} + p_2 x^{-3}\cdots\right).$$

In particular, equating coefficients shows us that

monomial	coefficients
x^{n-1}	$n = p_0$
x^{n-2}	$-(n-1)e_1 = p_1 - e_1 p_0$
x^{n-3}	$(n-2)e_2 = p_2 - e_1 p_1 + p_0 e_2$
x^{n-3}	$-(n-3)e_3 = p_3 - e_1 p_2 + e_2 p_1 - e_3 p_0$

This continues down in a recursion style. The general statement is that the coefficient of $x^{n-(d+1)}$ looks like

$$p_d - e_1 p_{d-1} + e_2 p_{d-2} - \cdots + (-1)^d e_d p_0 = \sum_{i=1}^{d} (-1)^i e_i p_{d-i} = \begin{cases} (-1)^d (n-d) e_d & d \le n \\ 0 & \text{else.} \end{cases}.$$

Using the fact that $p_0 = n$, we can move things around to get

$$\boxed{p_d = (-1)^{d+1} d e_d + \sum_{i=1}^{d-1} (-1)^{i+1} e_i p_{d-i}},$$

under the strong assumption that I have not made an error moving things around. ■

Example 4.101. Let's find the sum of the fifth powers of the roots of $x^3 + x + 1$. We have the following computations; note $e_1 = 0$ and $e_2 = 1$ and $e_3 = -1$ and $e_i = 0$ for $i > 3$.

- We have that $p_0 = 3$.
- We have that $p_1 = e_1 = 0$.
- We have that $p_2 = -2e_2 + e_1 p_1 = -2$.
- We have that $p_3 = 3e_3 + (e_1 p_2 - e_2 p_1) = -3$.
- We have that $p_4 = -4e_4 + (e_1 p_3 - e_2 p_2 + e_3 p_1) = 0 + (0 - 1 \cdot -2 + 0) = 2$.
- We have that $p_5 = 5e_5 + (e_1 p_4 - e_2 p_3 + e_3 p_2 - e_4 p_1) = 0 + (0 - 1 \cdot -3 + -1 \cdot -2) = \boxed{5}$.

Remark 4.102. Professor Borcherds is not sure why you would need the sum of the fifth powers.

4.3.3 Adams Operations

Let's talk about Adams operations in representation theory/K-theory for a little bit. Here we take V to be a finite-dimensional k-vector space with a linear G-action. Then we see that G also acts on $V \otimes V$ as well by

$$g \cdot (v_1 \otimes v_2) \coloneqq (g \cdot v_1) \otimes (g \cdot v_w). \tag{$*$}$$

Indeed, we can see that $\mu_g : V \times V \to V \otimes V$ defined componentwise is a bilinear map by writing down

$$\begin{aligned}\mu_g(v, b_1w_1 + b_2w_2) &= (gv) \otimes \big(b_1(gw_1) + b_2(gw_2)\big) \\ &= b_1 \cdot \big((gv) \otimes (gw_1)\big) + b_2 \cdot \big((gv) \otimes (gw_2)\big) \\ &= b_1\mu_g(v, w_1) + b_2\mu_g(v, w_2)\end{aligned}$$

and similar for the other side. So we do indeed have a linear map $\mu_g : V \otimes V \to V \otimes V$ defined by extending $(*)$ linearly.

Anyways, we see that the G-action on $V \otimes V$ in fact splits into

$$S^2(V) \oplus \Lambda^2(V),$$

where $S^2(V)$ is the symmetric part generated by $a \otimes b + b \otimes a$, and $\Lambda^2(V)$ is the antisymmetric part generated by $a \otimes b - b \otimes a$. In reality, what is happening is that the G-action on $V \otimes V$ more or less induces a $G \times S_2$-action on $V \otimes V$, where the S_2 swaps the two coordinates; this is legal because G acts on one coordinate at a time.[5]

Remark 4.103. We could generalize this to tensoring n copies of V to get a $G \times S_n$-action. Then the action on $V^{\otimes n}$ will split into something more complicated depending on some decomposition of S_n.

We now have the following definition.

Definition 4.104 (Adams operation)**.** Fix everything as above. We define the Adams operations in the ring of G-representations as

$$\psi^2(V) \coloneqq S^2(V) - \Lambda^2(V).$$

Yes, we can subtract by working in the "formal" representation ring. For here, what we need to know about the representation ring is that the addition is the direct sum $\oplus$.

We might be interested in how G acts on $\psi^2(V)$, which we can talk about via the trace. Of course, the trace on these formal representations might be poorly defined, but never fear—the trace is additive on the direct sum $\oplus$, so we can just subtract formally! Namely, if $T \in \mathrm{GL}(V)$ and $S \in \mathrm{GL}(W)$, then $T \oplus S \in \mathrm{GL}(V \oplus W)$ has

$$\operatorname{tr}(T \oplus S) = \operatorname{tr} T + \operatorname{tr} S.$$

For example, we can see this by writing everything out as matrices so that the matrix of representation of $T \oplus S$ looks like

$$T \oplus S = \begin{bmatrix} T & 0 \\ 0 & S \end{bmatrix},$$

which makes the diagonal sum indeed $\operatorname{tr} T + \operatorname{tr} S$.

Now, for concreteness, suppose that each $\mu_g \in \mathrm{GL}(V)$ is diagonalizable with eigenvalues $\alpha_1, \dots, \alpha_n$ and eigenbasis $\{v_1, \dots, v_n\}$. We now check the eigenvalues of μ_g for the action on each $V \otimes V$ and $S^2(V)$ and $\Lambda^2(V)$.

- On $V \otimes V$, we have a basis given by $v_i \otimes v_j$ for each i, j, so our eigenvalues on this eigenbasis look like $\alpha_i\alpha_j$ for each i, j because

$$\mu_g(v_i \otimes v_j) = (gv_i) \otimes (gv_j) = (\alpha_i v_i) \otimes (\alpha_j v_j) = (\alpha_i\alpha_j)(v_i \otimes v_j).$$

- On $S^2(V)$, we have a basis given by $v_i \otimes v_j + v_j \otimes v_i$ for each $i \geq j$. These elements do indeed span $S^2(V)$: for any $a, b \in V$, we can write

$$a = \sum_{i=1}^{n} a_i v_i \qquad \text{and} \qquad b = \sum_{i=1}^{n} b_i v_i$$

[5] I don't really feel like being more explicit about the decomposition of this representation, but one should also check that these spaces are G-invariant and orthogonal, which is not too hard to do.

so that

$$a \otimes b + b \otimes a = \sum_{i,j=1}^{n} (a_i b_j)(v_i \otimes v_j + v_j \otimes v_i),$$

so the basis hits all elements of $a \otimes b + b \otimes a$. Thus, we have $\dim_k S^2(V) \le \binom{n}{2} + n = \frac{n^2+n}{2}$, and we will check next that $\dim_k \Lambda^2(V) \le \frac{n^2-n}{2}$, so the fact that $\dim_k S^2(V) + \dim_k \Lambda^2(V) = \dim V \otimes V$ means that we must have equalities.

Anyways, we see that our eigenvalue for some $v_i \otimes v_j + v_j \otimes v_i$ upon applying μ_g becomes $\alpha_i \alpha_j$ again.

- Similarly, on $\Lambda^2(V)$, we have a basis given by $v_i \otimes v_j - v_j \otimes v_i$ for each $i > j$. Again, we can check that these span: for any $a, b \in V$, we can write

$$a = \sum_{i=1}^{n} a_i v_i \qquad \text{and} \qquad b = \sum_{i=1}^{n} b_i v_i$$

so that

$$a \otimes b - b \otimes a = \sum_{i,j=1}^{n} (a_i b_j)(v_i \otimes v_j - v_j \otimes v_i).$$

Now we can get to the restricted basis by noting that $i = j$ causes the term to collapse, so we don't care about those elements; and if $i < j$, then we can add a sign to make the basis element $v_j \otimes v_i - v_i \otimes v_j$ instead. So we get that $\dim_k \Lambda^2(V) \le \binom{n}{2} = \frac{n^2-n}{2}$, filling in the above argument.

Anyways, we see that our eigenvalue for some $v_i \otimes v_j - v_j \otimes v_i$ upon applying μ_g becomes $\alpha_i \alpha_j$ again.

In particular, the eigenvalues on $\psi^2(V)$ for μ_g sum to give trace

$$\operatorname{tr} \mu_g = \underbrace{\sum_{\substack{i,j=1 \\ i \ge j}}^{n} \alpha_i \alpha_j}_{S^2(V)} - \underbrace{\sum_{\substack{i,j=1 \\ i > j}}^{n} \alpha_i \alpha_j}_{\Lambda^2(V)} = \sum_{i=1}^{n} \alpha_i^2 = p_2,$$

which is sufficiently cute. More generally, we can define $\psi^m(V)$ to have trace p_m, more or less using Newton's identities in the sums directly.

4.3.4 Alternating Polynomials

Thus far we have studied invariants of S_n on $k[x_1, \dots, x_n]$. We might be interested in A_n-invariants.

Example 4.105. On $A_3 = \langle \sigma \rangle \cong \mathbb{Z}/3\mathbb{Z}$, we have the A_3-action on $k[x_1, x_2, x_3]$ determined by three-cycle generated by σ taking (say) $x_1 \to x_2 \to x_3 \to x_1$ in a cycle. Of course, the symmetric functions are invariants, but there are more: the function

$$(x_1 - x_2)(x_2 - x_3)(x_3 - x_1)$$

is A_3-invariant. It turns out that this is essentially the only other alternating polynomial, however.

Let's see this.

Exercise 4.106. We describe the A_n-invariants of $k[x_1, \dots, x_n]$, where k has characteristic not 2.

Proof. We briefly remark that if $f \in k[x_1, \dots, x_n]$ is A_n-invariant, then for any two odd permutations σ_1 and σ_2, we have that

$$\sigma_1 f = \sigma_2 \sigma_2^{-1} \sigma_1 f = \sigma_2 f$$

because $\sigma_2^{-1}\sigma_1 \in A_n$. The point is that the S_n-action on f is completely determined by what a single odd permutation does.

Now, we note that if some $\tau \in A_n$ swaps x_i and x_j, and if f is invariant on A_3, then we note we can write

$$f = \frac{f+\tau f}{2} + \frac{f-\tau f}{2}.$$

What is happening here is that the first term is S_n-invariant, and the second term is "antinvariant," meaning that it changes sign on odd permutations. We can check that $\frac{f+\tau f}{2}$ is actually S_n-invariant because any even permutation keeps the terms in place, and an odd permutation will swap them.

Formally, we have the following.

Definition 4.107 (Antinvariant)**.** We say that $f \in k[x_1, \dots, x_n]$ is *antinvariant* if and only if, for each $\sigma \in S_n$, we have that $\sigma f = (\operatorname{sgn}\sigma) f$.

So to check that $\frac{f-\tau f}{2}$ is antinvariant, we have the following checks.

- If σ is even, then $\sigma f = f$ so that

$$\sigma \cdot \frac{f-\tau f}{2} = \frac{\sigma f - (\sigma\tau)f}{2} = \frac{f-\tau f}{2}$$

  because τ and $\sigma\tau$ are both odd.

- If σ is odd, we see that $\sigma\tau$ is even implies that

$$\sigma \cdot \frac{f-\tau f}{2} = \frac{\sigma f - (\sigma\tau)f}{2} = \frac{\tau f - f}{2}$$

  because σ and τ are both odd.

So indeed, $\frac{f-\tau f}{2}$ is antinvariant.

The point of these computations is that we can write

$$k[x_1, \dots, x_n]^{A_n} = k[x_1, \dots, x_n]^{S_n} + \text{antinvariant polynomials}.$$

In fact, an element $f \in k[x_1, \dots, x_n]$ can be uniquely written as $f = g + h$ for g symmetric and h antisymmetric. Indeed, we have that

$$\begin{cases} f = g + h, \\ \tau f = g - h. \end{cases}$$

which we can solve to $g = \frac{f+\tau f}{2}$ and $h = \frac{f-\tau f}{2}$. (Here we use $\operatorname{char} k \neq 2$.)

Now, we have that

$$\prod_{i<j}(x_i - x_j)$$

is antinvariant because look at it. We then have the following sequence of observations; fix g any antinvariant polynomial.

- For any $x_i \neq x_j$, we see that g will vanish on setting $x_i = x_j$. If we let $\overline{g}$ be the polynomial after applying $x_i = x_j$, then we see that $g = -(i,j)g$ even though $\overline{(i,j)g} = \overline{g}$, so it follows that $\overline{g} = 0$.

- Now, given g vanishes on $x_i = x_j$, then ring theory lets us write $g = (x_i - x_j)h$. So each $x_i - x_j$ divides g, and we can see that these elements are irreducible (they are degree 1) and not off by a unit, so the product of these does indeed divide Δ.

- Thus, we can write
$$g = \Delta \cdot h$$
for some $h \in k[x_1, \dots, x_n]$, and in fact h is symmetric: for each $\sigma \in S_n$, we have
$$\sigma h = \frac{\sigma g}{\sigma \Delta} = \frac{(\operatorname{sgn}\sigma) g}{(\operatorname{sgn}\sigma)\Delta} = \frac{\sigma}{\Delta} = h.$$
The point is that the antinvariant polynomials are simply $\Delta \cdot k[x_1, \dots, x_n]^{S_n}$.

So brining this together, we see that
$$k[x_1, \dots, x_n]^{A_n} = k[x_1, \dots, x_n]^{S_n} \oplus \Delta \cdot k[x_1, \dots, x_n]^{S_n}.$$
But now if we wanted to generate $k[x_1, \dots, x_n]^{A_n}$ as a k-algebra, we note that $k[x_1, \dots, x_n] = k[e_1, \dots, e_n]$ because the $e_\bullet$ are algebraically independent[6], and then to add Δ, we see that
$$k[x, y, z]^{A_3} \cong \frac{k[e_1, \dots, e_n, \Delta]}{(\Delta^2 - \text{some polynomial in } k[e_1, e_2, e_3])}.$$
where the relation in the denominator comes from that $\Delta^2 \in k[e_1, \dots, e_n]$ is symmetric. It turns out that the relation for Δ^2 is quite annoying to compute; for example for $n = 3$ it is
$$\Delta^2 = 18e_1e_2e_3 - 4e_1^3e_3 + e_1^2e_2^2 - 4e_2^3 + 27e_3^2,$$
which is quite bad, but theoretically doable. ■

We would like to work out Δ^2 without tears.

Example 4.108. In two variables, our Δ^2 here is $(\alpha_1 - \alpha_2)^2 = e_1^2 - 4e_2$, which is really the discriminant of a monic quadratic.

So we more or less want the "discriminant" of a cubic.

Definition 4.109 (Discriminant)**.** Fix $f(x)$ a polynomial with roots $\alpha_1, \dots, \alpha_n$ in algebraic closure, the *discriminant* is defined as
$$\prod_{k \neq \ell} (\alpha_k - \alpha_\ell).$$

So how do we compute this? The answer is resultants.

4.3.5 Resultants

Set-Up

As usual, fix a field k and a polynomial
$$f(x) = x^n - e_1 x^{n-1} + \cdots \in k[x]$$
with roots $\alpha_1, \dots, \alpha_n$. We would like to compute
$$\Delta^2 = \prod_{k < \ell} (\alpha_k - \alpha_\ell)^2 \in k[e_1, \dots, e_n].$$
The main point is that $\Delta^2 = 0$ if and only if f has a multiple root.

[6] This is present in Lang; roughly speaking, the point is to do an induction on the number of variables so that an algebraic relation becomes a polynomial in one of the $x_\bullet$. Then the constant term must vanish by inductive hypothesis, but then we can divide out by $x_\bullet$ to force the entire relation to vanish.

Remark 4.110. Sylvester chose the name "discriminant" for this reason.

Going further, we see that f has a multiple root in the algebraic closure if and only if $\gcd(f, f') \neq 1$. We check this briefly.

Lemma 4.111. Fix k a field and $f \in k[x] \setminus \{0\}$. Then f has a double root in the algebraic closure of f if and only if $\gcd(f, f')$ has positive degree.

Proof. We have the following checks.

- In one direction, $(x-\alpha)^2 \mid f(x)$ implies $f(x) = (x-\alpha)^2 g(x)$ for some g implies $f'(x) = 2(x-\alpha)g(x) + (x-\alpha)^2 g(x)$ implies $(x-\alpha) \mid f, f'$. We can show the other direction purely formally.
- In the other direction, if $\gcd(f, f') \neq 1$, then checking for roots of $\gcd(f, f')$ in the algebraic closure, we get α such that $f(\alpha) = f'(\alpha) = 0$. Then $f(\alpha) = 0$ lets us write $f(x) = (x-\alpha)g(x)$ for some g so that
$$f'(x) = g(x) + (x-\alpha)g'(x),$$
and when we plug α one more time, we see that $g(\alpha) = 0$ so that we can write $g(x) = (x-\alpha)h(x)$ for some h. It follows $(x-\alpha)^2 \mid f$. ∎

Remark 4.112. There is a tricky thing that f' might vanish with f non-constant. For example, the derivative of $f(x) = x^p - a$ is 0 in $\mathbb{F}_p$. This is okay with multiple roots because, indeed, $x^p - a$ has only one root in $\overline{\mathbb{F}_p}$: if b is some root with $b^p = a$, then $x^p - a = x^p - b^p = (x-b)^p$, so we indeed only have the root b.

The point is that we are interested when f and f' have a root in common.

The Sylvester Matrix

Of course, there isn't much special about f', so we can just ask when two polynomials f and g have any root in common. For concreteness, we fix

$$f(x) = \sum_{k=0}^{\deg f} a_k x^k \qquad \text{and} \qquad g(x) = \sum_{k=0}^{\deg g} b_k x^k.$$

Supposing that these have a common root of α, we write $f(x) = (x-\alpha)p(x)$ and $g(x) = (x-\alpha)q(x)$. The key equation, now is that

$$f(x)q(x) = (x-\alpha)q(x)p(x) = g(x)p(x),$$

where $\deg p < \deg f$ and $\deg q < \deg g$. Indeed, the existence of such p and q is equivalent to f and g having a common root.

Lemma 4.113. Fix k a field and $f, g \in k[x] \setminus \{0\}$. Then f and g have a common root in the algebraic closure if and only if there exist $p, q \in k[x] \setminus \{0\}$ with $\deg p < \deg f$ and $\deg q < \deg g$ such that $fq = gp$.

Proof. The above argument gave the forward direction. In the reverse direction, we note that in $fq = gp$, we may assume p and q are coprime, else we could divide out by their greatest common divisor. But then $p \mid gp = fq$ implies that $p \mid g$, and similarly $q \mid fq = gp$ implies that $q \mid f$ so that

$$\frac{f}{p} = \frac{g}{q}$$

is a valid equation in $k[x]$. Now, $\deg p < \deg f$ and $\deg g < \deg q$ implies that both sides have positive degree, so both sides have a common root in the algebraic closure of k. It follows that f and g have a common root in the algebraic closure. ∎

But at this point we see that solving

$$fq = gp$$

with $\deg p < \deg f$ and $\deg q < \deg p$ is essentially some massive set of linear equations in p and q where the coefficients come from f and g. So we can check for nontrivial solutions for p and q by checking if the determinant of the corresponding coefficient matrix vanishes.

Let's see this. For concreteness, fix $m \coloneqq \deg f$ and $n \coloneqq \deg g$ with the coefficients

$$p(x) = \sum_{\ell=0}^{m-1} y_\ell x^\ell \qquad \text{and} \qquad q(x) = \sum_{\ell=0}^{n-1} z_\ell x^\ell.$$

Then

$$fq = \left(\sum_{k=0}^{m} a_k x^k\right)\left(\sum_{\ell=0}^{n-1} z_\ell x^\ell\right) = \sum_{d=0}^{n+m-1}\left(\sum_{k+\ell=d} a_k z_\ell\right) x^d$$

while

$$gp = \left(\sum_{k=0}^{\deg g} b_k x^k\right)\left(\sum_{\ell=0}^{m-1} y_\ell x^\ell\right) = \sum_{d=0}^{n+m-1}\left(\sum_{k+\ell=d} b_k y_\ell\right) x^d.$$

Comparing coefficients, we have the system

$$\sum_{k+\ell=d} a_k z_\ell - \sum_{k+\ell=d} b_k y_\ell = 0$$

for each $0 \le d < n+m-1$. Written out in matrix form, we get an $(n+m)\times(n+m)$ matrix which looks like

$$\begin{array}{l}
d=0 \\ d=1 \\ d=2 \\ \vdots \\ d=m \\ d=m+1 \\ d=m+2 \\ d=m+3 \\ \vdots \\ d=n+m-1
\end{array}
\begin{bmatrix}
a_0 & & & & & -b_0 & & & & \\
a_1 & a_0 & & & & -b_1 & -b_0 & & & \\
a_2 & a_1 & a_0 & & & -b_2 & -b_1 & -b_0 & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_m & a_{m-1} & a_{m-2} & \cdots & a_{m-n+1} & -b_m & -b_{m-1} & -b_{m-2} & \cdots & -b_1 \\
 & a_m & a_{m-1} & \cdots & a_{m-n+2} & -b_{m+1} & -b_m & -b_{m-1} & \cdots & -b_2 \\
 & & a_m & \cdots & a_{m-n+3} & -b_{m+2} & -b_{m+1} & -b_m & \cdots & -b_3 \\
 & & & \cdots & a_{m-n+4} & -b_{m+3} & -b_{m+2} & -b_{m+1} & \cdots & -b_4 \\
 & & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & & & & a_m & & & & & -b_n
\end{bmatrix}
\begin{bmatrix} z_0 \\ z_1 \\ z_2 \\ \vdots \\ z_{n-1} \\ y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{m-1} \end{bmatrix},$$

where the blank spaces are zeroes. We are interested in if this matrix has determinant zero, so we note that it does not matter if we make the $b_\bullet$ columns positive and transpose the matrix. Additionally, we can flip the columns/rows and rearrange them as we please while keeping the status of being zero unchanged, possibly introducing a sign here or there.

Doing all of this along with some aesthetic choices gives us the Sylvester matrix.

Definition 4.114 (Sylvester matrix)**.** Fix $f, g \in k[x]$ as above. Then the *Sylvester matrix* of f and g is the $(n+m) \times (n+m)$ matrix

$$\begin{bmatrix}
a_m & a_{m-1} & a_{m-2} & \cdots & a_0 & & & & & \\
 & a_m & a_{m-1} & \cdots & a_1 & a_0 & & & & \\
 & & a_m & \cdots & a_2 & a_1 & a_0 & & & \\
 & & & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \\
 & & & \cdots & a_{n-1} & a_{n-2} & a_{n-3} & a_{n-4} & \cdots & a_0 \\
b_n & b_{n-1} & b_{n-2} & \cdots & b_{n-m} & b_{n-m-1} & b_{n-m-2} & b_{n-m-3} & \cdots & \\
 & b_n & b_{n-1} & \cdots & b_{n-m+1} & b_{n-m} & b_{n-m-1} & b_{n-m-2} & \cdots & \\
 & & b_n & \cdots & b_{n-m+2} & b_{n-m+1} & b_{n-m} & b_{n-m-1} & \cdots & \\
 & & & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \\
 & & & \cdots & b_{n-1} & b_{n-2} & b_{n-3} & b_{n-4} & \cdots & b_0
\end{bmatrix},$$

where the first n rows shift the $a_\bullet$ across, and the next m rows shift the $b_\bullet$ across.

Definition 4.115 (Resultant)**.** Fix $f, g \in k[x]$ as above. Then the *resultant* of f and g, notated $\operatorname{Res}(f, g)$, is the determinant of their Sylvester matrix.

Now we have that f and g have a common root if and only if the determinant is zero. In particular, we have the following.

Proposition 4.116. Fix $f, g \in k[x]$. Then f and g have a common root in the algebraic closure of k if and only if the resultant of f and g is zero.

Proof. This follows from the above discussion. ∎

Remark 4.117. Our choice of m and n to be the degrees guaranteed that $a_m \neq 0$ and $a_n \neq 0$, which parts of what make the above argument work. Sometimes by convention we might take $a_m = 0$ and $b_n = 0$ inducing a "common root at ∞."

Examples

Now let's work out our original example: take $f \in k[x]$ of degree m, and we will require k to have characteristic 0 for psychological reasons. From our work with the resultant, we see that $\Delta^2 = 0$ if and only if $\operatorname{Res}(f, f') = 0$. So with the above notation, we will have

$$f(x) = \sum_{k=0}^{m} a_k x^k \qquad \text{and} \qquad f'(x) = \sum_{k=0}^{m-1} \underbrace{(k+1)a_{k+1}}_{b_k :=} x^k.$$

The point is that Δ^2 and $\operatorname{Res}(f, f')$ share the same roots, so certainly $\Delta \mid \operatorname{Res}(f, f')$, but in fact $\operatorname{Res}(f, f')$ is symmetric (it is a function of the coefficients of f, which are symmetric in the roots), so we get that

$$\Delta^2 \mid \operatorname{Res}(f, f').$$

Looking at the matrix for $\operatorname{Res}(f, f')$, we can look at each column of the $(2m-1) \times (2m-1)$ Sylvester matrix and note that their degrees in $k[\alpha_1, \dots, \alpha_m]$ fill in as follows.

$$\begin{bmatrix}
0 & 1 & 2 & \cdots & m-1 & m & & & & \\
 & 0 & 1 & \cdots & m-2 & m-1 & m & & & \\
 & & 0 & \cdots & m-3 & m-2 & m-1 & m & & \\
 & & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
 & & & & 0 & 1 & 2 & 3 & \cdots & m \\
1 & 2 & 3 & \cdots & m & & & & & \\
 & 1 & 2 & \cdots & m-1 & m & & & & \\
 & & 1 & \cdots & m-2 & m-1 & m & & & \\
 & & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
 & & & & & 1 & 2 & 2 & \cdots & m
\end{bmatrix}$$

The point is that when we are choosing a permutation for the determinant, one entry from each row/column, we will multiply them together and hence add these degrees.

Now, we see that each degree term increases linearly across a row, so the correct thing to do is to imagine adding in terms of extraneous degrees in the blank spaces—these terms of extraneous degree will have no effect on the determinant afterwards because their coefficient is zero and hence will all vanish. Anyways, we get the following.

$$\begin{bmatrix}
0 & 1 & 2 & \cdots & m-1 & m & m+1 & m+2 & \cdots & 2m-1 \\
-1 & 0 & 1 & \cdots & m-2 & m-1 & m & m+1 & \cdots & 2m-2 \\
-2 & -1 & 0 & \cdots & m-3 & m-2 & m-1 & m & \cdots & 2m-3 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \cdots \\
-m+1 & -m+2 & -m+3 & \cdots & 0 & 1 & 2 & 3 & \cdots & m \\
1 & 2 & 3 & \cdots & m & m+1 & m+2 & m+3 & \cdots & 2m \\
0 & 1 & 2 & \cdots & m-1 & m & m+1 & m+2 & \cdots & 2m+1 \\
-1 & 0 & 1 & \cdots & m-2 & m-1 & m & m+1 & \cdots & 2m+2 \\
-2 & -1 & 0 & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
-m+1 & -m+2 & -m+2 & \cdots & 0 & 1 & 2 & 2 & \cdots & m
\end{bmatrix}$$

The idea behind this is that each row has some specified "shift" from the top row, so if we imagine going vertically row-by-row to select our permutation, the accumulated degree will simply be the sum of all shifts plus the sum of the entries of the top row. In particular, the sum of degrees does not depend on our exact permutation, so $\operatorname{Res}(f, f')$ is in fact homogeneous.

We quickly note that we can compute $\deg \operatorname{Res}(f, f')$ explicitly by just choosing some random permutation: it is $1 \cdot m + m \cdot m = m(m+1)$ by choosing m of the 1s along the lower diagonal followed by all m of the ms along the top.

But now Δ^2 has the same degree in $k[\alpha_1, \dots, \alpha_n]$! So we find that $\Delta^2 = c \operatorname{Res}(f, f')$ for some nonzero $c \in k^\times$. A more sophisticated argument (say, present in Lang) is able to pin down what the coefficient c should be and can find that it ought be ± 1, but getting the exact sign is somewhat annoying.

Anyways, let's get to the examples.

Example 4.118. We compute the discriminant of $x^3 + bx + c$, which makes things easier to look at, and this is legal in characteristic not 3. We can compute the Sylvester matrix as

$$\begin{bmatrix} 1 & 0 & b & c & 0 \\ 0 & 1 & 0 & b & c \\ 3 & 0 & b & 0 & 0 \\ 0 & 3 & 0 & b & 0 \\ 0 & 0 & 3 & 0 & b \end{bmatrix}.$$

We can compute that this determinant is $4b^3 + 27c^2$. So is our discriminant the positive or negative sign? Well, we can look at a particular polynomial to pin this down. For example, the discriminant of $x^3 - x$ has $\prod(\alpha_k - \alpha_\ell)^2$ is positive because all the roots are positive, so $b = -1$ forces us to use the negative discriminant: $\boxed{-4b^3 - 27b^2}$.

Example 4.119. We work in $\mathbb{Z}[\alpha]$, where $\alpha^3 + \alpha + 1 = 0$. The discriminant of the number field is the discriminant of the polynomial is $-4(1)^3 - 27(1)^2 = -31$, so we see that 31 is our only ramified prime.

Example 4.120. We can ask when $y^2 = x^3 + bx + c$ is an elliptic curve. This requires testing for singularities, which happens when $x^3 + bx + c$ has multiple roots. So we are interested in testing for $4b^3 + 27c^2 \neq 0$.

We close with some remarks.

Remark 4.121. There is a geometric meaning for the discriminant: given two homogeneous polynomials $f, g \in k[z_1, \dots, z_m][x, y]$ (meaning the degrees of $x^a y^b$ are stable). Then we see that the vanishing set for f and g are going to define hypersurfaces H_f and H_g in $k^m \times \mathbb{P}^1$ respectively. Then the resultant is the projection of $H_f \cap H_g$ onto k^m.

For example, this is a closed set by definition of our Zariski topology, so we can fun things like that the projection $X \times \mathbb{P}^1 \to X$ takes closed sets to closed sets. In general, these projections do not have to be closed sets.

The discriminant is an example of a "syzygy," which is a word with no vowels. More seriously, a syzygy describes a relation between invariants. Namely, the discriminant gave the relation between the A_3-invariants Δ and the other symmetric polynomials. More generally the syzygies can be numerous and difficult to keep track of, so we might have second-order syzygies to keep track of these. This can get quite complex.

Example 4.122. Take $\langle g \rangle \cong \mathbb{Z}/n\mathbb{Z}$ acting on $\mathbb{C}[x, y]$ by $g \cdot x \coloneqq \zeta x$ and $g \cdot y \coloneqq \zeta y$, where ζ is a primitive nth root of unity. We saw last time that we have the invariants

$$x^n, \qquad x^{n-1}y, \qquad x^{n-2}, \qquad \dots.$$

We can label these $a_0, a_1, \dots$ by their degree of y, and we get lots of syzygies like $a_0 a_2 = a_1^2$ and $a_1 a_3 = a_2^2$.

Remark 4.123. Any word ending in "-ant" is probably an invariant, probably named by Sylvester. For example, the determinant, discriminant, bezoutiant, catalectiant, and so on.

4.4 October 28

I am heartbreak.

4.4.1 Formal Power Series

By way of example, our elements of the formal power series of $\mathbb{C}[[x]]$ are of the form

$$\sum_{k=0}^{\infty} a_k x^k,$$

where we don't care about convergence.

Example 4.124. For example,

$$\sum_{k=0}^{\infty} k! x^k$$

converges nowhere except for $x = 0$, but this is okay for $\mathbb{C}[[x]]$.

Anyways, we have the following.

Definition 4.125 (Formal power series, I)**.** Fix R a ring. A *formal power series* in $R[[x]]$ is a sequence of numbers $\{a_k\}_{k\in\mathbb{N}}$ represented by

$$\sum_{k=0}^{\infty} a_k x^k.$$

The operations of addition and multiplication are defined purely formally and work.

We can also construct this as an inverse limit.

Definition 4.126 (Formal power series, II)**.** We construct $R[[x]]$ as the completion of the ring of polynomials $R[x]$ at the ideal (x).

Wait, what is the completion?

Definition 4.127 (Completion)**.** The *completion* of a ring R at an ideal $\mathfrak{p}$ is the inverse limit of

$$\hat{R} \coloneqq \varprojlim R/\mathfrak{p}^{\bullet}.$$

Namely, we are constructing $R[[x]]$ as a sequence of compatible elements in the system

$$R[x]/(x) \leftarrow R[x]/\left(x^2\right) \leftarrow R[x]/\left(x^3\right) \leftarrow \cdots,$$

where these maps are defined by projectivity. In practice, this looks like a series of polynomials $\{a_k\}_{k\in\mathbb{N}}$ such that $a_k \equiv a_\ell \pmod{x^\ell}$ for each $k \geq \ell$. If we think about the exact monomials we are adding each time, this is really a formal power series.

Remark 4.128 (Nir)**.** In practice, Definition 4.126 might appear more awkward than the more physical power series, but in practice, this definition tells us that $R[[x]]$ is a ring effectively for free, which expedites a lot of our checks.

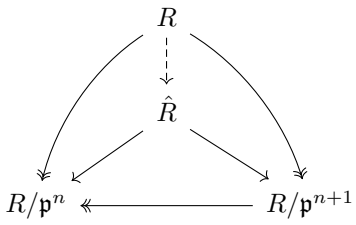
Remark 4.129 (Nir)**.** We briefly explain why this is called a "completion." Using $R[[x]]$ as an example, we note that $R[x]$ has a size function given by

$$|f|_{(x)} \coloneqq c^{-\text{order of vanishing of } f \text{ at } x=0}.$$

It turns out that $d(f,g) \coloneqq |f-g|_{(x)}$ forms a metric, and $R[[x]]$ is (canonically) isomorphic to this metric completion, justifying why we are calling $R[[x]]$ a completion—it is actually a metric completion.

Warning 4.130. There is a natural map $R \to \hat{R}$ induced by the natural projections $R \twoheadrightarrow R/\mathfrak{p}^\bullet$ and the universal property of the inverse limit. However, this need not be injective; it will be injective, for example, when R is a commutative, Noetherian integral domain.

To be more explicit, the map $R \to \hat{R}$ is induced by the following diagram.



Let's see some examples of the map $R \to \hat{R}$.

Example 4.131. The map $R[x] \hookrightarrow R[[x]]$ is in fact injective. With respect to the inverse limit definition, this comes down to the fact that a nonzero polynomial needs to have a nonzero coefficient $c_k x^k$ somewhere, and then the map into $R/\left(x^k\right)$ will not go to 0.

Non-Example 4.132. Consider the ring $R \coloneqq C^\infty(\mathbb{R})$ of smooth functions $\mathbb{R} \to \mathbb{R}$ and $I \subseteq R$ the R-ideal of smooth functions vanishing at 0; this is an ideal because $0 \in I$, and, for $r, s \in S$ and $f, g \in I$, we have $rf + sg \in I$ because

$$(rf + sg)(0) = r \cdot f(0) + s \cdot g(0) = 0.$$

The main problem with the map $R \to \hat{R}$ is that there are nonzero functions which go to 0 under each map $R \hookrightarrow R/I^\bullet$, which roughly corresponds with having a zero of "infinite order" at $x = 0$. For example, $e^{-1/\left(nx^2\right)} \in I$ for each $n \in \mathbb{Z}^+$, so

$$e^{-1/x^2} = \left(e^{-1/\left(nx^2\right)}\right)^n \in I^n$$

for each $n \in \mathbb{Z}^+$. So the function e^{-1/x^2} goes to 0 under the canonical map $R \to \hat{R}$.

Non-Example 4.133 (Miles)**.** In the ring $\mathbb{Z} \times \mathbb{Z}$, completing with respect to the (prime) ideal $I \coloneqq \{0\} \times \mathbb{Z}$ still has $\mathbb{Z} \times \mathbb{Z} \to \widehat{\mathbb{Z} \times \mathbb{Z}}$ not an injective map. For example, $(0, 1) \in I^n$ for each n.

Here is another important example of the completion.

Example 4.134. Fix p a rational prime. The ring $\mathbb{Z}_p$ of "p-adic numbers" is the completion of $\mathbb{Z}$ at the ideal (p). Namely, $\mathbb{Z}_p$ is the inverse limit of

$$\mathbb{Z}/p\mathbb{Z} \leftarrow \mathbb{Z}/p^2\mathbb{Z} \leftarrow \mathbb{Z}/p^3\mathbb{Z} \leftarrow \cdots,$$

where the leftwards maps are the canonical projections. These look quite similar to power series: if we write out compatible system of elements in "base p," this looks like

$$a_0 \in \mathbb{Z}/p\mathbb{Z}, \quad a_0 + a_1 p \in \mathbb{Z}/p^2\mathbb{Z}, \qquad a_0 + a_2 p + a_2 p^2 \in \mathbb{Z}/p^3\mathbb{Z}, \qquad \ldots.$$

In base p, this looks like an infinite sequence of digits going off to the left, which might look problematic but is fine as long as our addition and multiplication is purely formal.

Remark 4.135. Algebraic topologists have a bad habit of using $\mathbb{Z}_p$ to mean $\mathbb{Z}/p\mathbb{Z}$, essentially adding ambiguity for no good reason.

Remark 4.136. It is not advisable to let infinitely many digits go off to the left and right, then multiplication is no longer well-defined.

Remark 4.137 (Nir)**.** The intuition that $\mathbb{Z}_p$ is essentially "power series in p" can be rigorized in the isomorphism

$$\frac{\mathbb{Z}[[x]]}{(x-p)} \cong \mathbb{Z}_p,$$

where quotienting by $(x-p)$ is more or less the rigorization of plugging in p.

There is an important analogy between $\mathbb{Z}_p$ and $R[[x]]$ and especially $\mathbb{C}[[x]]$ or $\mathbb{F}_p[[x]]$. However, there is a difference in that our "p-digits" can induce strange carries in our arithmetic. For example,

$$1 + (p-1) = 0 + 1 \cdot p$$

is not something that can happen for formal power series. At a high level, the problem as the digits for $\mathbb{Z}_p$ are in $\{1, \dots, p-1\}$ which has not been given a ring structure in the same way that the coefficients of $R[[x]]$ have.

4.4.2 Ideals of Completions

Now fix k a field, and we will ask for the maximal ideals of $k[[x]]$.

Proposition 4.138. Fix k a field. The only maximal ideal of $k[[x]]$ is (x).

Proof. Suppose we have a formal power series

$$f(x) \coloneqq \sum_{i=0}^{\infty} a_i x^i.$$

The main point is that $a_0 \neq 0$ implies that f has an inverse. Indeed, we may write $a_0^{-1} f(x) = 1 + xg(x)$ for some $g(x) \in k[[x]]$. Then we have

$$\frac{1}{a_0^{-1} f(x)} = \frac{1}{1 + xg(x)} = \sum_{i=0}^{\infty} (-1)^i g(x)^i x^i,$$

which is a well-defined power series. Namely, we can envision $\frac{1}{f(x)}$ as the compatible sequence

$$\left\{ \sum_{i=0}^{n} (-1)^i g(x)^i x^i \right\}_{i \in \mathbb{N}}$$

because

$$\sum_{i=0}^{n+1} (-1)^i g(x)^i x^i \equiv \sum_{i=0}^{n} (-1)^i g(x)^i x^i \pmod{p^n}.$$

So we have succeeded in inverting $a_0^{-1} f$, which tells us that f is also invertible.

The point is that $k[[x]] / \setminus (x)$ are all units, and any unit will have to live in $k[[x]] \setminus (x)$. In fact, the following is true.

Lemma 4.139. Fix R a commutative ring with identity with an ideal $\mathfrak{m}$. Then $\mathfrak{m}$ is the unique maximal ideal if and only if $R \setminus \mathfrak{m} \subseteq R^\times$.

Proof. We have two claims.

- If $\mathfrak{m}$ is the unique maximal ideal, then we show that $R \setminus \mathfrak{m} = R^\times$. On one hand, $\mathfrak{m} \neq R$ implies that each $a \in \mathfrak{m}$ has $(a) \neq R$ so that $a \notin R^\times$, so $\mathfrak{m} \subseteq R \setminus R^\times$ so that $R \setminus \mathfrak{m} \subseteq R^\times$.

  On the other hand, if $a \in R^\times$, then $(a) \neq R$ is an ideal and must be contained in some maximal ideal, so it follows $a \in (a) \subseteq \mathfrak{m}$. So indeed, $R^\times \subseteq R \setminus \mathfrak{m}$.
- Take $\mathfrak{m}$ an ideal with $R \setminus \mathfrak{m} \subseteq R^\times$; note each $a \in R^\times$ also has $a \notin \mathfrak{m}$ because $\mathfrak{m} \neq R$, so in fact $R \setminus \mathfrak{m} = R^\times$. Now, for any other maximal ideal $\mathfrak{m}'$, we see that any $a \in \mathfrak{m}'$ has $a \notin R^\times$ (else $\mathfrak{m}' = R$), so $a \in \mathfrak{m}$. It follows that
$$\mathfrak{m}' \subseteq \mathfrak{m} \subsetneq R.$$
By maximality of $\mathfrak{m}'$, we see that $\mathfrak{m} = \mathfrak{m}'$ is forced. Note that this argument tells us that $\mathfrak{m}$ is a maximal for free because there exists at least one maximal ideal $\mathfrak{m}'$. ■

So the above tells us that, because $k[[x]] \setminus (x) \subseteq k[[x]]^\times$, we have that (x) is the unique maximal ideal. ■

In fact, stronger is true.

Proposition 4.140. The only ideals of $k[[x]]$ are (0) or $(x^\bullet)$.

Proof. The main point is that, for any nonzero $f \in k[[x]] \setminus \{0\}$, there exists some k for which the coefficient $c_i x^i$ of f is nonzero, so we may set
$$\nu\left(\sum_{i=0}^{\infty} c_i x^i\right) \coloneqq \min\{i \in \mathbb{N} : c_i \neq\}.$$
With this in mind, we see that, for any $f \in k[[x]]$, we have $f/x^{\nu(f)}$ has nonzero constant term, so $f/x^{\nu(f)}$ is a unit. In particular, $\left(f/x^{\nu(f)}\right) = (1)$ so that $(f) = \left(x^{\nu(f)}\right)$.

More generally, for any nonzero ideal I, we have that
$$I = \bigcup_{f \in I \setminus \{0\}} (f) = \bigcup_{f \in I \setminus \{0\}} \left(x^{\nu(f)}\right) = \left(x^{\min\{\nu(f) : f \in I \setminus \{0\}\}}\right),$$
so indeed, all nonzero ideals take the form $(x^\bullet)$. ■

There is something similar which happens for $\mathbb{Z}_p$.

Proposition 4.141. Fix p a rational prime. The only ideals of $\mathbb{Z}_p$ are (0) or $(p^\bullet)$.

Proof. This was more or less on the homework, and it is qiute similar to the case of $k[[x]]$; we take on faith that (p) is the unique maximal ideal of $\mathbb{Z}_p$ because we showed it on the homework. Again, the main point is that nonzero elements $a \in \mathbb{Z}_p$ can be given a "valuation"
$$\nu(a) \coloneqq \min\left\{n \in \mathbb{N} : a \in (p^n)\right\}.$$
In particular, $a/p^{\nu(a)} \notin (p)$, but $\mathbb{Z}_p \setminus (p) = \mathbb{Z}_p^\times$ because (p) is the unique maximal ideal, as shown on the homework. From this it follows $a/p^{\nu(a)}$ will be a unit, so $(a) = \left(p^{\nu(a)}\right)$. Thus, for any nonzero ideal I, we can write
$$I = \bigcup_{a \in I \setminus \{0\}} (a) = \bigcup_{a \in I \setminus \{0\}} \left(p^{\nu(a)}\right) = \left(p^{\min\{a \in I \setminus \{0\} : \nu(a)\}}\right),$$
which is what we wanted. ■

4.4.3 More Variables

We would like to define $k[[x,y]]$. There are a couple ways to do this.

Definition 4.142 (Multivariable power series, I)**.** We can define $k[[x,y]]$ can be defined as the completion of $k[x,y]$ with respect to the (maximal) ideal (x,y). Here our system of compatible elements more or less looks like

$$\sum_{ij,j=0}^{\infty} a_{i,j}x^i y^j.$$

More generally, we can define $k[[x_1,\dots,x_n]]$ as the completino of $k[x_1,\dots,x_n]$ with respect to the (maximal) ideal $(x_1,\dots,x_n)$.

We can also do the following.

Definition 4.143 (Multivariable power series, II)**.** We can define $k[[x,y]]$ as $k[[x]][[y]]$, which is essentially the formal power series in y with coefficients in $k[[x,y]]$. More generally, we can inductively define

$$k[[x_1,\dots,x_n]] \coloneqq k[[x_1,\dots,x_{n-1}]][[x_n]].$$

I don't really care about proving that these definitions are equivalent, so we won't.

We also get the similar property as before.

Proposition 4.144. The only maximal ideal of $k[[x_1,\dots,x_n]]$ is $(x_1,\dots,x_n)$.

Proof. The main claim is that $k[[x_1,\dots,x_n]] \setminus (x_1,\dots,x_n) \subseteq k[[x_1,\dots,x_n]]^\times$ again, which will be enough by Lemma 4.139. To see, this we start by noting that $f \in (x_1,\dots,x_n)$ if and only if its constant term is 0.[7] So suppose we have $f \in k[[x_1,\dots,x_n]]$ with nonzero constant term. This means that we may write

$$c_0^{-1}f = 1 + \sum_{i=1}^{n} x_i g_i$$

for some $g_\bullet \in k[[x_1,\dots,x_n]]$. As before, we may formally invert this as

$$\frac{1}{c_0^{-1}f} = \frac{1}{1 - -\sum_{i=1}^n x_i g_i} = \sum_{d=0}^{\infty} \left(-\sum_{i=1}^{n} x_i g_i\right)^d.$$

Again these partial sums form a valid compatible sequence because, for any N,

$$\sum_{d=0}^{N+1} \left(-\sum_{i=1}^{n} x_i g_i\right)^d = \sum_{d=0}^{N} \left(-\sum_{i=1}^{n} x_i g_i\right)^d + \left(-\sum_{i=1}^{n} x_i g_i\right)^{N+1},$$

where the second term is in $(x_1,\dots,x_n)^N$ because each $-\sum_{i=1}^n x_i g_i \in (x_1,\dots,x_n)$. This finishes. ∎

However, the other ideals are more complicated than before, essentially due to the extra dimension.

4.4.4 Getting Noetherian

Given Hilbert's basis theorem showing that $R[x]$ is Noetherian from R Noetherian, we might hope that $R[[x]]$ is Noetherian. Let's see this.

[7] If $f \in (x_1,\dots,x_n)$, then write out a representation. If f has constant term 0, then each monomial of f with nonzero coefficient has nonzero degree, so group them in any reasonable way.

Theorem 4.145. If R is Noetherian, then $R[[x]]$ is Noetherian.

Proof. We can essentially copy the proof for $R[x]$ by using the coefficient of least degree instead of largest degree. Namely, with $R[x]$ we looked at the leading coefficients, but $R[[x]]$ doesn't have largest coefficients. To salvage this, we look at the smallest nonzero power in some element of $R[[x]]$. Fix I an ideal $R[[x]]$ which we would like to finitely generate. Then we define

$$I_0 \coloneqq \{\text{constant terms for } f \in I\}.$$

More generally, we have

$$I_n \coloneqq \left\{ a_n : \sum_{k=n}^{\infty} a_k x^k \in I \right\}.$$

Roughly the same reasoning gets us the ascending chain of R-ideals.

$$I_0 \subseteq I_1 \subseteq \cdots.$$

Indeed, we have the following checks.

- To see that I_n is an R-ideal, we note that $0 \in R[[x]]$ has $0x^n$ for its x^n coefficient, so $0 \in I_n$. Then for any $c_n, d_n \in I$ and $r, s \in R$, there exist $f, g \in R[[x]]$ with their x^n coefficient equal to c_n and d_n respectively, with no terms of smaller degree. Then
$$rf + sg$$
will have leading their x^n coefficient equal to $rc_n + sc_n \in I$, again with no terms of smaller degree. So I is nonempty and closed under R-linear combination.
- To see that $I_n \subseteq I_{n+1}$, we note that if f has c_n for its x^n coefficient and no terms of smaller degree, then xf has c_n for its x^{n+1} coefficient and no terms of smaller degree, so $c_n \in I_{n+1}$.

Anyways, the point is that we get our ascending chain of ideals will stabilize to some I_N because R is Noetherian. Then each of I_k for $0 \le k \le N$ is finitely generated, so we say that

$$I_k = (c_{k,1}, c_{k,2}, \ldots, c_{k,n_k}).$$

Now, for each $c_{k,\ell} \in I_d$, there exists a polynomial $f_{k,\ell}$ with that coefficient and no terms of smaller degree, by definition of I_k. We claim that the $f_{k,\ell}$ for $0 \le k \le N$ and $1 \le \ell \le n_k$ (of which there are finitely many) generate I.

For clarity, define, for $f \in R[[x]] \setminus \{0\}$,

$$\deg f = \deg \left(\sum_{k=0}^{\infty} a_k x^k \right) \coloneqq \min\{k \in \mathbb{N} : a_k \neq 0\},$$

which is well-defined because $f \neq 0$ requires some coefficient to be nonzero. We now show that the $f_{k,\ell}$ generate f directly. There are two steps.

1. If $d \coloneqq \deg f < N$, then we claim that we can find $\{r_{d,\ell}\}_{\ell=1}^{n_d} \subseteq R$ so that $\deg f < \deg\left(f - \sum_\ell r_{d,\ell} f_{d,\ell}\right)$. Indeed, write
$$f(x) = \sum_{k=d}^{\infty} c_k x^k$$
so that $c_d \in I_d$. This implies that there exists $\{r_{d,\ell}\}_{\ell=1}^{n_d} \subseteq R$ such that
$$c_d = \sum_{\ell=1}^{n_d} r_{d,\ell} c_{d,\ell}$$

so that

$$f - \sum_{\ell=1}^{n_d} r_{d,\ell} f_{d,\ell}$$

has no terms of degree smaller than x^d and also has the x^d term vanish. This is what we wanted. (Here we have used the fact that I_d is only represented by polynomials which have no term of degree smaller than x^d.)

Inductively repeating the above process will give us elements $r_{k,\ell}$ for $1 \le k < N$ and $1 \le \ell \le n_k$ such that

$$f - \sum_{k=1}^{N-1} \sum_{\ell=1}^{n_k} r_{k,\ell} f_{k,\ell}$$

has degree at least N.

2. So now we may take $d := \deg f \ge N$. We claim that there exists $\{r_{d,\ell}\}_{\ell=1}^{n_N} \subseteq R$ so that $\deg f < \deg\left(\sum_\ell r_{d,\ell} f_{N,\ell}\right)$ again. Write

$$f(x) = \sum_{k=d}^{\infty} c_k x^k$$

so that $c_d \in I_d = I_N$. This implies that there exists $\{r_{d,\ell}\}_{\ell=1}^{n_N} \subseteq R$ such that

$$c_d = \sum_{\ell=1}^{n_N} r_{d,\ell} c_{N,\ell}$$

so that

$$f - \sum_{\ell=1}^{n_N} r_{d,\ell} x^{d-N} f_{N,\ell}$$

has no terms of degree smaller than x^d and also has the x^d term vanish. This is what we wanted.

If we combine the two steps, we see that, for any $f \in I$, we have found coefficients $r_{k,\ell}$ so that

$$f - \sum_{k=1}^{N-1} \sum_{\ell=1}^{n_k} r_{k,\ell} f_{k,\ell} - \sum_{k=N}^{\infty} \sum_{\ell=1}^{n_N} r_{k,\ell} x^{k-N} f_{N,\ell}$$

vanishes completely. (Technically, we ought truncate the second sum and show that the truncated sum vanishes as we add more terms. We will not do this because I already have a headache.) In other words,

$$f = \sum_{k=1}^{N-1} \sum_{\ell=1}^{n_k} r_{k,\ell} f_{k,\ell} + \sum_{\ell=1}^{n_N} \left(\sum_{k=N}^{\infty} r_{k,\ell} x^{k-N} \right) f_{N,\ell},$$

so indeed, we have represented f as an $R[[x]]$-linear combination of the $f_{k,\ell}$. Importantly, those are power series at the end sum, and they do converge. ■

Remark 4.146. This proof does not work for polynomials because the induction at the end technically need not terminate. All that the proof required is that the induction creates a power series linear combination.

4.4.5 Unique Factorization

We would like to have unique factorization. Of course, $k[[x]]$ is safe because it is a principal ideal domain. Well, what about $k[[x, y]]$? In theory, we should be able to show that if R is a unique factorization domain, then $R[[x]]$ is a unique factorization domain. However, we won't do this because it is false; the exact counterexample is relatively uninteresting.

Remark 4.147. Some early versions of Lang's *Algebra* claimed that this was true. Early versions are notorious for this.

Roughly speaking, the proof for polynomials used the content of a polynomial, which makes sense because polynomials are nice finite objects. Namely, given a polynomial $f \in k[x]$, we could find some $c(f) \in k$ such that

$$\frac{f}{c(f)} \in R[x]$$

and have coprime coefficients. However, this is not possible for power series.

Example 4.148. Over $\mathbb{Z}[[x]]$, the power series

$$1 + \frac{1}{p}x + \frac{1}{p}^2 x^2 + \cdots \in \mathbb{Q}[[x]]$$

will have no content to get it into $\mathbb{Z}[[x]]$.

The correct thing to do here is to use the Weierstrass preparation theorem.

Theorem 4.149 (Weierstrass preparation)**.** An element $f \in k[[x_1, \dots, x_n]]$ can be made to look like a polynomial in x_1. More precisely, in $k[[x, y]]$, we assert that any $f \in k[[x, y]]$ can be written as $y^\bullet ug$ where $u \in k[[x, y]]^\times$, and g has the form

$$\sum_{k=0}^{n} a_i x^k i,$$

where $a_k \in k[[y]]$ and $a_n = 1$. We call g a *Weierstrass polynomial in* x, and it is unique.

Remark 4.150. According to Professor Borcherds, units are kind of harmless.

Proof. The idea is to turn f into g by repeatedly multiplying by some harmless units of the form $1 + x^i y^j$, which will let us kill various coefficients of f. Namely, here are the monomials for $f \in k[x, y]$.

$$\begin{array}{ccccc}
\vdots & \vdots & \vdots & \vdots & \iddots \\
x^3 & x^3y & x^3y^2 & x^3y^3 & \cdots \\
x^2 & x^2y & x^2y^2 & x^2y^3 & \cdots \\
x & xy & xy^2 & xy^3 & \cdots \\
1 & y & y^2 & y^3 & \cdots
\end{array}$$

To begin, we write

$$f(x, y) = \sum_{i,j=0}^{\infty} c_{i,j} x^i y^j \neq 0,$$

and we see that dividing out by some power of y will eventually make one of the $x^\bullet$ coefficients nonzero, so without loss of generality take $f \notin (y)$. If the 1 coefficient is nonzero, then f is already a unit, so we are done.

So suppose, by way of example, that the 1 and x have coefficients of 0, and we focus on x^2. Without loss of generality, assert that x^2 has coefficient 1.

We now note that all of the x^iy^j for $i<2$ can be thrown into our Weierstrass polynomial $g(x,y)\in k[[y]][x]$ as

$$f(x,y)=x^0\underbrace{\sum_{\ell=0}^{\infty}c_{0,\ell}y^\ell}_{a_0}+x^1\underbrace{\sum_{\ell=0}^{\infty}c_{1,\ell}y^\ell}_{a_1}+x^2\left(1+\cdots\right),$$

but now the rest of them need to be killed. We are going to kill these in column-by-column, starting with the y^0 column, then moving to the y^1 column, and so on.

We will recursively kill our monomials.[8] Let's say that the current smallest nonzero monomial is cx^iy^j, where $c\neq 0$ is some constant. Explicitly, we have found a unit u such that

$$f-ug$$

has cx^iy^j as its least monomial. The point here is that ug also has the small x^2 term coming from f because u will have a nonzero constant coefficient. Then we observe that multiplying

$$f-g\cdot u\left(1+cx^{i-2}y^j\right)=(f-gu)-gu\cdot cx^{i-2}y^j$$

Here we see that the $cx^{i-2}y^j$ moves the $1x^2$ term in gu term up to the term cx^iy^j term we need to kill, so indeed, we have killed this term. Additionally, we see that multiplying all terms by $x^{i-2}y^j$ means that any other monomial in gu will get moved to the upper-right of the x^iy^j multiplication, meaning we have not altered any of the previously killed monomials.

There is some care here because we have infinite product for the unit u. These will converge, however, because our factors are of the form $\left(1+x^iy^j\right)$, so our coefficients will converge somewhat rapidly.

Lastly, it remains to check that this g is unique. Well, suppose we have that

$$y^au\sum_{i=0}^{m}a_ix^i=y^bv\sum_{j=0}^{n}b_jx^j,$$

where u,v are units and $a_i,b_j\in k[[y]]$ with $a_m=b_n=1$. We see that $a=b$ is forced by $a_m=b_n=1$ because the $y^\bullet$ here describes the largest power of y dividing into either side. Then, rearranging, we have that

$$v^{-1}u\sum_{i=0}^{m}a_ix^i=\sum_{j=0}^{n}b_jx^j.$$

At this point uniqueness follows for reasons I don't really understand, but we can kind of see this because the right-hand side has limied degree in x. ■

Remark 4.151. We can do this for any number of variables, but it requires tears.

Now let's show our unique factorization.

Theorem 4.152. We have that $k[[x,y]]$ is a unique factorization domain.

Proof. We know that $k[[x,y]]$ is Noetherian, so any element has an irreducible factorization. The hard part is getting uniqueness, which requires knowing that irreducibles are prime. Well, fix f irreducible dividing the product gh, and we want to show that $f\mid g$ or $f\mid h$. By the Weierstrass preparation theorem, we may get rid of units, and we may also remove powers of y without making our lives easier, so we assume that f,g,h are Weierstrass polynomials.

[8] Technically this will require a Zorn's lemma argument to show that recursively going up any column can move all the way across. But I don't want to write this out, so let's pretend we don't have to do this.

Now, $f \mid gh$ lets us write

$$fr = gh.$$

But now fr must be a Weierstrass polynomial, so the key point is that we may deduce that r is a Weierstrass polynomial. But now the above equality lives in $k[[y]][x]$ (!), we may reduce to unique factorization here, and f is irreducible in $k[[x,y]]$ gives f irreducible in $k[[y]][x]$, which implies that f is prime in $k[[y]][x]$, so in the above we have that $f \mid g$ or $f \mid h$ in $k[[x,y]]$. ■

At a high level, we have had two main steps.

- Use Weierstrass preparation to make things a polynomial in one variable.
- Use unique factorization in a polynomial ring to finish.

Remark 4.153. Here are some traps in this proof.

- Again, for R a unique factorization domain, $R[[x]]$ is not necessarily a unique factorization domain.
- If $f \mid g$ in $k[[x,y]]$, then we do not necessarily have $f \mid g$ in $k[[y]][x]$, even in $f, g \in k[[y]][x]$. For example, $g = 1$ adn $f = 1 + x$ are both units in $k[[x,y]]$, but f is not a unit in $k[[y]][x]$.
- Irreducible polynomials in $k[x,y]$ need not be irreducible in $k[[x,y]]$. For example, $f = 1 + x + y$ is irreducible in $k[x,y]$ but not in the formal power series. Or the elliptic curve $y^2 - x^2 - x^3$ is irreducible in $k[x,y]$ but we can write $y^2 - x^2(1+x)$ as

  $$\left(y - x\sqrt{1+x}\right)\left(x + x\sqrt{1+x}\right).$$

  At a high level, this is because the curve $y^2 - x^3 - x^2$ looks reducible when we look locally at $(0,0)$; namely, it looks like $y^2 - x^2$ close to $(0,0)$, which certainly reduces.

4.4.6 Hensel's lemma

Let's talk about Hensel's lemma. There are lots of variations, but they are essentially just about factorization of polynomials in $\hat{R}[x]$, where $\hat{R}$ is some completion. At a high level, the statement is that factorization in R/I^n can occasionally be lifted directly to a factorization of $\hat{R}[x]$.

The most common case, for number theory, is to take $\hat{R} = \mathbb{Z}_p$ the p-adic numbers. Here is the simplest possible case.

Theorem 4.154 (Hensel)**.** Fix $f \in \mathbb{Z}_p[x]$. Suppose $f \pmod p$ has a root $f(\alpha) \equiv 0 \pmod p$ but $f'(\alpha) \not\equiv 0 \pmod p$. Then α lifts to a root in $\mathbb{Z}_p$.

Non-Example 4.155. It is not always possible to lift roots up to $\mathbb{Z}_p$. For example, for $x^2 - 1 \in \mathbb{Z}_2[x]$, we cannot lift the solution $3 \pmod 8$: no number which is $3 \pmod 4$ is a solution to $x^2 - 1 \equiv 16$. The issue here is that $x^2 - 1$ completely vanishes $\pmod 2$.

Let's see a real example of this.

Exercise 4.156. We lift the root $x \equiv 1 \pmod 3$ of $x^2 = 7$ in $\mathbb{Z}_3$.

Proof. We start by lifting to $\mathbb{Z}/9\mathbb{Z}$. Well, set $x_0 = 1$ and then we want to find some a for which $x_1 := 1 + 3a_1$ has $x_1^2 \equiv 7 \pmod 9$. Expanding we want

$$1 + 6a + 9a_1^2 \equiv 7 \pmod 9,$$

which reduces to requiring $a \equiv 1 \pmod 3$. So we set $x_1 \coloneqq 1 + 3 \cdot 1$. Continuing, we want some b for which $x_2 \coloneqq 1 + 3 \cdot 1 + 9a_2$ and $x_2^2 \equiv 7 \pmod{27}$. Expanding, we want

$$7 + 9 + 2 \cdot 9a_2 \equiv 7 \pmod{27}.$$

This rearranges to $2a_2 \equiv -1 \pmod 3$, or $a_2 \equiv 1 \pmod 3$, giving $x_2 = 1 + 3 + 9$. We can keep doing this because the coefficient in front of the $a_\bullet$ is nonzero, which comes from the fact that the derivative of $x^2 - 7$ at $x_0 = 1$ is nonzero.

More explicitly, suppose by way of induction that we have x_n for which $x_n^2 \equiv 7 \pmod{3^{n+1}}$ so that we want to lift x_n to $x_{n+1} \equiv x_n \pmod{3^{n+1}}$ such that $x_{n+1}^2 \equiv 7 \pmod{3^{n+2}}$. Well, we want $x_{n+1} = x_n + 3^{n+1}a$, so writting this out means we want

$$x_n^2 + 2 \cdot 3^{n+1}a \equiv 7 \pmod{3^{n+2}},$$

where we can see the derivative of $x^2 - 7$ at $x_0 = 1$ is that 2. Anyways, this rearranges to

$$a \equiv \frac{7 - x_n^2}{2 \cdot 3^{n+1}} \pmod 3,$$

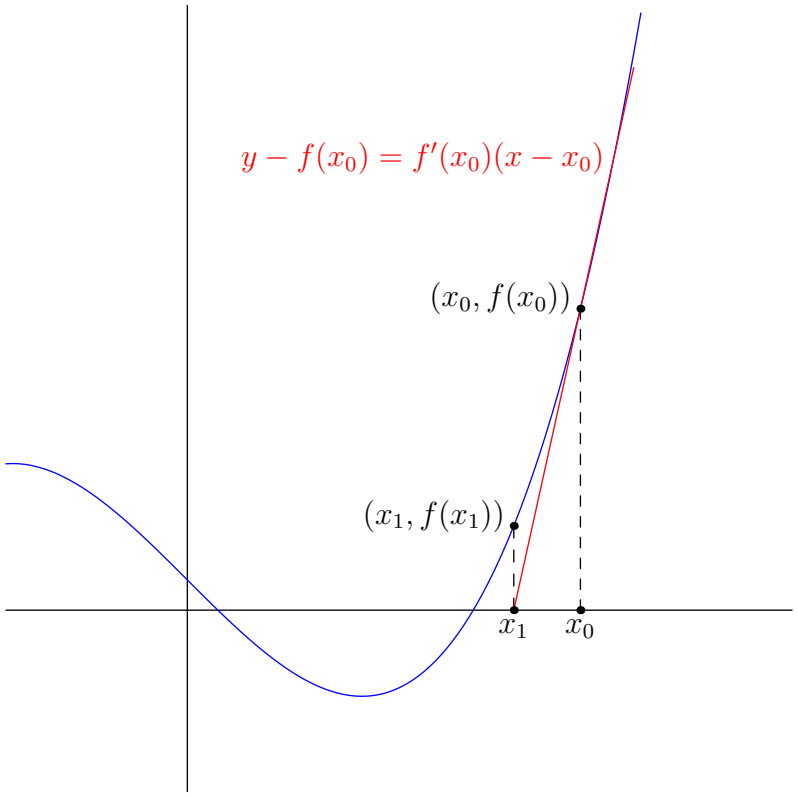
which is perfectly valid, finishing our lifting. The point is that this induction gives us a compatible sequence[9]

$$(x_0, x_1, \ldots) \in \mathbb{Z}_p,$$

which is the root we wanted because it squares to 7 in every $\pmod{p^\bullet}$. ∎

The above can be turned into a formal proof by just continuing it by force. We're going to give a different proof.

Proof of Theorem 4.154. We show that this is essentially a special case of Newton's method for finding roots. Recall Newton's method of finding a root of $f(x)$ by just guessing somewhere x_n, drawing the tangent line, and finding where it intersects the x-axis, and use that for our new x_{n+1}. Here is the image.



[9] Perhaps I should mutter something about Zorn's lemma, but I won't.

Writing this out, we find that our recursion should be

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

For concreteness, we have $f \in \mathbb{Z}_p[x]$ and have some x_0 for which $f(x_0) \equiv 0 \pmod{p}$. For our induction, suppose that we have found $x_n \equiv x_0$ for which $f(x_n) \equiv 0 \pmod{p^n}$, and by way of induction, say that $f'(x_n) \equiv f'(x_0) \not\equiv 0 \pmod{p}$.

Now, we can expand $f(x_{n+1})$ as a Taylor series about x_n, as shown in the homework. Indeed, we have

$$f(x_{n+1}) = f(x_n) - \frac{f'(x_n)}{1}\left(\frac{f(x_n)}{f'(x_n)}\right) + \frac{f''(x_n)}{2}\left(\frac{f(x_n)}{f'(x_n)}\right)^2 + \cdots.$$

The point here is that the first two terms will cancel, which in fact one reason Newton's method is good. So we see that

$$f(x_{n+1}) \equiv \sum_{k=2}^{\infty} \frac{f^{(n)}(x_n)}{n!}\left(\frac{f(x_n)}{f'(x_n)}\right)^k \equiv 0 \pmod{p^{2n}}$$

because all later terms have at least $2n$ many powers of p coming from $\left(\frac{f(x_n)}{f'(x_n)}\right)^k$.

There is some concern that perhaps $\frac{f^{(n)}(x_n)}{n!}$ is not a well-defined element of $\mathbb{Z}_p$ because of the denominator. However, we can see that a monomial $f(x) = x^d$ for $d \geq n$ will have

$$\frac{f^{(n)}(x)}{n!} = \frac{d(d-1)(d-2)\cdot\ldots\cdot(d-n+1)}{n!}x^{n-d} = \binom{d}{n}x^{n-d},$$

so the coefficient is perfectly well-defined as an integer and hence in $\mathbb{Z}_p$. From here we can extend lienarly out to all polynomials f. ■

Remark 4.157. In the exercises above, we get approximately one digit each time. In contrast, as we showed in the proof, Newton's method will square our accuracy/double the number of digits, and we don't have any of the problems of Newton's method for real numbers.

Remark 4.158 (Nir)**.** I am under the impression that the above proof works as long as the largest power of p dividing $f(\alpha)$ exceeds the largest power of p dividing $f'(\alpha)^2$. Observe that we are essentially requiring the first remainder term in our Taylor expansion

$$\frac{f''(x_n)}{2}\left(\frac{f(x_n)}{f'(x_n)}\right)^2$$

to be small.

It is also true that Newton's method works for power series.

Example 4.159. Consider $y^2 - x^2 - x^3$. It factors in $k[[x,y]]/(x,y)^3$ as $(y-x)(y+x)$. Then Hensel's lemma for power series lets this factorization lift all the way upwards to $k[[x,y]]$.

Non-Example 4.160. We have the factorization $y^2 - x^3 \equiv y \cdot y \pmod{(x,y)^3}$, but this factorization does not lift to $k[[x,y]]$. The reason is that the given factorization induces multiple roots, which will influence having derivative zero. Intuitively, Hensel's lemma "doesn't know" which root we are supposed to lift.

Remark 4.161. The condition $f'(\alpha) \not\equiv 0 \pmod{p}$ more or less is telling us that we are only lifting simple roots.

Again geometrically, the point is that $y^2 - x^2 - x^3$ will look like a cross, but $y^2 - x^3$ has a cusp at $(0,0)$, so geometrically this does not obviously reduce.

Theme 5
Galois Gossip

Usually mathematicians have to shoot somebody to get this much publicity.

—Thomas R. Nicely

5.1 November 2

A few hours grace before the madness begins again.

5.1.1 Algebraic Extensions

So we're talking about fields and Galois theory for the last third of the class. Today we're mostly doing a field review.

Definition 5.1 (Field extension)**.** A *field extension* L/K is a field L containing a field K.

We are interested in field extensions rather than the field itself because oftentimes we can decompose some complicated field M into its subfields and be able to study M in this more controlled way.

Here is an important invariant.

Definition 5.2 (Degree)**.** The *degree* of a field extension L/K, denoted $[L : K]$, is the dimension of L as a K-vector spaces.

Remark 5.3. Yes, field containments induce vector spaces. This is a good thing to check once.

Example 5.4. The degree of $\mathbb{C}/\mathbb{R}$ is $[\mathbb{C} : \mathbb{R}] = 2$, where our basis is (say) $\{1, i\}$.

We have the following definition from this.

Definition 5.5 (Algebraic)**.** An element $\alpha \in L$ is *algebraic* over K if and only if α is the root of a polynomial in $K[x]$. We say that L is an *algebraic extension over* K if all of its elements are algebraic over K.

Example 5.6. The number $i \in \mathbb{C}$ is algebraic over $\mathbb{R}$ and $\mathbb{Q}$.

Non-Example 5.7. The number π is not algebraic over $\mathbb{Q}$, and the proof is hard.

Non-Example 5.8. Look at the extension $\mathbb{Q} \subseteq \mathbb{Q}(x)$, where $\mathbb{Q}(x)$ is the field of rational functions. Then, by construction essentially, x is not algebraic over $\mathbb{Q}$.

"Not algebraic" elements have a name.

Definition 5.9 (Transcendental)**.** An element $\alpha \in L$ which is not algebraic over K is called *transcendental*.

In Galois theory, we mostly care about finite, algebraic extensions.

5.1.2 Constructing Algebraic Extensions

To construct an algebraic extension, we have the following proposition.

Proposition 5.10. Start with a field K and some polynomial $\pi \in K[x]$. Then it happens that

$$L \coloneqq \frac{K[x]}{(p)}$$

is a ring, and it is a field if and only if p is irreducible.

Proof. We know that $K[X]$ is a ring, and $(p) \subseteq K[X]$ is an ideal, so $K[X]/(p)$ is the quotient ring.

The key point to getting a field is that p is irreducible if and only if all nonzero elements have inverses. Indeed, fix q nonzero in $K[x]/p(x)$. Then because p is irreducible and $p \nmid q$, we have that p and q are coprime, so $(p) + (q) = (1)$, where we are using the fact that $K[x]$ is a principal ideal domain. It follows that there are polynomials a and b such that

$$ap + bq = 1,$$

so $bq \equiv 1 \pmod{p}$, finishing.

In the reverse direction, We can ask what happens if p is not irreducible. Well, if we can write $p = fg$ where f and g are coprime nonconstant polynomials, then the Chinese remainder theorem lets us write

$$\frac{K[x]}{(p)} \cong \frac{K[x]}{(f)} \times \frac{K[x]}{(g)}.$$

In particular, it follows that this has zero-divisors ($(1,0) \cdot (0,1) = (0,0)$) and hence is not a field. ∎

Remark 5.11. The above proof is actually effective for finding inverses: we can use the (extended) Euclidean algorithm to find the a and b such that $ap + bq = 1$, and then we can extract our inverse like that.

Example 5.12. We have that $\mathbb{C} \cong \mathbb{R}[x]/\left(x^2+1\right)$.

We give some remarks in the case where p is not irreducible in Proposition 5.10. In general, if

$$p = \prod_{\pi | p} \pi^{\alpha_\pi}$$

so that

$$\frac{K[x]}{(p)} \cong \prod_{\pi | k} \frac{K[x]}{(\pi^{\alpha_\pi})}.$$

If $\alpha_\pi = 1$, then we get a field, which is nice, and when everything has $\alpha = 1$, then we just have a product of fields. But when $\alpha_\pi > 1$, then we get nilpotent elements, which is very not good.

We also have the following statement.

Proposition 5.13. Fix L/K an extension. Then $\alpha \in L$ is algebraic over K if and only if α is contained in a finite sub-extension.

Proof. In one direction, if α is algebraic, then α is the root of some (without loss of generality) irreducible $p \in K[x]$. Then we can place

$$K[\alpha] \cong \frac{K[x]}{(p)},$$

which we can place inside of L, and this is our finite extension.

In the reverse direction, if $\alpha \in M$, for $[M : K]$ finite, then the infinitely many elements

$$1, \alpha, \alpha^2, \dots$$

cannot all be linearly independent, so there is some linear relation present here. ∎

To count degrees, we have the following.

Proposition 5.14. Suppose we have a tower of fields $M/L/K$. Then

$$[M : K] = [M : L][L : K].$$

In other words, the degree is multiplicative.

Proof. In brief, the idea is to pick a basis $\{a_k\}_{k=1}^m$ for L/K and a basis $\{b_\ell\}_{\ell=1}^n$ for M/L, and we can check that the $\{a_k b_\ell\}$ are a basis for M/K. This gives the result because there are $[M : L] \cdot [L : K]$ basis elements. ∎

This gives us the following.

Proposition 5.15. If $\alpha, \beta \in L$ are algebraic over a base field K, then $\alpha + \beta, \alpha\beta, \alpha - \beta, \alpha/\beta$ are also all algebraic, where the last case requires $\beta \neq 0$.

Proof. The point is that we have the following tower of fields.

$$\begin{array}{c} K[\alpha, \beta] \\ | \\ K[\alpha] \\ | \\ K \end{array}$$

The degree of $[K[\alpha]:K]$ is finite by hypothesis, and the degree of $[K[\alpha,\beta]:K[\alpha]]$ is less than $[K[\beta]:K]$ by checking the polynomial, so the entire extension is going to be finite with degree bounded above by

$$[K[\alpha,\beta]:K] = [K[\alpha,\beta]:K[\alpha]]\cdot[K[\alpha]:K] \le [K[\beta]:K]\cdot[K[\alpha]:K].$$

This gives the result. ■

It is actually quite hard to find these polynomials, which is why we are giving these abstract degree arguments.

Example 5.16. We could try to find the irreducible polynomial

$$\sqrt{2}+\sqrt[3]{2}+\sqrt[5]{2},$$

but it is of degree 30.

Here are some open problems.

Example 5.17. We don't know if either $e+\pi$ or $e\pi$ is algebraic, and you'll be very famous if you can solve either of them. Let's solve one of them, but we won't know which. Indeed, e and π are roots of the polynomial

$$x^2-(e+\pi)x+e\pi.$$

So if $e+\pi$ and $e\pi$ were both algebraic, then we could use the following statement to conclude that both e and π would have to be algebraic, which is false.

Remark 5.18. This argument really annoys intuitionist/constructivist mathematicians because technically we haven't actually showed either $e+\pi$ or $e\pi$ is algebraic.

Proposition 5.19. Fix $p(x)\in L[x]$. If the coefficients of p are algebraic over K, then the roots of polynomial are also algebraic.

Proof. Fix

$$p(x)=\sum_{k=0}^{n} a_k x^k.$$

Then let α be a root off p, and we see that the chain

$$K\subseteq K[a_0]\subseteq K[a_0,a_1]\subseteq\cdots\subseteq K[a_0,\dots,a_n]\subseteq K[a_0,\dots,a_n,\alpha]$$

is a finite chain of finite extensions (the last extension is finite because $p\in K[a_0,\dots,a_n]$, so we are only adjoining the root α here), so the entire extension is finite, so the final root α in the last field is algebraic. ■

Remark 5.20. Again, it is difficult to find the polynomial in the above argument. For example, we won't try to find the explicit polynomial for a root for $x^3-\sqrt{3}x+\sqrt{5}$.

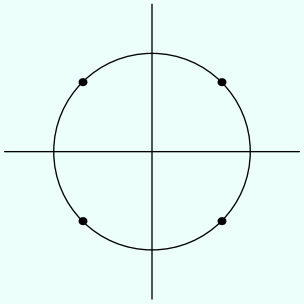
5.1.3 Splitting Fields

Here, suppose that $\pi\in K[x]$ is an irreducible polynomial. Then, looking in

$$L=\frac{K[x]}{(\pi)},$$

we note that π has a root in L, but does it fully factor?

Example 5.21. Take $K = \mathbb{Q}$ and $p(x) = x^4 + 1$. We can check that p is irreducible because $p(x+1) = x^4 + 4x^3 + 6x^2 + 4x + 2$ satisfies Eisenstein's criterion. Now, the roots of p are the primitive 8th roots of 1, roughly graphed as follows.



So if we let $L = \mathbb{Q}(\zeta) \cong K[x]/(p)$, then we see that the roots of p are simply $\zeta, \zeta^3, \zeta^5, \zeta^7$, so indeed we get all of our roots.

Example 5.22. Take $K = \mathbb{Q}$ and $p(x) = x^3 - 2$, which is irreducible because it has no linear factor. Taking $\sqrt[3]{2}$ to be one of its roots, we have that

$$L \coloneqq \mathbb{Q}(\sqrt[3]{2}) \cong \frac{K[x]}{(p)}$$

does not contain the other roots of $x^3 - 2$. Explicitly, the other roots are $\omega\sqrt[3]{2}$ and $\omega^2\sqrt[3]{2}$ where ω is a primitive third root of unity, but $\mathbb{Q}(\sqrt[3]{2}) \subseteq \mathbb{R}$ cannot contain those complex roots.

In the above example, we can manifest the problem by writing

$$x^3 - 2 = \left(x - \sqrt[3]{2}\right)\left(x^3 + \sqrt[3]{2} - \sqrt[3]{4}\right)$$

as our irreducible factorization in $x^3 - 2$ in $\mathbb{Q}(\sqrt[3]{2})$.

We would like our polynomials to fully factor, so we have the following definition.

Definition 5.23 (Splittin field)**.** A *splitting field* of a polynomial $p \in K[x]$ is an extension L/K such that p splits into linear factors in L, and L is actually generated by these roots.

Remark 5.24. We should probably call this a splitting extension, but so it goes.

The main theorem here is as follows.

Theorem 5.25. Splitting fields exist and are isomorphic as K-extensions. In other words, given two splitting fields L_1 and L_2, there is a field isomorphism $L_1 \cong L_2$, and this field isomorphism is also a K-linear map.

Example 5.26. Given $x^4 + 1 \in \mathbb{Q}[x]$, we have that $\mathbb{Q}(\zeta_8)$, where ζ_8 is our primitive eight root of unity, is our splitting field.

Example 5.27. Given $x^3 - 2 \in \mathbb{Q}[x]$, we have that $\mathbb{Q}(\sqrt[3]{2})$ is not out splitting field because we are still missing the roots $\omega\sqrt[3]{2}$ and $\omega^2\sqrt[3]{2}$ in this extension. Namely, we still have to factor the quadratic polynomial in the factorization

$$x^3 - 2 = \left(x - \sqrt[3]{2}\right)\left(x^2 + \sqrt[3]{2}x - \sqrt[3]{4}\right),$$

which we do by looking at

$$\frac{\mathbb{Q}[\sqrt[3]{2}][x]}{\left(x^2 + \sqrt[3]{2}x - \sqrt[3]{4}\right)},$$

which is now a perfectly fine splitting field.

This last example gives the idea behind the proof of Theorem 5.25.

Proof of existence in Theorem 5.25. We proceed inductively; set $K_0 \coloneqq K$ and $p \coloneqq p_0$. We start with some polynomial p. If it has no irreducible factors of degree larger than 1, then we are done. Otherwise, fix $\pi_1 \in K_0[x]$ an irreducible factor of p of degree larger than 1. Now we look at

$$K_1 \coloneqq \frac{K[\alpha_1]}{\pi_1(\alpha)}.$$

Now we can factor p with at least one root α from π, so p will at least partially factor in K_1. So we can factor

$$p(x) = (x - \alpha_1)p_1(x),$$

and because $\deg p_1 < \deg p_0$. Repeating this process (find an irreducible factor π_2 of p_1, and then look at $K_2 \coloneqq K_1[\alpha_2]/(\pi_2)$, and continue) makes the degree continue to decrease, so we finish by induction. At each point we are only adding roots to K_0, so we have that this field we made is also only generated by the roots of p. ∎

Remark 5.28. The above proof in fact gives us a bound of $n!$ for the degree of the splitting field. To be explicit, the extension $[K_{m+1} : K_m]$ will have degree at most $n - m$ because the polynomial p_m at this step has degree $n - m$.

And here is our uniqueness.

Proof of uniqueness in Theorem 5.25. Suppose that we have a splitting field L', and we build our own splitting field using the above algorithm with the chain

$$K \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq L.$$

Now the point is that $K_1 \cong K_0[\alpha_1]/(\pi_1)$ has a root in L', and we can send this α_1 to L' to find a subfield of L isomorphic to K_1. Continuing this process will eventually give us an embedding $L \hookrightarrow L'$.

Building a similar chain for L' (via the roots generating L') will induce an embedding $L' \hookrightarrow L$, so $[L : K] = [L' : K]$ follows, forcing the $L \hookrightarrow L'$ to be bijective and hence an isomorphism. This finishes. ∎

Remark 5.29. The above proof of uniqueness is somewhat problematic because the isomorphism between splitting fields is not unique, which turns out to cause problems. For example, suppose the math department denotes $\mathbb{C}$ by $\mathbb{R}[i]$, and the engineering department denotes this by $\mathbb{R}[j]$, and the chemistry department denotes this by $\mathbb{R}[k]$. The issue is that it is very possible for $i = -j$, and $j = -k$, but then we need to have $k = i$, even though there is another isomorphism $(k \mapsto -i)$ present.

Remark 5.30. The point here is that having a unique isomorphisms are very nice.

5.1.4 Finite Fields

Now let's do number theory because why else would we study algebra? We start with the following small step.

Proposition 5.31. Any finite field F contains some finite field $\mathbb{F}_p$, for a unique prime p.

Proof. Look at the image of the map $\mathbb{Z} \to F$. The kernel here must be a prime ideal because it quotients into an integral domain, so it is either (0) or (p), but (0) would force F to be infinite. So we have an embedding of $\mathbb{Z}/(p) \hookrightarrow F$. ■

So F contains $\mathbb{F}_p$ of finite degree say n, so F will be some n-dimensional vector space, so it will have $q \coloneqq p^n$ elements.

Now, the main statement is as follows.

Theorem 5.32. For each prime-power q, there is one finite field of order q, up to (non-unique) isomorphism.

Proof. The point is that F, being an finite field of order p^n, is equivalent to being the splitting field of the polynomial $x^{p^n} - x$. This will make the given statement follow from existence and uniqueness of splitting fields.

In one direction, fix F a splitting field of $f(x) \coloneqq x^{p^n} - x$. We would like to show that F has order p^n. For this, we show that

$$F = \left\{ \alpha : \alpha \text{ is a root of } x^{p^n} - x \right\}.$$

Surely $f(x) = x^{p^n} - x$ has p^n roots because $f'(x) = -1$, so f has no multiple roots in the algebraic closure. So the above works at least set-theoretically. However, we do need to show that these roots form a subfield structure to show the other inclusion, getting that the field F generated by these roots.

- Given roots α and β, we see that $\alpha\beta$ is a root because $(\alpha\beta)^{p^n} = \alpha^{p^n}\beta^{p^n} = \alpha\beta$.
- For closure under addition, we fix α and β roots, and the point is that

  $$(\alpha+\beta)^p = \sum_{k=0}^{p} \binom{p}{k} \alpha^k \beta^{p-k} = \alpha^p + \beta^p,$$

  where the point is that the middle binomial coefficients vanish $\pmod p$. Repeating this map enough times, we see that

  $$(\alpha+\beta)^{p^n} = \alpha^{p^n} + \beta^{p^n} = \alpha + \beta,$$

  so we have closure under addition.
- 1 and 0 are roots by simply plugging them in.

So indeed, the roots for a subfield of F with p^n elements, so the roots must make up all of F.

Now, in the reverse direction, we need to show that any field F of p^n elements is a splitting field for this polynomial. Well, the point is that each $x \in F$ either has $x = 0$ or $x \in F^\times$ so that

$$x^{p^n - 1} - 1 = 0$$

by Lagrange's theorem. So in all cases, the elements of F are roots of $x^{p^n} - x$, so indeed F will be generated by these roots and is apparently a field. This finishes. ■

Example 5.33. We can find a field of order 2^4 by finding the splitting field of $x^{16} - x$ in $\mathbb{F}_2$. How do we factor this polynomial? Well, it factors as

$$\left(x^4 + x + 1\right)\left(x^4 + x^3 + 1\right)\left(x^4 + x^3 + x^2 + x + 1\right)\left(x^2 + x + 1\right)(x + 1)\,x.$$

We now see that there are three irreducible factors of degree 4, and we notice that the finite field

$$\frac{\mathbb{F}_2[x]}{(x^4 + x + 1)}$$

will have the required dimension.

In general, we see from the above that we are really searching for irreducible polynomials of prescribed degree $\pmod p$. However, proving the existence of such polynomials is somewhat hard.

As an aside, there does not appear to be a "canonical" choice for the irreducible polynomial to construct our finite fields. We could just choose according to lexicographic order, but there is no good reason to do this.

More manifestly, we can see this as the fact that there is no good choice for a square root of -1 in $\mathbb{F}_5$; do we choose 2 or 3? So essentially this is made worse by the fact that even if we were to choose an irreducible polynomial for $\mathbb{F}_{16}$, this might not communicate well with the polynomial generating its $\mathbb{F}_4$ subfield.

We remark that we also have the following statement.

Proposition 5.34. In $\mathbb{F}_p[x]$, we have the irreducible factorization

$$x^{p^n} - x = \prod_{\substack{\pi \text{ irred.} \\ \deg \pi \mid n}} \pi(x).$$

Proof. We have a few things to show here.

- We show that
$$\prod_{\substack{\pi \text{ irred.} \\ \deg \pi \mid n}} \pi(x)$$
divides into $x^{p^n} - x$. Each of these factors are distinct irreducibles and hence coprime, so it suffices to show that, if $\pi \in \mathbb{F}_p[x]$ is an irreducible polynomial of degree $d \mid n$, then $\pi(x) \mid x^{p^n} - x$. Indeed, we see that
$$\frac{\mathbb{F}_p[x]}{(\pi)}$$
is a field with $\mathbb{F}_{p^d}$ elements and hence a subfield of our field $\mathbb{F}_{p^n}$. More explicitly, elements which are roots of π will be roots of $x^{p^d} - x$, which turn into roots of $x^{p^n} - x$ by taking higher powers.
It follows that all roots of π in the algebraic closure are roots of $x^{p^n} - x$. Thus, $\gcd\left(\pi(x), x^{p^n} - x\right) \in \mathbb{F}_p[x]$ will be a polynomial with nonzero degree dividing π, so it must be equal to π, so it follows $\pi \mid x^{p^n} - x$.
- We can compute the exponent of each irreducible π with $\deg \pi \mid n$ dividing into $x^{p^n} - x$. Indeed, we recall $x^{p^n} - x$ has all of its roots of multiplicity 1, so π cannot have multiplicity greater than 1 dividing into $x^{p^n} - x$.
- Lastly, we classify the irreducibles dividing into $x^{p^n} - x$. Namely, if π is some irreducible dividing $x^{p^n} - x$, then we show that the $d := \deg \pi$ must divide into n. Indeed, fixing any root α of π, we see that α is a root of $x^{p^n} - x$, so $\alpha \in \mathbb{F}_{p^n}$. But also
$$\mathbb{F}_p[\alpha] \cong \frac{\mathbb{F}_p[x]}{(\pi)}$$
is a field of size p^d, so it follows from degree arguments that $d \mid n$. ∎

This lets us answer fun questions.

Example 5.35. We can compute the number of irreducible polynomials N_d of degree d in $\mathbb{F}_2[x]$. For example, summing over the degrees given in the factorization of Proposition 5.34, we have

$$2^6 = 6N_6 + 3N_3 + 2N_2 + 1N_1,$$

which gives $N_6 = 9$. One could imagine doing this recursively to solve for the number of irreducibles of degree 6.

Remark 5.36 (Nir)**.** More generally, in $\mathbb{F}_p[x]$, we can let N_d be the umber of irreducible polynomials of degree d so that the factorization in Proposition 5.34 implies

$$p^n = \sum_{d|n} dN_d.$$

Applying Möbius inversion to this implies the "prime number theorem in $\mathbb{F}_p[x]$" by

$$N_n = \frac{1}{n}\sum_{d|n} \mu(d)p^{n/d} = \frac{p^n}{n} + O\left(\frac{p^{n/2}}{n}\right).$$

We also remark that, if d is the largest squarefree divisor of n (so that all squarefree divisors of n divide into d), then

$$nN_n \equiv \mu(d)p^{n/d} \not\equiv 0 \pmod{p^{n/d+1}}$$

because all other terms of the sum will vanish. It follows there is indeed an irreducible of degree n in $\mathbb{F}_p[x]$. (One could also see this by directly bounding the sum for N_n by $\frac{1}{n}\left(p^n - \sum_{d=1}^{n-1} p^{n/d}\right) > 0$.)

We close with a remark.

Remark 5.37. We are able to construct a splitting field of any finite set of polynomials, simply by iterating. We can extend to a countable set of polynomials using a transfinite induction (read: Zorn's lemma). For example, if we take the splitting field of all polynomials, we get the algebraic closure of our field.

5.2 November 9

Despite the severity of his injury, the child was conscious, and in terrible pain.

5.2.1 Algebraic Closure

Let's quickly finish this off so that we can talk about Galois extensions. Briefly recall that we have a notion of "splitting field."

Definition 5.38 (Splitting field)**.** Given a set of polynomials $\{p_\alpha\}_{\alpha\in\lambda} \in K[x]$, the *splitting field* L/K is a field in which all the p_k split fully into linear factors, and the corresponding roots generate the field L.

We saw last lecture that splitting fields exist and are unique up to (non-canonical/non-unique) isomorphism.

Remark 5.39. Technically we showed that splitting fields exist for a single polynomial, but this construction can be extended to any set of polynomials by some kind of transfinite induction.

Remark 5.40. The lack of uniqueness of the isomorphism here induces major headaches later in life.

Anyways, we have the following definition.

Definition 5.41 (Algebraic closure)**.** Given a field K, the *algebraic closure* $\overline{K}$ of K is an algebraic extension of K such that all polynomials in $\overline{K}[x]$ will fully factor in $\overline{K}$.

Of course, it is not immediately obvious that such a thing should exist, nor that it is unique up to some isomorphism (which justifies the use of the word "the" in the above definition). Let's see this.

Proposition 5.42. Fix a field K. Then an algebraic closure of K exists and is unique.

We present two proofs of the existence, and we will use the second proof to show uniqueness.

Lazy proof of existence in Proposition 5.42. We start with a lazy proof of existence. Set $K_0 \coloneqq K$ and then define K_1 to be the splitting field of the set of all polynomials in $K_0[x]$ over K_0. However, we might have $K_1 \neq \overline{K}$ if there is a polynomial in $K_1[x]$ without roots in K_1, so we inductively define K_{n+1} to be the splitting field of the set of all polynomials in $K_n[x]$ over K_n. This creates the chain

$$K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots.$$

So we claim that we can define

$$\overline{K} \coloneqq \bigcup_{n \geq 0} K_n.$$

We can check that this is a field (closed under addition, multiplication, and inverses) by hand using the chain condition; for example, this is closed under addition because any $\alpha, \beta \in \overline{K}$ have some N for which $\alpha, \beta \in K_N$, so $\alpha + \beta \in K_N \subseteq \overline{K}$.

Remark 5.43. This is a common idea in mathematics: just inductively build up and do a big union to finish.

So now we want to check that $\overline{K}$ is algebraically closed. Well, any polynomial

$$\sum_{k=0}^{n} a_k x^k \in \overline{K}[x]$$

will have some N for which $a_k \in K_N$ for each k because we constructed $\overline{K}$ as a chain, and there are only finitely many of the $a_\bullet$. It follows that

$$\sum_{k=0}^{n} a_k x^k \in K_N[x],$$

so this polynomial fully splits in $K_{N+1} \subseteq \overline{K}$, so indeed, this polynomial fully splits in $\overline{K}$. ∎

An issue with the above proof is that it makes uniqueness a bit difficult to prove uniqueness, and we haven't even showed that $\overline{K}$ defined above is actually algebraic over K.. To solve this, we decide to be a little less lazy.

Proof of Proposition 5.42. To be less lazy, we actually do field theory. The main idea is the following lemma, which essentially reduces the check of algebraic closure to just polynomials in $K[x]$. In the vocabulary of the previous proof, we are essentially saying that K_1 defined above is in fact algebraically closed, so our chain stops after one step.

Lemma 5.44. Fix a field K. A field $\overline{K}$ is an algebraic closure of K if and only if it is a splitting field of the set of all polynomials $K[x]$ over K.

Proof. Define K_1 to be the splitting field of all polynomials $K[x]$ over K. The main point is to recall that K_1 is algebraically closed: for any polynomial

$$p(x) \coloneqq \sum_{k=0}^{n} a_k x^k \in K_1[x],$$

the elements a_k are algebraic over K, so the extension $K(a_0, \dots, a_n)$ is of finite degree over K. Intuitively, any root α of the above polynomial will still have $K(a_0, \dots, a_n, \alpha)$ a finite extension, implying that α is algebraic over K, implying $\alpha \in K_1$. Rigorously, we may (without loss of generality) take p to be irreducible so that

$$\frac{K_1[\alpha]}{(p(\alpha))}$$

is still a field. But now α is a root of the above polynomial, so we can use our intuitive argument so show that α is algebraic over K, so $\alpha \in K_1$. But then $(x - \alpha)$ is a factor of $p(x)$, so we must have $(p) = (x - \alpha)$, which makes p fully factor over $K_1[x]$.

So we see that K_1 is in fact algebraically closed, and by construction is algebraic over K. So indeed, K_1 is an algebraic closure of K. Now fix $\overline{K}$ any algebraic closure of K. Certainly $\overline{K}$ must contain all the roots of polynomials in $K[x] \subseteq \overline{K}[x]$, so there is a subfield

$$\overline{K}_1 \subseteq \overline{K}$$

generated by these roots; i.e., $\overline{K}_1$ is a splitting field of the set of all polynomials in $K[x]$. However, $\overline{K}$ is an algebraic extension of K, so all elements of K are roots of some polynomial in $K[x]$, so we also get $\overline{K} \subseteq \overline{K}_1$. So we see

$$\overline{K} = \overline{K}_1 \cong K_1,$$

where the isomorphism is by uniqueness of the splitting field. This finishes. ∎

So we see that uniqueness and existence of splitting fields establishes the existence and uniqueness of the algebraic closure automatically. So we are done. ∎

Examples of the algebraic closure are somewhat annoying to look directly at, for example because the splitting field of so many polynomials is a bit annoying to keep track of.

Example 5.45. The complex numbers $\mathbb{C}$ are an algebraic closure of $\mathbb{R}$, which we'll prove later in an algebraic way.

Example 5.46. The field $\overline{\mathbb{Q}}$ of algebraic numbers, which are the elements of $\mathbb{C}$ algebraic over $\mathbb{Q}$, is the algebraic closure of $\mathbb{Q}$.

Example 5.47. The field of Laurent power series $\mathbb{C}((t))$ with coefficients in $\mathbb{C}$ is not algebraically closed, but its algebraic closure is

$$\bigcup_{n \geq 1} \mathbb{C}((t^{1/n})).$$

This result is more or less due to Newton, who gave an algorithm to solve polynomials in $\mathbb{C}((t))$ which implies that the above is algebraically closed.

Example 5.48. The algebraic closure $\overline{\mathbb{F}_p}$ of $\mathbb{F}_p$ is more or less the infinite union

$$\bigcup_{n \geq 1} \mathbb{F}_{p^n},$$

which is actually a direct limit with the embeddings $\mathbb{F}_{p^k} \hookrightarrow \mathbb{F}_{p^{k\ell}}$. However, the non-uniqueness of these embeddings makes this description annoying to work with.

5.2.2 Galois Advertisement

We're going to build towards Galois extensions.

Idea 5.49. A *Galois extension* of fields L/K is an extension which is "as symmetric as possible." For an extension L/K, we may define $\mathrm{Gal}(L/K)$ as the set of automorphisms of L fixing K, and it will happen that $\mathrm{Gal}(L/K)$ "controls" the extension.

As an example, subgroups of the Galois group will correspond with intermediate extensions.

Anyways, let's see a definition.

Definition 5.50 (Galois extension, I)**.** An extension L/K is *Galois* if and only if it is normal and separable.

Wait, what do "normal" and "separable" mean?

5.2.3 Normal Extensions

We have the following definition.

Definition 5.51 (Normal extension)**.** An algebraic extension L/K is *normal* if and only if every irreducible polynomial in $K[x]$ which has a root in L will fully split into linear factors in L.

Remark 5.52 (Nir)**.** Here is another way to state this definition: fixing some embedding $K \hookrightarrow \overline{K}$, any embedding $\sigma : L \hookrightarrow \overline{K}$ is actually an embedding into L. Indeed, any $\alpha \in L$ with irreducible polynomial $\pi \in K[x]$ will have

$$\pi(\sigma\alpha) = \sigma(\pi(\alpha)) = 0,$$

so $\sigma\alpha$ is another root of π (!). But π, with one root in L, will fully split in L because L/K is normal, so $\sigma\alpha \in L$. It follows that $\sigma : L \hookrightarrow \overline{K}$ does indeed restrict to L.

In fact, σ is also an automorphism of L: it remains to check that σ is surjective. Well, all the roots of π live in L as discussed, so π restricts to an injective mapping of the set of roots of π to itself, which is bijective because there are only finitely many roots. In particular, there is an element $\beta \in L$ mapping to our $\alpha \in L$.

Remark 5.53 (Nir)**.** In fact, the converse of the above remark is also true: suppose all embeddings $\sigma : L \hookrightarrow \overline{K}$ fixing K actually output into L. Now, fix any irreducible polynomial $\pi \in K[x]$ with a root $\alpha \in L$; we show that all roots of π in $\overline{K}$ (where π certainly fully splits) are in fact elements of L.

Well, if $\beta \in \overline{K}$ is a root of π, then there is an embedding fixing K given by

$$K(\alpha) \cong \frac{K[x]}{(\pi)} \cong K(\beta) \hookrightarrow \overline{K}.$$

By a Zorn's lemma argument, we can extend (!) this up to $L \hookrightarrow \overline{K}$ (even when L is of infinite degree over K), so we have an embedding $L \hookrightarrow \overline{K}$ fixing K sending $\alpha \mapsto \beta$. However, this embedding must output into L, so $\beta \in L$.

It is hard to prove that a particular extension is normal from the above definition because checking all the irreducibles in $K[x]$ is difficult; however, here are some examples.

Example 5.54. Fix L/K an algebraic extension of degree 2. Now suppose $f \in K[x]$ is irreducible and has a root in L. However, $[L : K] = 2$ implies that $\deg f \leq 2$, so f has at most 2 roots, and the sum of the two roots is an element of K by Vieta's formulae, so the other root will still be in L.

Non-Example 5.55. The extension $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is not normal because $x^3 - 2$ has one root in $\mathbb{Q}(\sqrt[3]{2})$ but not all roots. Namely, the other roots of $x^3 - 2$ are not real and so do not live in $\mathbb{Q}(\sqrt[3]{2})$.

Luckily, there is an easier classification of normal extensions.

Proposition 5.56. Fix L/K an algebraic extension. Then L/K is a normal extension if and only if L is the splitting field of some set of polynomials.

Proof. We show the directions one at a time.

- Fix L/K a normal extension and $\overline{K}$ an algebraic closure of K with a chosen embedding $L \subseteq \overline{K}$. (The point of doing this is so that we don't need to worry about uniqueness of isomorphisms of splitting fields anymore.) The main idea is to look at

  $$S \coloneqq \{\pi \in K[x] : \pi \text{ is irreducible and has a root in } L\}.$$

  Now set $L' \subseteq \overline{K}$ equal to the splitting field of S over K; we claim that $L = L'$. Because L/K is an algebraic extension, all elements of L are the root of some irreducible polynomial over K, so L is certainly a subset of L'.

  But conversely, any irreducible polynomial $\pi \in K[x]$ with a root in L will fully split in L and in particular have all of its roots in L, so L will contain all the generators of the splitting field of S over K. Thus, L contains L', finishing.

- Fix L/K a splitting field of some set of polynomials S. Additionally, fix $\overline{K}$ some algebraic closure of K with chosen embedding $L \subseteq \overline{K}$; the main point is to show that any embedding $\sigma : L \hookrightarrow \overline{K}$ fixing K actually embeds into L, which is implies L/K is normal by Remark 5.53.

  Well, fixing some polynomial $p \in S$, we note that if $\alpha \in L$ is a root of p, then

  $$p(\sigma\alpha) = \sigma(p(\alpha)) = 0,$$

  so σ takes roots of p to roots of p. However, L has all the roots of p in its list of generators over K, so $\sigma\alpha \in L$.

  Thus, $\sigma : L \hookrightarrow \overline{K}$ sends the generators of L into L, so it follows that σ just sends L to L because the fact σ is a homomorphism means that any expression involving the generators of L will still go to L. So indeed, any embedding $L \hookrightarrow \overline{K}$ outputs into L, finishing. ∎

The point of the above proposition is that it lets us construct lots and lots of normal extensions: choose your favorite polynomial and then look at its splitting field.

Non-Example 5.57. The extension $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is not a normal extension, even though we can write down

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt[4]{2}),$$

which is a chain of degree-2 and hence normal extensions. Namely, $x^4 - 2$ has $i\sqrt[4]{2} \notin \mathbb{Q}(\sqrt[4]{2}) \subseteq \mathbb{R}$ as a root, which is a problem.

Remark 5.58. The word "normal" in Galois theory will turn into normal subgroups of the Galois group, which is nice. For example, any degree-2 extension being normal corresponds to the statement that any subgroup of index 2 is normal. And the above non-example corresponds to the statement that a chain of normal subgroups

$$A \subseteq B \subseteq C$$

does not necessarily have $A \subseteq C$ normal. For example, $\mathbb{Z}/2\mathbb{Z} \subseteq (\mathbb{Z}/2\mathbb{Z})^2 \subseteq D_8$.

5.2.4 Separable Extensions

Let's talk about separable extensions next. We have the following definition.

Definition 5.59 (Separable)**.** Fix L/K an algebraic extension. Then $\alpha \in L$ is *separable* if and only if its irreducible polynomial $\pi \in K[x]$ is *separable*, which means π has no multiple roots. Then L/K is *separable* if and only if each element $\alpha \in L$ is separable.

The following statements show that most fields we care about will have separable extensions.

Exercise 5.60. If L/K is algebraic extension of fields with characteristic 0, then L/K is separable.

Proof. Fix $\pi \in K[x]$ the irreducible polynomial of any $\alpha \in L$. Then we recall that π has multiple roots if and only if π and π' have a nonconstant common factor $g \mid \gcd(\pi, \pi')$. However, in characteristic zero, we have that

$$\deg \pi' = \deg \pi - 1,$$

so in particular any common factor $g \mid \gcd(\pi, \pi')$ has $\deg g < \deg \pi$. Thus, g is a factor of π of smaller degree but the irreducibility of π forces g to be constant. So indeed, π and π' have no nonconstant common factors, so π has no multiple roots. ∎

Exercise 5.61. If L/K is an extension of finite fields, then the extension is separable.

Proof. If $\#L = q$, then we see that L consists of the roots of $x^q - x = 0$. In particular, for any $\alpha \in L$ with irreducible polynomial $\pi \in K[x]$, we showed last time that $\pi(x) \mid x^q - x$. So multiple roots of π would induce multiple roots of $x^q - x$, but $x^q - x$ has no multiple roots because its derivative is

$$qx^{q-1} - 1 = -1$$

in $L[x]$, so $x^q - x$ has no common factors with its derivative. ∎

In the early days of field theory, the above were our only examples, but inseparable extension do exist!

Non-Example 5.62. Consider the field $L \coloneqq \mathbb{F}_p(t)$ of rational functions over $\mathbb{F}_p$, and set $K \coloneqq \mathbb{F}_p\left(t^p\right)$ so that $[L : K] = p$; in particular, we have the power basis $\left\{1, t, \dots, t^{p-1}\right\}$ for L/K. Now we see that t is the root of the polynomial

$$x^p - t^p \in K[x],$$

which must be our minimal and hence irreducible polynomial because it is has the correct degree of p. However,

$$x^p - t^p = (x - t)^p,$$

so $x^p - t^p$ has multiple roots at t up in L.

Remark 5.63. Essentially all problems in positive characteristic come from this example of inseparable extensions.

Like with normal extensions, we would like a nice classification of separable extensions; here it is.

Proposition 5.64. Fix L/K a finite algebraic extension. Then the following are equivalent.

(a) L is generated by separable elements of K.

(b) The embedding $K \hookrightarrow \overline{K}$ into the algebraic closure has exactly $[L : K]$ extensions $L \hookrightarrow \overline{K}$.

(c) All elements of L/K are separable; i.e., L/K is a separable extension.

Essentially the above shows that we can check separable extensions by only checking if a set of generating elements are separable, which is nice.

Proof. We show our implications separately.

- That (c) implies (a) is because generators are elements.
- For (a) implies (b), pick up some $\alpha \in L$ a separable element. We consider the chain

  $$K \subseteq K(\alpha) \subseteq L.$$

  Now suppose $n \coloneqq [K(\alpha) : K]$ and π is the irreducible polynomial of α. We see that there are exactly $n = \deg \pi$ extensions of $K \hookrightarrow \overline{K}$ to $K(\alpha) \hookrightarrow \overline{K}$: there are at most that many we can have to send α to some root of π, of which there are n, and each of these defines at most one mapping $K(\alpha) \hookrightarrow \overline{K}$. But each choice of root β of π does indeed induce an embedding

  $$K[\alpha] \cong \frac{K[x]}{(\pi)} \overset{x \mapsto \beta}{\hookrightarrow} \overline{K},$$

  where the embedding is well-defined because $x \mapsto \beta$ induces a map $K[x] \to \overline{K}$ with kernel (π). So indeed, there are n extensions of $K \hookrightarrow \overline{K}$ to $K(\alpha) \hookrightarrow \overline{K}$, for any separable element $\alpha \in L$.

  So we assert, as promised, that

  $$L = K(\alpha_1, \dots, \alpha_m),$$

  where $\{\alpha_1, \dots, \alpha_n\}$ are separable. Now we consider the chain of fields

  $$K \subseteq K(\alpha_1) \subseteq K(\alpha_1, \alpha_2) \subseteq \cdots \subseteq K(\alpha_1, \dots, \alpha_m) = L$$

  and inductively count the number of embeddings into $\overline{K}$. Simply extending automorphisms one at a time gives

  $$[K(\alpha_1) : K] \cdot [K(\alpha_1, \alpha_2) : K] \cdot \ldots \cdot [L : K(\alpha_1, \dots, \alpha_{m-1})] = [L : K]$$

  total embeddings, where we have applied the tower law. (These embeddings are distinct because distinct extensions of $K(\alpha_1, \dots, \alpha_\ell) \hookrightarrow \overline{K}$ to $K(\alpha_1, \dots, \alpha_{\ell+1}) \hookrightarrow \overline{K}$ will send $\alpha_{\ell+1}$ different places, so we can track our automorphisms by where they send the generators.)

But these inductively constructed embeddings are in fact all of our embeddings: any embedding $L \hookrightarrow \overline{K}$ will induce embeddings $K(\alpha_1, \dots, \alpha_\ell) \hookrightarrow \overline{K}$ which extend into each other, so it must come from the above extending process. So indeed, there are exactly $[L : K]$ total embeddings $L \hookrightarrow \overline{K}$.

Remark 5.65. The above argument can more generally show that, for a given finite algebraic extension L/K and another field M, there are at most $[L : K]$ extensions of some $K \hookrightarrow M$ to L. Namely, for a fixed element $\alpha \in L$, removing the condition that α is separable implies that there are at most $[K(\alpha) : K]$ extensions of $K \hookrightarrow M$ to $K(\alpha) \hookrightarrow M$ because there are at most $[K(\alpha) : K]$ roots for α to go to.

Then the tower law argument still applies, but now it only shows there are at most

$$[K(\alpha_1) : K] \cdot [K(\alpha_1, \alpha_2) : K] \cdot \ldots \cdot [L : K(\alpha_1, \dots, \alpha_{m-1})] = [L : K]$$

extensions of $K \hookrightarrow M$ to $L \hookrightarrow M$.

- For (b) implies (c), fix $\alpha \in L$ so that we want to show α is separable. We again focus on the chain

$$K \subseteq K(\alpha) \subseteq L.$$

By hypothesis, there are exactly $[L : K]$ embeddings $L \hookrightarrow \overline{K}$. Further, we see that there are at most $[L : K(\alpha)]$ extensions of a chosen embedding $K(\alpha) \hookrightarrow \overline{K}$ to $L \to \overline{K}$ by Remark 5.65. So there are at least

$$\frac{[L : K]}{[L : K(\alpha)]} = [K : K(\alpha)]$$

embeddings $K(\alpha) \hookrightarrow \overline{K}$ if we are to be able to extend these automorphisms to $[L : K]$ total embeddings $L \hookrightarrow \overline{K}$.

However, an embedding $K(\alpha) \hookrightarrow \overline{K}$ must send α to some root of π, and the embedding is completely determined by where it sends α, so the fact there are at least $\deg \pi = [K(\alpha) : K]$ embeddings implies that there are at least $\deg \pi$ distinct roots of π. So there are exactly $\deg \pi$ distinct roots of π by Lagrange's theorem on polynomials. ∎

One of the major headaches with the above proofs is that our finite extensions are often generated by many elements, which means we are forced to look at chains of fields. Life would be easier if our extensions were generated by single elements, and it turns out being separable is the correct condition.

Non-Example 5.66. Consider the extension $L \coloneqq \mathbb{F}_p(t, u)$ of rational functions of two variables over $\mathbb{F}_p$. Then we let $K \coloneqq \mathbb{F}_p\left(t^p, u^p\right)$. Now, L/K is an extension of degree p^2, but for any $f \in L$, we have $f^p \in K$ by the Frobenius automorphism, so the degree of $[K(x) : K]$ is at most p. It follows that $K(x) \neq L$ for any $x \in L$.

The issue above is that L/K is an inseparable extension, as we discussed earlier. We do get the result for separable extensions.

Theorem 5.67 (Primitive element)**.** If L/K is finite and separable, then there exists $\alpha \in L$ with $L = K(\alpha)$.

Proof. If K is a finite field, then $[L : K] < \infty$ implies L is also finite. So $L^\times$ is a cyclic group (it's a finite multiplicative subgroup of $L^\times$), so choose any generator g of $L^\times$ to give $L = K[g]$.

So now we may assume that K is infinite.

Remark 5.68. It is very strange that we have to talk about finite fields and infinite fields differently, but we will really use that K is infinite below.

Because L/K is a finite extension, we know that L is generated by a finite number of elements. So by induction, it suffices to show that $L := K(\alpha, \beta)$ is generated by a single element. (In particular, the induction functions because intermediate extensions of separable extensions are separable.)

The main idea, now, is to study embeddings $K(\alpha,\beta) \hookrightarrow \overline{K}$. For any distinct maps $\sigma, \tau : K(\alpha,\beta) \hookrightarrow \overline{K}$, we claim that there is at most one $c \in K$ giving

$$\sigma(\alpha + c\beta) = \tau(\alpha + c\beta).$$

Indeed, this is because the previous equation implies

$$(\sigma\beta - \tau\beta)c = \tau\alpha - \sigma\alpha.$$

Now, we have two cases.

- If $\sigma\beta = \tau\beta$, then we must have $\sigma\alpha \neq \tau\alpha$ if we are to have $\sigma \neq \tau$, so in this case the given equation will have no solutions.
- If $\sigma\beta \neq \tau\beta$, then we simply solve
$$c = \frac{\tau\alpha - \sigma\alpha}{\sigma\beta - \tau\beta}$$
as our only solution for c.

Because there are only finitely many embeddings $K(\alpha,\beta) \hookrightarrow \overline{K}$ (in particular, at most $[K(\alpha,\beta) : K]$), it follows that we can find $c \in K$ such that

$$\sigma(\alpha + c\beta) \neq \tau(\alpha + c\beta)$$

for each pair of distinct embeddings $\sigma, \tau \in K(\alpha,\beta) \hookrightarrow \overline{K}$. Indeed, each such pair throws out at most one element of K, but K is infinite (!), so we must have elements $c \in K$ left over.

In particular, this implies that $\alpha + c\beta$ has at least $[K(\alpha,\beta) : K]$ distinct images under embeddings into $\overline{K}$—here we are using the fact that $K(\alpha,\beta)/K$ is separable to imply there are $[K(\alpha,\beta) : K]$ distinct embeddings $K(\alpha,\beta) \hookrightarrow \overline{K}$. Thus, by tracking the generator there are at least $[K(\alpha,\beta) : K]$ embeddings

$$K(\alpha + c\beta) \hookrightarrow \overline{K}.$$

By Remark 5.65, we see that $[K(\alpha + c\beta) : K]$ is at least the number of embeddings $K(\alpha + c\beta) \hookrightarrow \overline{K}$, so $[K(\alpha + c\beta) : K] \geq [K(\alpha,\beta) : K]$ by chaining inequalities. But of course $K(\alpha + c\beta) \subseteq K(\alpha,\beta)$, so $K(\alpha + c\beta) = K(\alpha,\beta)$ follows. This finishes the proof. ∎

5.2.5 Galois Extensions

So lastly let's talk about Galois extensions. We have the following definition.

Definition 5.69 (Galois group)**.** Fix L/K a finite extension of fields. Then we define the *Galois group* of L/K

$$\mathrm{Gal}(L/K) := \{\sigma \in \mathrm{Aut}(L) : \sigma|_K = \mathrm{id}_K\}$$

to be automorphisms of L fixing K.

Checking that $\mathrm{Gal}(L/K)$ is actually a group is as usual: we need to show that $\mathrm{Gal}(L/K)$ is a subgroup of $\mathrm{Aut}(L)$, for which it suffices to see that $\mathrm{id}_L \in \mathrm{Gal}(L/K)$ because $\mathrm{id}_L|_K = \mathrm{id}_K$ and $\sigma, \tau \in \mathrm{Gal}(L/K)$ implies that $(\sigma\tau^{-1})|_K = \sigma|_K \cdot (\tau|_K)^{-1} = \mathrm{id}_K$.

Anyways, here are some examples.

Example 5.70. The Galois group $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ is simply $\mathrm{id}_{\mathbb{C}}$ and $z \mapsto \overline{z}$. Essentially this is why complex conjugation is so important in analysis.

Example 5.71. The Galois group $\mathrm{Gal}(\mathbb{F}_4/\mathbb{F}_2)$ has the nontrivial automorphism $x \mapsto x^p$ where $p := 2$ which is an automorphism because $(a+b)^p = a^p + b^p$ and $(ab)^p = a^p b^p$, and we see that $(a^p)^p = a^{p^2} = a$, so we have injectivity and hence surjectivity. It follows we have at least 2 elements.

The above examples technically only exhibit elements of the Galois group without showing that we have found all of them; the following bound establishes that the above examples do indeed find the entire Galois group.

Proposition 5.72. Fix L/K a finite and hence algebraic extension. Then we have that $\#\,\mathrm{Gal}(L/K) \le [L : K]$.

Proof. We could apply another chain argument, where we set $L = K(\alpha_1, \dots, \alpha_n)$ for some $\{\alpha_k\}_{k=1}^n \subseteq L$ and consider the chain

$$K \subseteq K(\alpha_1) \subseteq K(\alpha_1, \alpha_2) \subseteq \cdots \subseteq K(\alpha_1, \dots, \alpha_n) = L.$$

Inductively considering the number of extensions, there are at most $[L : K]$ total extensions by the tower law and using the argument from earlier.

Alternatively, we could optimize out the chain argument: Remark 5.65 implies that there are at most $[L : K]$ extensions of $K \hookrightarrow M$ up to $L \hookrightarrow M$, so setting $M = L$ implies that there are most $[L : K]$ extensions of $K \hookrightarrow L$ up to $L \hookrightarrow L$. In other words, there are at most $[L : K]$ automorphisms of L fixing K. ■

Note that we really can have less than or equal to in this bound.

Example 5.73. We exhibited 2 elements of $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$, so we have found all of them.

Example 5.74. The size of $\mathrm{Gal}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$ is $1 < [\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}]$ because the root $\sqrt[3]{2}$ must stay fixed. Namely, an automorphism $\sigma : \mathbb{Q}(\sqrt[3]{2}) \to \mathbb{Q}(\sqrt[3]{2})$ must send $\sqrt[3]{2}$ to a root of $X^3 - 2$, but the other roots of this polynomial are

$$\zeta_3 \sqrt[3]{2},\ \zeta_3^2 \sqrt[3]{2} \notin \mathbb{R},$$

which are not real and hence not in $\mathbb{Q}(\sqrt[3]{2})$. So σ must fix $\sqrt[3]{2}$, so σ must be the identity on $\mathbb{Q}(\sqrt[3]{2})$.

The issue of the previous example is that $\mathbb{Q}(\sqrt[3]{2})$ cannot see the other roots of $X^3 - 2$; Galois extensions are defined to nullify this problem. Here is our definition.

Definition 5.75 (Galois extension, II)**.** A finite extension of fields L/K is a *Galois extension* if and only if $\#\,\mathrm{Gal}(L/K) = [L : K]$. Namely, there are as many symmetries as possible.

This definition makes it difficult to tell if a particular extension is Galois. For this, we bring in our machinery.

Proposition 5.76. Fix L/K a finite extension of fields. Then the following are equivalent.

(a) L/K is the splitting field of some separable polynomials.

(b) L/K is normal and separable.

(c) L/K is a Galois extension: $\#\,\mathrm{Gal}(L/K) = [L : K]$.

(d) K is the fixed field of some subgroup $G \subseteq \mathrm{Aut}(L)$.

All of these criteria are giving us nice ways of generating Galois extensions.

Remark 5.77. Many books define Galois as being normal and separable, though this is somewhat unintuitive because the two definitions seems somewhat orthogonal. We have defined as above so that Galois means "the most symmetric possible," which is a bit more motivated according to Professor Borcherds.

Proof of Proposition 5.76. We take these one at a time.

- For (a) implies (b), we showed that splitting fields are normal, and roots of separable polynomials will generate separable extensions, so this follows.
- For (b) implies (c), we start by seeing $\#\operatorname{Gal}(L/K) \le [L:K]$ from the above, so we merely need to exhibit $[L:K]$ different elements of $\operatorname{Gal}(L/K)$.

  Again, we could do a chain argument, but we've done enough theory to be able to optimize it out: because L/K is separable, there are exactly $[L:K]$ extensions of $K \hookrightarrow \overline{K}$ up to $L \hookrightarrow \overline{K}$. Because L/K is normal, Remark 5.52 implies that each embedding $L \hookrightarrow \overline{K}$ is in fact an automorphism of L fixing K. So we have found $[L:K]$ elements of $\operatorname{Gal}(L/K)$.
- For (c) implies (d), we set $L^G \subseteq L$ to be the elements fixed by $G \coloneqq \operatorname{Gal}(L/K)$. The point is that $K \subseteq L^G$ by definition, and we see that we can bound

  $$[L:K] = \#G \le \#\operatorname{Gal}\left(L/L^G\right) \le \left[L:L^G\right],$$

  where we have used the fact that L/K is Galois in the first equality. However, $K \subseteq L^G$ gives $\left[L:L^G\right] \le \left[L:L^G\right]\left[L^G:K\right] = [L:K]$, so in fact $\left[L:L^G\right] = [L:K]$, giving $K = L^G$.
- Lastly, we will show (d) implies (a) at the start of next lecture. ∎

Remark 5.78 (Nir). Technically the "natural" definition Definition 5.75 only works for finite extensions L/K, if $[L:K]$ is to make sense. However, we see from Proposition 5.76 above that we can extend this to all extensions by way of Definition 5.50.

5.3 November 16

I'm gonna die to the sound of that noise.

5.3.1 Galois Loose Ends

Last lecture we were in the middle of proving the following statement.

Proposition 5.76. Fix L/K a finite extension of fields. Then the following are equivalent.

(a) L/K is the splitting field of some separable polynomials.

(b) L/K is normal and separable.

(c) L/K is a Galois extension: $\#\operatorname{Gal}(L/K) = [L:K]$.

(d) K is the fixed field of some subgroup $G \subseteq \operatorname{Aut}(L)$.

Proof. To finish off, we have to show that (d) implies (a). For this, pick some element $\alpha \in L$ and look at the G-conjugates of α, namely $G\alpha = \{g\alpha : g \in G\}$. To find a polynomial with α as a root, we take

$$f_\alpha(x) = \prod_{\beta \in G\alpha} (x - \beta).$$

This polynomial is separable because we took the product over the set of roots in $G\alpha$, so there will be no repetition. Further, notice that the coefficients are fixed by G because we can induce a G-action on $L[x]$ by fixing x, upon which we see

$$g \cdot f_\alpha(x) = \prod_{\beta \in G\alpha} (gx - g\beta) = \prod_{\beta \in G\alpha} (x - g\beta) = \prod_{\beta \in G\alpha} (x - \beta)$$

because the G-action induces a bijection on $G\alpha$. Thus, $f_\alpha(x) \in L^G[x] = K[x]$.

Generating the polynomial f_α for each element $\alpha \in L$ gives the set

$$\{f_\alpha(x) \in K[x] : \alpha \in L\}$$

of separable polynoials, whose splitting field is $K(\{\alpha\}_{\alpha \in L}) = L$. ■

Remark 5.79. In the original statement, it is not at all obvious that any of the above are equivalent, but they are, which is nice.

Anyways, Proposition 5.76 gives us lots of examples of Galois extensions.

Example 5.80. The splitting field of any separable polynomial, as in (a), will form a Galois extension. For example, the splitting field of $x^7 - 3x^4 + 2$ over $\mathbb{Q}$ makes a Galois extension, but actually finding what this splitting field is not easy; e.g., what is the degree? This probably has Galois group S_7, but this is hard to prove.

Example 5.81. We use (d): for example, set $L = \mathbb{Q}(x_1, \dots, x_n)$ to be rational functions in n variables. This has an $G \coloneqq S_n$-action of permuting the coordinates, so we find that L/L^G is a Galois extension. Recall that, by the Fundamental theorem of symmetric polynomials, we have

$$L^G = \mathbb{Q}(e_1, \dots, e_n),$$

where the $e_\bullet$ are the symmetric polynomials. (Explicitly, we see that any element $p/q \in L^G$ can have $p, q \in \mathbb{Q}[e_1, \dots, e_n]$, so $p/q \in \mathbb{Q}(e_1, \dots, e_n)$.)

Example 5.82. In general, take G a finite group, and we see that we can embed $G \subseteq S_n$ with $n = \#G$ (say). So now if we take $L = \mathbb{Q}(x_1, \dots, x_n)$, we find that L/L^G will have Galois group G. Indeed, $G \subseteq \mathrm{Gal}(L/L^G)$ because each element $\sigma \in G$ does act on L in a way fixing L. And conversely, any element $\tau \in \mathrm{Gal}(L/L^G)$ will have to fix

$$\alpha \coloneqq \sum_{\sigma \in G} x_{\sigma 1} \in L^G,$$

which we can see implies that $\tau \in G$. To be explicit, α is fixed by G because the G-action merely permutes the $x_{\sigma 1}$. And $\tau\alpha = \sum_{\sigma \in \tau G} x_{\sigma 1}$ is the sum over a coset, so $\tau\alpha = \alpha$ implies that $\tau \in G$. The point is that any finite group comes from some finite field extension.

As an aside, actually writing down what L^G in Example 5.82 is somewhat difficult (say) in terms of generators. This is more or less the same difficulty we were feeling in Example 5.80: actually describing our extension is hard, though we know it exists.

Remark 5.83. It is an open problem in Galois theory that, given a finite group G, if there exists an extension $K/\mathbb{Q}$ with the prescribed Galois group. If we let the base field vary, the answer yes; if we fix the base field to be $\mathbb{C}(t)$ or $\mathbb{Q}_p(t)$, the answer is still yes.

5.3.2 Advertisement for the Galois Correspondence

Here is the main theorem in Galois theory.

Theorem 5.84. Fix M/K a finite Galois extension of fields with Galois group $G \coloneqq \mathrm{Gal}(M/K)$. Then we get a one-to-one correspondence between intermediate extensions $K \subseteq L \subseteq M$ and subgroups $H \subseteq G$. To be explicit, our maps are as follows.

$$\begin{array}{ccc} M^H & \longmapsto & H \\ L & \longmapsto & \mathrm{Gal}(M/L) \end{array}$$

Additionally, we have $[M : M^H] = \#H$ and $[M : L] = \#\,\mathrm{Gal}(M/L)$. In fact, this mapping is inclusion reversing: if we have subgroups $H_1 \subseteq H_2 \subseteq G$, then $M^{H_2} \subseteq M^{H_1}$; and $K \subseteq L_1 \subseteq L_2 \subseteq M$ implies that $\mathrm{Gal}(M/L_2) \subseteq \mathrm{Gal}(M/L_1) \subseteq G$.

Remark 5.85. Intuitively, inclusion-reversing means that small extensions become big subgroups, and big extensions becomes smalls subgroups. This is quite confusing.

Remark 5.86. The extension L corresponds to $\mathrm{Gal}(M/L)$, not $\mathrm{Gal}(L/K)$. As an example of a reason this is bad, L/K might not be a Galois extension, so this automorphism group need not be "good."

Let's give some examples before proving Theorem 5.84.

Exercise 5.87. We work out the Galois correspondence for the splitting field of $x^3 - 2$ over $\mathbb{Q}$.

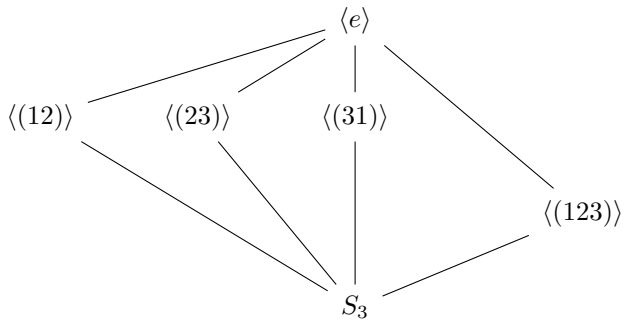
Proof. The roots of $x^3 - 2$ are $\{\sqrt[3]{2}, \omega\sqrt[3]{2}, \omega^2\sqrt[3]{2}\}$, where ω is a primitive third root of unity. Thus, our splitting field is $K \coloneqq \mathbb{Q}(\sqrt[3]{2}, \omega\sqrt[3]{2}, \omega^2\sqrt[3]{2})$, and it is not hard to see that this is $K = \mathbb{Q}(\sqrt[3]{2}, \omega)$.

Now, we see that $[K : \mathbb{Q}] = 6$ because we have the chain

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt[3]{2}) \subseteq \mathbb{Q}(\sqrt[3]{2}, \omega).$$

Namely, $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$ because the irreducible polynomial for $\sqrt[3]{2}$ is $x^3 - 2$, and $[\mathbb{Q}(\sqrt[3]{2}, \omega) : \mathbb{Q}(\sqrt[3]{2})] = 2$ because the irreducible polynomial for ω is $x^2 + x + 1$, which is irreducible over $\mathbb{Q}(\sqrt[3]{2})$ because it is quadratic and has no roots in $\mathbb{Q}(\sqrt[3]{2}) \subseteq \mathbb{R}$.

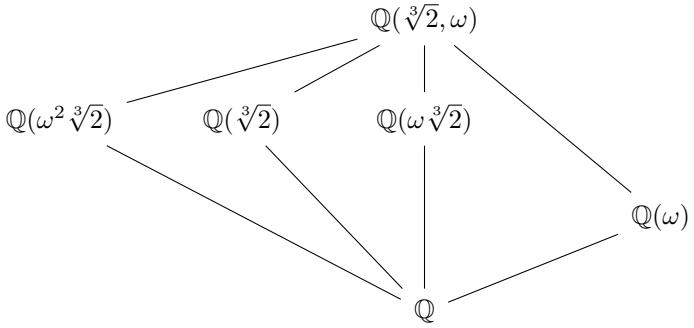
Thus, $\#\,\mathrm{Gal}(K/\mathbb{Q}) = [K : \mathbb{Q}] = 6$ and so must be S_3, acting on the three roots $\{\sqrt[3]{2}, \omega\sqrt[3]{2}, \omega^2\sqrt[3]{2}\}$ of $x^3 - 2$. (Amusingly, these fit on an equilateral triangle in the complex plane, though this is not a necessary picture.) Let's write out the lattice diagram of subgroups for S_3.



And we can write down the tower of fields. To be explicit, we number off $\{\sqrt[3]{2}, \omega\sqrt[3]{2}, \omega^2\sqrt[3]{2}\}$ by $\{1, 2, 3\}$.

The point is that (for example) $\langle(23)\rangle$ will only fix $\sqrt[3]{2}$ because it swaps the other two roots, so $\mathbb{Q}(\sqrt[3]{2})$ is an example of such a field fixed by these automorphisms, and it is not fixed by the other automorphisms, so this fixed field must be $\mathbb{Q}(\sqrt[3]{2})$.

Similarly, $\langle(12)\rangle$ corresponds to $\mathbb{Q}(\omega^2\sqrt[3]{2})$, and $\langle(31)\rangle$ corresponds to $\mathbb{Q}(\omega\sqrt[3]{2})$. Lastly, we need to find the field corresponding to $\langle(123)\rangle$. Well, this subgroup has index 2, so we need to find a quadratic subfield of $\mathbb{Q}(\sqrt[3]{2},\omega)$, which we see must be $\mathbb{Q}(\omega)$. This gives us the following lattice.



We remark that the Galois correspondence now tell us, automatically, that these are all of the intermediate fields. ■

Exercise 5.88. We work out the Galois correspondence for $\mathbb{F}_{64}/\mathbb{F}_2$.

Proof. We see that $\sigma : x \mapsto x^2$ is our Frobenius automorphism, and this automorphism σ has order 6: the order k of this automorphism is the smallest k such that

$$x^{2^k} = \sigma^k(x) = x.$$

Because we are working in $\mathbb{F}_{64} = \mathbb{F}_{2^6}$, certainly $k = 6$ suffices, and $k \geq 6$ because all of $\mathbb{F}_{64}$ must be the root of $x^{2^k} - x$.

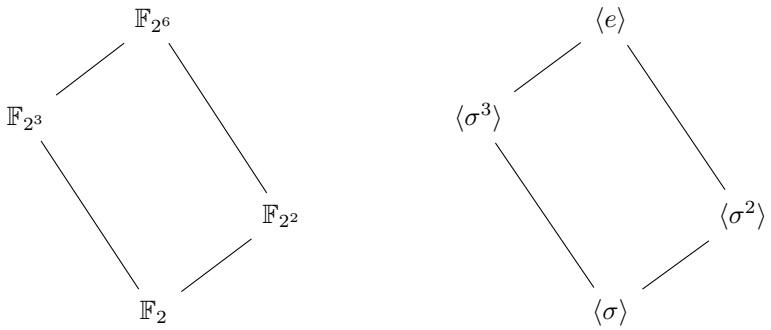
Additionally, we can see that $[\mathbb{F}_{64} : \mathbb{F}_2] = 6$, so in fact the Galois group must be generated by σ, giving $\mathrm{Gal}(\mathbb{F}_{64}/\mathbb{F}_2) \cong \langle\sigma\rangle \cong \mathbb{Z}/6\mathbb{Z}$.

Remark 5.89 (Nir)**.** For any prime-power q and positive integer r, the above argument can be used to show that $\mathrm{Gal}(\mathbb{F}_{q^r}/\mathbb{F}_q)$ is cyclic of order r generated by the Frobenius automorphism $x \mapsto x^q$.

Now, because $\mathbb{Z}/6\mathbb{Z}$ is cyclic, all of its subgroups are cyclic genreated by σ^d for various $d \mid 6$. Then we can see that the fixed field of $\langle\sigma^d\rangle \cong d\mathbb{Z}/6\mathbb{Z}$ consists of the elements such that

$$x^{2^d} = \sigma^d(x) = x,$$

which is exactly $\mathbb{F}_{2^d}$. Running this correspondence through gives the following lattices.

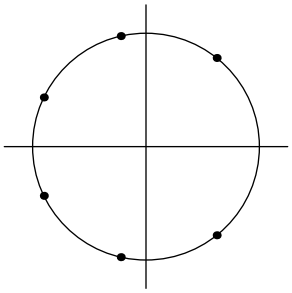


This finishes. ■

In the above examples, we more or less knew what the subfields and the subgroups were in advance, and it was nice to see the lattice diagrams correspond. In the next example, the subfields are less obvious.

Exercise 5.90. We work out the Galois correspondence for $\mathbb{Q}(\zeta_7)/\mathbb{Q}$ where ζ_7 is a primitive seventh root of unity.

Proof. The minimal polynomial for ζ_7 is $\Phi_7(x) = 1+x+\cdots+x^6 = \frac{x^7-1}{x-1}$, which has degree 6 and is irreducible because $\Phi_7(x+1)$ satisfies Eisenstein's criterion with the prime 7.[1] Visually, we can see all of the roots of Φ_7 as follows.



In particular, any root of $\Phi_7(x) = \frac{x^7-1}{x-1}$ must be a seventh root of unity which is not 1, so it must be a a primitive seventh root of unity. Conversely, we can see that all the primitive seventh roots of unity are indeed roots because they satisfy $x^7 - 1 = 0$ but $x - 1 \neq 0$.

Now, any automorphism in the Galois group must send ζ_7 to some other root of Φ_7, say ζ_7^k for $k \in (\mathbb{Z}/7\mathbb{Z})^\times$. We can check that each of these maps does indeed induce a unique automorphism

$$\mathbb{Q}(\zeta_7) \cong \frac{\mathbb{Q}[x]}{(\Phi_7(x))} \cong \mathbb{Q}(\zeta_7^k)$$

because the minimal polynomial of ζ_7^k is still Φ_7. So each of the constraints $\zeta_7 \mapsto \zeta_7^k$ induces a unique automorphism, so in fact we get that

$$\mathrm{Gal}(\mathbb{Q}(\zeta_7)/\mathbb{Q}) \cong (\mathbb{Z}/7\mathbb{Z})^\times$$

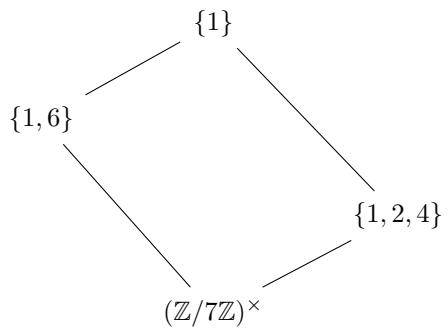
by taking the automorphism $\sigma_k : \zeta_7 \mapsto \zeta_7^k$ to k. Technically, we shuold check that this is well-defined (it is because ζ_7 exponents only matter $\pmod 7$) and that $\sigma_k \circ \sigma_\ell = \sigma_{k\ell}$ to be a homomorphism (it is because). So we get our isomorphism.

Remark 5.91 (Nir)**.** The above argument can be carried out essentially verbatim by replacing 7 with any prime. It is fact that, for any positive integer n, we have

$$\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \cong (\mathbb{Z}/n\mathbb{Z})^\times,$$

which we can see using the last half of the argument above, but some amount of care is required to show that Φ_n is in fact irreducible.

Noting that 3 is a generator of $(\mathbb{Z}/7\mathbb{Z})^\times \cong \mathbb{Z}/6\mathbb{Z}$, we can write down our subgroup lattice as follows.



[1] More generally, we showed that $\Phi_p(x)$ is irreducible whenwe first introduced Eisenstein's criterion.

So now let's try and find our fixed fields. Of course $\{1\}$ corresponds to $\mathbb{Q}(\zeta_7)$, and $(\mathbb{Z}/7\mathbb{Z})^\times$ corresponds to $\mathbb{Q}$. We now do the harder ones.

- For $\{1,6\}$, we note that the automorphism $\sigma_6 : \zeta_7 \mapsto \zeta_7^6 = \overline{\zeta_7}$ is simply the conjugation automorphism: for any element $\alpha = \sum_{k=1}^{6} a_k \zeta_7^k$, we see that $\sigma\alpha = \overline{\alpha}$ by direct expansion.

  Anyways, the point is that the fixed field of conjugation is $\mathbb{R}$, so restricting our view to $\mathbb{Q}(\zeta_7)$, we are interested in $\mathbb{Q}(\zeta_7) \cap \mathbb{R}$. It is not too hard to see that this field is $\mathbb{Q}(\zeta_7 + \zeta_7^{-1})$ (e.g., $[\mathbb{Q}(\zeta_7) : \mathbb{Q}(\zeta_7+\zeta_7^{-1})] = 2$ and then use the Galois correspondence), but this observation does not matter very much.

- For $\{1,2,4\}$, this subgroup has index 2, so we are looking for a field of degree 2 over $\mathbb{Q}$. Namely, we want a quadratic subextension of $\mathbb{Q}(\zeta_7)$; with all the 7s floating around, it is reasonable to hope that we get $\mathbb{Q}(\zeta 7)$ or $\mathbb{Q}(\sqrt{-7})$.

  Anyways, to find a generator, we pick up some random element fixed by $\{\sigma_1, \sigma_2, \sigma_4\}$, say

$$\alpha \coloneqq \sigma_1(\zeta_7) + \sigma_2(\zeta_7) + \sigma_4(\zeta_7) = \zeta_7 + \zeta_7^2 + \zeta_7^4.$$

  We hope that this "generic" element will generate our subextension. Well, we can square α to get

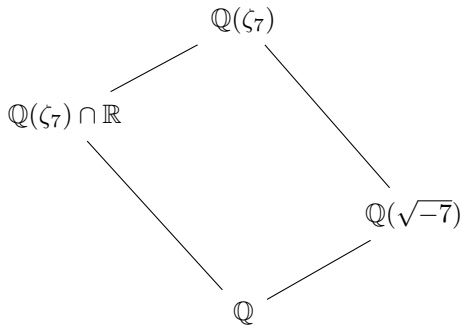
$$\begin{aligned} \alpha^2 &= \zeta_7^2 + \zeta_7^4 + \zeta_7^8 + 2\left(\zeta_7^3 + \zeta_7^6 + \zeta_7^5\right) \\ &= \zeta_7 + \zeta_7^2 + 2\zeta_7^3 + \zeta_7^4 + 2\zeta_7^5 + 2\zeta_7^6 \\ \alpha^2 + \alpha &= 2\zeta_7 + 2\zeta_7^2 + 2\zeta_7^3 + 2\zeta_7^4 + 2\zeta_7^5 + 2\zeta_7^6 \\ &= 2 \cdot -1, \end{aligned}$$

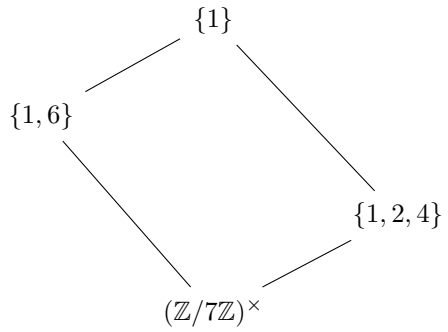
  so we find that $\alpha^2 + \alpha + 2 = 0$. Thus,

$$\alpha = \frac{-1 \pm \sqrt{-7}}{2},$$

  so our corresponding field here is $\mathbb{Q}(\sqrt{-7})$.

In total, we see that we have the following lattices.





This finishes. ∎

Remark 5.92. In the above, one might object that α is set to a specific number, but we only showed that $\alpha \in \left\{\frac{-1\pm\sqrt{-7}}{2}\right\}$. However, this was good enough for our purposes, so we don't need to figure out which one is α, and thankfully so—actually figuring out which one is α requires much more effort.

Remark 5.93. If we used 5 instead of 7 in the above exampe, our quadratic subextension would have been $\mathbb{Q}(\sqrt{+5})$. Whether or not the minus sign is added has to do with quadratic reciprocity.

5.3.3 Proof of the Galois Correspondence

Anyways, let's prove our theorem.

Theorem 5.84. Fix M/K a finite Galois extension of fields with Galois group $G \coloneqq \mathrm{Gal}(M/K)$. Then we get a one-to-one correspondence between intermediate extensions $K \subseteq L \subseteq M$ and subgroups $H \subseteq G$. To be explicit, our maps are as follows.

$$\begin{array}{ccc} M^H & \longmapsto & H \\ L & \longmapsto & \mathrm{Gal}(M/L) \end{array}$$

Additionally, we have $[M : M^H] = \#H$ and $[M : L] = \#\,\mathrm{Gal}(M/L)$. In fact, this mapping is inclusion reversing: if we have subgroups $H_1 \subseteq H_2 \subseteq G$, then $M^{H_2} \subseteq M^{H_1}$; and $K \subseteq L_1 \subseteq L_2 \subseteq M$ implies that $\mathrm{Gal}(M/L_2) \subseteq \mathrm{Gal}(M/L_1) \subseteq G$.

Proof. For concreteness, we label our maps by $f : L \mapsto \mathrm{Gal}(M/L)$ and $g : H \mapsto M^H$. To show that f and g are inverses and bijective, it suffices to just show that they are inverses.

We show that $g \circ f = \mathrm{id}$. Indeed, we start with some subgroup $H \subseteq G$ and take

$$H \overset{g}{\longmapsto} M^H \overset{f}{\longmapsto} \mathrm{Gal}(M/M^H).$$

Certainly $H \subseteq \mathrm{Gal}(M/M^H)$ because any $\sigma \in H$ will fix M^H by definition of M^H. We would like to get the equality.

Remark 5.94. This is trivial, but it is easy to get it wrong.

It suffices to show that these have the same size, so we claim that

$$\#H \overset{?}{=} \#\,\mathrm{Gal}(M/M^H).$$

But we've done this: M/M^H is a Galois extension by Proposition 5.76 part (d), so $\#H = \#\,\mathrm{Gal}(M/M^H)$ by Proposition 5.76 part (c).

We now show that $g \circ f = \mathrm{id}$. Indeed, we start with some intermediate field L and take

$$L \overset{f}{\longmapsto} \mathrm{Gal}(M/L) \overset{g}{\longmapsto} M^{\mathrm{Gal}(M/L)}.$$

We hope to show that $L = M^{\mathrm{Gal}(M/L)}$. Certainly $L \subseteq M^{\mathrm{Gal}(M/L)}$ because each $\sigma \in \mathrm{Gal}(M/L)$ will fix L by definition of $\mathrm{Gal}(M/L)$. We would like to show the equality.

Well, again by size arguments, it suffices to show that both of these fields have the same "size." Explicitly, we claim that

$$[L : K] \overset{?}{=} [M^{\mathrm{Gal}(M/L)} : K].$$

Indeed, dividing both sides from $[M : K]$, it suffices to show that $[M : L] \overset{?}{=} [M : M^{\mathrm{Gal}(M/L)}]$. But here we see that $M/M^{\mathrm{Gal}(M/L)}$ is Galois by Proposition 5.76, so $[M : M^{\mathrm{Gal}(M/L)}] = \#\,\mathrm{Gal}(M/L)$. So it suffices to show that $[M : L] \overset{?}{=} \#\,\mathrm{Gal}(M/L)$.

Warning 5.95. It is not true that the size of $\#\,\mathrm{Gal}(M/L)$ equals $[L : K]$ directly. However, we do have $[M : L] = \#\,\mathrm{Gal}(M/L)$.

At this point, the claim that $[M : L] \overset{?}{=} \#\,\mathrm{Gal}(M/L)$ is more internal to just the extension M/L, so we hope that it is more tractable. Observe that, by Proposition 5.76, we are essentially showing that M/L is a Galois extension.

Of course, the only reason we have to believe that M/L is Galois is that M/K is Galois, so we will have to use this fact. Certainly we do get $[M : L] \geq \#\,\mathrm{Gal}(M/L)$. To get the other inequality, we see that we can

count elements $\sigma \in \mathrm{Gal}(M/K)$ by counting embeddings $L \hookrightarrow M$ fixing K and multiplying by the number of ways to extend these maps $L \hookrightarrow M$ up to $M \hookrightarrow M$. Bounding both of these quantities,[2] we see that

$$\#\,\mathrm{Gal}(M/K) \le [L:K] \cdot [M:L],$$

but equality must hold because both sides here are $[M:K]$ because M/K is Galois. In particular, there are $[M:L]$ maps extending the embedding $L \hookrightarrow L \subseteq M$, which is the same as saying there are $[M:L]$ elements in $\mathrm{Gal}(M/L)$. ■

Remark 5.96. If M/K is not Galois, we do get something: there is a correspondence between subgroups $G \subseteq \mathrm{Gal}(M/K)$ and subextensions containing $M^G \supseteq K$. One way to see this is to throw out K and just work with the Galois extension M/M^G instead.

Remark 5.97 (Nir)**.** Technically we have not shown the inclusion-reversing in the above argument. We do this quickly.

- If $H_1 \subseteq H_2 \subseteq G$, then we know that each $\alpha \in M^{H_2}$ will be fixed by each element of $H_1 \subseteq H_2$, so $M^{H_2} \subseteq M^{H_1}$.
- If $K \subseteq L_1 \subseteq L_2 \subseteq M$, then each $\sigma \in \mathrm{Gal}(M/L_2)$ will fix each element of $L_1 \subseteq L_2$, so $\mathrm{Gal}(M/L_2) \subseteq \mathrm{Gal}(M/L_1)$.

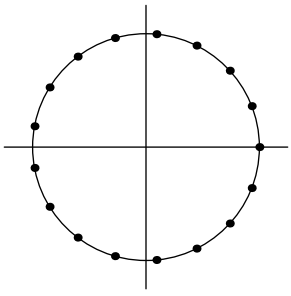
5.3.4 Applications of the Galois Correspondence

Let's do some more applications.

Exercise 5.98. We construct the heptadecagon, the regular 17-gon.

Remark 5.99. Gauss did this when he was a teenager. The Greeks had known about 2^n-gons and $2^n \cdot 3$-gons and $2^n \cdot 5$-gons and $2^n \cdot 3 \cdot 5$-gons. But the 17-gon made Gauss somewhat famous.

Proof. Fix ζ_{17} a primitive seventeenth root of unity in $\mathbb{C}$; here is the picture, to establish that this will in fact give us a 17-gon on the unit circle.



By taking powers, we essentially have to construct one of these because the ruler-and-compass constructions correspond to algebraic constructions with $+, -, \times, \div, \sqrt{\cdot}$. Well, if we can take square roots, then we essentially need to find a sequence of quadratic extensions

$$\mathbb{Q} \subseteq F_1 \subseteq F_2 \subseteq F_3 \subseteq \mathbb{Q}(\zeta_{17}).$$

[2] We are using the fact, given an extension L/K, the number of embeddings $L \hookrightarrow X$ fixing K is bounded above by $[L:K]$. One way to see this is to use a chain argument as we did in the case of $X = \overline{K}$.

Well, this is not that bad: we know that

$$\mathrm{Gal}(\mathbb{Q}(\zeta_{17})/\mathbb{Q}) \cong (\mathbb{Z}/17\mathbb{Z})^\times \cong \mathbb{Z}/16\mathbb{Z}$$

as we discussed earlier with $\mathbb{Q}(\zeta_7)$. So in the Galois correspondence, we can use the sequence of index-2 subgroups

$$\mathbb{Z}/16\mathbb{Z} \supseteq 2\mathbb{Z}/16\mathbb{Z} \supseteq 4\mathbb{Z}/16\mathbb{Z} \supseteq 8\mathbb{Z}/16\mathbb{Z} \supseteq 16\mathbb{Z}/16\mathbb{Z}.$$

Thus, at least abstractly, we see that it is possible to construct ζ_{17} and hence the 17-gon.

Let's actually find some of these fields explicitly. Well, note that $(\mathbb{Z}/17\mathbb{Z})^\times \cong \mathbb{Z}/16\mathbb{Z}$ is cyclic generated by 3 (again), but I won't write out the log table here. Now let's find our subgroups and so the subextensions.

- The entire set $(\mathbb{Z}/17\mathbb{Z})^\times$ correponds to $\mathbb{Q}$.
- The squares of index 2 becomes $\{1, 9, \ldots\}$ of order 8.
- The fourth-powers are the next index-2 subgroup, which are $\{1, 13, 16, 4\}$.
- The next index-2 subgroup is $\{1, 16\}$.
- Lastly, we are left with $\{1\}$.

Building our tower of fields as follows by taking the powrs of ζ_{17}, as we did earlier. For example, set

$$\alpha = \zeta_{17}^1 + \zeta_{17}^9 + \cdots$$

and

$$\beta = \zeta_{17}^3 + \zeta_{17}^{10} + \cdots$$

to be the the two cosets of the subgroup of order 8, namely the squares. Well, we see that $(x-\alpha)(x-\beta)$ will be fixed by the $(\mathbb{Z}/17\mathbb{Z})^\times$-action, so it will be in $\mathbb{Q}[x]$. We can check by hand that $\alpha+\beta = -1$ and $\alpha\beta = -4$, which we leave as an exercise; then we find that α and β are the roots of

$$x^2 + x - 4 = 0,$$

which gives

$$\alpha, \beta = \frac{-1 \pm \sqrt{17}}{2}.$$

So we have that $F_1 = \mathbb{Q}(\sqrt{17})$.

So next let's look at the cosets of the subgroup of order 4 in the subgroup of order 8. Namely, we set

$$\gamma = \zeta_{17} + \zeta_{17}^3 + \zeta_{17}^{16} + \zeta_{17}^4,$$

and

$$\delta = \zeta_{17}^4 + \zeta_{17}^{15} + \zeta_{17}^8 + \zeta_{17}^2.$$

Again, we can find that $(x-\gamma)(z-\delta)$ is fixed by the right Galois group to get that this will live in $F_1 = \mathbb{Q}(\sqrt{17})$. We can check that by hand that $\gamma + \delta = \frac{-1+\sqrt{17}}{2}$ and $\gamma\delta = -1$, which means that we can write down γ and δ using the quadratic formula. In theory, we could do everything explicitly, but it is somewhat tedious. ■

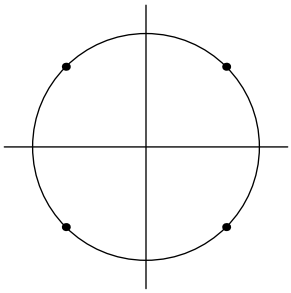
Remark 5.100. Gauss wanted the 17-gon on his tombstone. This did not occur.

Remark 5.101. In general, any prime of the form $1 + 2^n$ will work using the above construction. For example, we can do 257 and 65537. Folklore says that somewhat worked out an explicit construction of 65537, but this is a somewhat useless exercise.

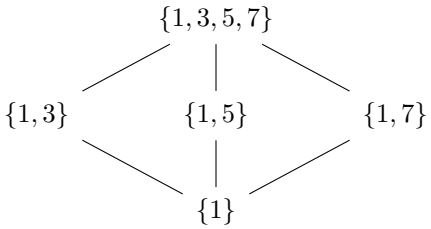
Let's do some more examples.

Exercise 5.102. We work out the Galois correspondence for the splitting field of x^4+1 over $\mathbb{Q}$.

Proof. We find that $x^4+1 = \frac{x^8-1}{x^4-1}$, so graphically its roots look as follows.



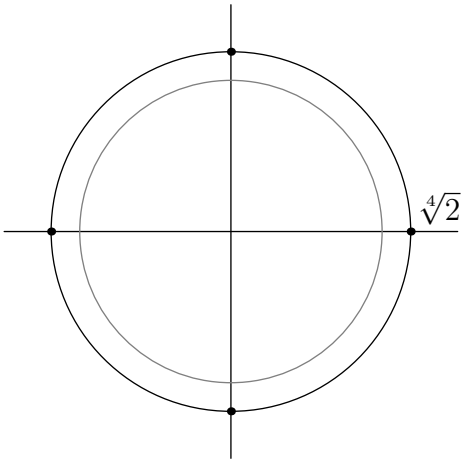
So we find that the roots are primitive 8th roots of unity, so our splitting field is $\mathbb{Q}(\zeta_8)$, which will have Galois group $\mathrm{Gal}(\mathbb{Q}(\zeta_8)/\mathbb{Q}) \cong (\mathbb{Z}/8\mathbb{Z})^\times$. As a warning, this is not cyclic because all of the elements of $(\mathbb{Z}/8\mathbb{Z})^\times$ have expoent 2. In particular, here is our lattice of subgroups.



We could work out the corresponding lattice of subfields, but it is difficult to do live, so we leave it as an exercise. ■

Exercise 5.103. We work out the Galois correspondnece for the splitting field of x^4-2 over $\mathbb{Q}$.

Proof. We note that $\sqrt[4]{2}$ is certainly a root, but $\mathbb{Q}(\sqrt[4]{2})$ is not the splitting field because we are missing the roots $i\sqrt[4]{2}$ and $i^3\sqrt[3]{2}$. So to get the full splitting field, we want $\mathbb{Q}(\sqrt[3]{2}, i)$. Here is our picture.



It turns out that the Galois group must preserve the above square, which we can check algebraically, so our Galois group is D_8. If we take D_8 generated by 90° rotation σ and a reflection τ we get the following lattice, which I won't write out because it is complicated.

However, we can see that we get three quadratic extensions of $\mathbb{Q}(\sqrt[4]{2}, i)$, which correspond to $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{-2})$. then our extensions of degree 4 are $\mathbb{Q}(i, \sqrt{2})$ and $\mathbb{Q}(\sqrt[4]{2})$ and $\mathbb{Q}(i\sqrt[4]{2})$ and $\mathbb{Q}((1+i)\sqrt[4]{2})$ and $\mathbb{Q}((1-i)\sqrt[4]{2})$. ■

Remark 5.104. Again, some of these subfields are not obvious: namely, $\mathbb{Q}((1+i)\sqrt[4]{2})$ is somewhat subtle. But we can find it from the Galois correspondence.

5.3.5 Intermediate Normal Extensions

As an aside, we can see from the above lattices that it appears normal subgroups correspond to normal extensions.

Example 5.105. In the above examples, we see that any of our index-2 subgroups correspond to quadratic subextensions, and both of these objects are normal for the corresponding definitions of normal.

To be explicit, we have the following statement.

Proposition 5.106. Fix M/K a Galois extension with Galois group $G \coloneqq \mathrm{Gal}(M/K)$. Then we have the following.

- If $K \subseteq L \subseteq M$ is an intermediate extension such that L/K is normal, $\mathrm{Gal}(M/L) \subseteq G$ is a normal subgroup, andL/K is a Galois extension such that
$$\mathrm{Gal}(L/K) \cong \frac{\mathrm{Gal}(M/K)}{\mathrm{Gal}(M/L)}.$$
- Take M/K finite. If $H \subseteq G$ is a normal subgroup, then the corresponding fixed field M^H has M^H/K a normal extension. In fact, M^H/K is a Galois extension with Galois group G/H.

Proof. We show the claims one at a time.

- We are given that L/K is normal, and we see that L/K is separable because $L \subseteq M$, so all elements of L are separable over K because all elements of M are separable over K.

  So L/K is a Galois extension. To compute its Galois group, we construct $\varphi : \mathrm{Gal}(M/K) \to \mathrm{Gal}(L/K)$ by restriction, taking
  $$\varphi : \sigma \mapsto \sigma|_L.$$
  Indeed, it is not hard to see that $\sigma|_L$ is in fact an automorphism: it is at least an embedding $L \hookrightarrow M$, and because L is normal, this embedding $L \hookrightarrow M \hookrightarrow \overline{K}$ must be an automorphism. Note that here is the only place in this argument where we use the fact that L is normal: it makes φ well-defined.

  Now, φ is surjective because any $L \to L$ fixing K becomes an embedding $L \to L \hookrightarrow M$ fixing K, which can then be lifted up to an automoprhism $M \hookrightarrow M$ fixing K by using a chain argument. And lastly, we see that the kernel of φ is
  $$\{\sigma \in \mathrm{Gal}(M/K) : \sigma|_L = \mathrm{id}_L\},$$
  which is simply the automorphisms of M fixing K which also fix L. But $K \subseteq L$, so $\ker \varphi = \mathrm{Gal}(M/L)$.

  Thus, $\mathrm{Gal}(M/L)$ is indeed a normal subgroup of G because it is the kernel of φ, and we find that
  $$\mathrm{Gal}(L/K) \cong \frac{\mathrm{im}\,\varphi}{\ker \varphi} = \frac{\mathrm{Gal}(M/K)}{\mathrm{Gal}(M/L)}.$$
  This is what we wanted.

- Now take $H \subseteq G$ a normal subgroup, and we want to show that M^H/K is a Galois extension. Surely this extension is separable because element of $M \supseteq M^H$ is separable over K.

  So the hard part is showing that M^H/K is normal. Well, assign an embedding $M^H \subseteq \overline{K}$, and suppose that we have some other embedding $\sigma : M^H \hookrightarrow \overline{K}$ so that it suffices to show $\sigma(M^H) \subseteq M^H$, which will imply that σ is an automorphism.

  Well, to show $\sigma(M^H) \subseteq M^H$, we need to show that $\sigma(M^H)$ is fixed by H. So pick up some $m \in M^H$, and then we note, for any $h \in H$, we have $\sigma^{-1}h\sigma \in H$ (here we use the fact that H is normal), so $\left(\sigma^{-1}h\sigma\right)(m) = m$, so
  $$h(\sigma m) = \sigma m.$$
  Thus, each $h \in H$ fixes each $\sigma m \in \sigma\sigma(M^H)$. So indeed, $\sigma(M^H) \subseteq M^H$, so each embedding $M^H \hookrightarrow \overline{K}$ is an automorphism, so M^H/K is normal.

  To finish, we see that
  $$\mathrm{Gal}\left(M^H/K\right) \cong \frac{\mathrm{Gal}\left(M/K\right)}{\mathrm{Gal}\left(M/M^H\right)} = \frac{G}{H},$$
  where we have used the previous part for $\cong$ and the Galois correspondence for $=$. This finishes. ■

Remark 5.107. This is where the term "normal subgroup" came from: first normal was used for field extensions, and then second it was pushed into group theory from this correspondence.

5.4 November 18

This is a roadkill song about a kid who followed the bouncing ball in a singalong.

5.4.1 Normal Loose Ends

Last time we were cut off discussing normal extensions. We will give an alternate proof of the fact that normal extensions correspond to normal subgroups. Here is the key lemma.

Lemma 5.108. Fix M/K a Galois extension with Galois group $G \coloneqq \mathrm{Gal}(M/K)$ and L an intermediate extension with corresponding subgroup $H \coloneqq \mathrm{Gal}(M/L) \subseteq G$ fixing L. Then, for any $\sigma \in G$, the subgroup fixing σL is $\mathrm{Gal}(M/\sigma L) = \sigma H \sigma^{-1}$.

Proof. Fix $\sigma \in G$. We start by noting that σL is a field because σ is an automorphism, so the field structure of L will carry over to σL.[3]

We are interested in computing $\mathrm{Gal}(M/\sigma L)$. Well, certainly $\mathrm{Gal}(M/\sigma L) \subseteq \mathrm{Gal}(M/K)$ because $K \subseteq \sigma L$, so it suffices to check which $\tau \in G$ fix σL. But now any element of σL takes the form $\sigma\alpha$ for some $\alpha \in L$, so $\tau \in G$ fixes each $\sigma\alpha \in L$ if and only if
$$\tau\sigma\alpha = \sigma\alpha$$
if and only if $\left(\sigma^{-1}\tau\sigma\right)(\alpha) = \alpha$ if and only if $\sigma^{-1}\tau\sigma$ fixes L. But the subgroup fixing L is H, so this is equivalent to $\tau \in \sigma H \sigma^{-1}$, so indeed the subgroup of G fixing L is $\sigma H \sigma^{-1}$. ■

Remark 5.109 (Nir)**.** Mnemonically, we have
$$\left(\sigma H \sigma^{-1}\right)(\sigma L) = \sigma\left(H \cdot \sigma^{-1}\sigma L\right) = \sigma(H \cdot L) = \sigma L,$$
where we have commited heavy abuse of notation mutliplying the subgroup H by a field L.

[3] For example, σL has inverses: for any $\sigma\alpha \in (\sigma L) \setminus \{0\}$, we have $\alpha \neq 0$, so $(\sigma\alpha)^{-1} = \sigma\left(\alpha^{-1}\right)$ provides an inverse.

Another way to phrase the above lemma is that the set of intermediate extensions of M/K and subgroups of G are isomorphic as G-sets, where the bijection is by the Galois correspondence by

$$\varphi : L \mapsto \mathrm{Gal}(M/L).$$

To be explicit, the G-action, for some $\sigma \in G$, on the intermediate extensions is by $\sigma \cdot L = \sigma L$ and on the subgroups by conjugation.

Indeed, we only need to show that this mapping is a G-set homomorphism because we already know it is bijective. Well, fixing some $\sigma \in G$, we see that

$$\varphi(\sigma \cdot L) = \varphi(\sigma L) = \mathrm{Gal}(M/\sigma L) = \sigma \,\mathrm{Gal}(M/L)\sigma^{-1} = \sigma \cdot \mathrm{Gal}(M/L) = \sigma \cdot \varphi(L),$$

which is what we wanted.

In particular, to show that normal subgroups correspond to normal extensions, we note that normal subgroups are exactly the subgroups fixed by the G-action (by conjugation), so by the above isomorphism as G-sets, it suffices to talk about the intermediate extensions fixed by the G-action.

Lemma 5.110. Fix M/K a (finite) Galois extension with Galois group $G \coloneqq \mathrm{Gal}(M/K)$. Then an intermediate extension L is normal if and only if $\sigma L = L$ for each $\sigma \in G$.

Proof. This is somewhat techincal; we more or less showed this last time. Fix an algebraic closure $\overline{K}$. In one direction, if L is normal, then we note that each $\sigma \in G$ restricts to a map

$$L \xrightarrow{\sigma} \sigma L \subseteq M \subseteq \overline{K}.$$

But any embedding $L \hookrightarrow \overline{K}$ must output into L because L is normal, so the composite of the above maps into L, so $\sigma L \subseteq L$. A similar argument shows $\sigma^{-1} L \subseteq L$, so $L \subseteq \sigma L$ as well, so $L = \sigma L$.

For the other direction, suppose $\sigma L = L$ for each $\sigma \in G$. Fix any embedding $\sigma : L \hookrightarrow \overline{K}$ so that we want to show $\sigma L \subseteq L$. Well, by using some chain argument, we can extend $\sigma : L \hookrightarrow \overline{K}$ to

$$\sigma : M \hookrightarrow \overline{K},$$

but now we know that M is normal, so σ must output into M. In particular, $\sigma : M \hookrightarrow M$ fixing K, so we claim $\sigma \in \mathrm{Gal}(M/K)$. The only concern is for surjectivity, but we notice that $\sigma M \subseteq M$ while $[M : K] = [\sigma M : K]$ by tracking a basis, so $M = \sigma M$.[4] So $\sigma \in \mathrm{Gal}(M/K)$, so $\sigma L = L$, so L is indeed normal. ∎

So we get the following result.

Proposition 5.111. Fix M/K a (finite) Galois extension with Galois group $G \coloneqq \mathrm{Gal}(M/K)$. Then normal extensions correspond to normal subgroups. Explicitly, we have the following.

- Fix L/K a normal intermediate extension. Then $\mathrm{Gal}(M/L) \subseteq G$ is normal.
- FIx $H \subseteq G$ a normal subgroup. Then M^H/K is a normal extension.

Proof. These more or less follow directly from the above discussion.

- The extension L/K is normal if and only if L is a fixed point of the G-action on intermediate extensions if and only if $\mathrm{Gal}(M/L) \subseteq G$ is a fixed point of the G-action on subgroups if and only if $\mathrm{Gal}(M/L) \subseteq G$ is normal.
- The subgroup $H \subseteq G$ is the subgroup fixing M^H with $\mathrm{Gal}(M/M^H) = H$ by the Galois correspondence, so the previous part promises that M^H/K is normal because $\mathrm{Gal}(M/M^H) \subseteq G$ is a normal subgroup. ∎

[4] We have used the fact that M/K is finite (or at lease profinite) here.

To close off, we note that

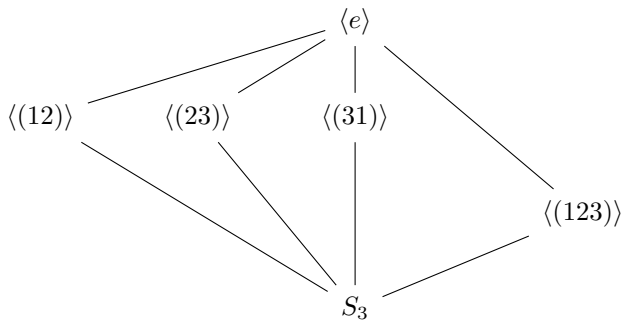
$$\mathrm{Gal}(L/K) \cong \frac{\mathrm{Gal}(M/K)}{\mathrm{Gal}(M/L)}$$

as we claimed last time. Indeed, as we said last time, we have a map $\mathrm{Gal}(M/K) \to \mathrm{Gal}(L/K)$ by restriction, and it has kernel $\mathrm{Gal}(M/L)$.

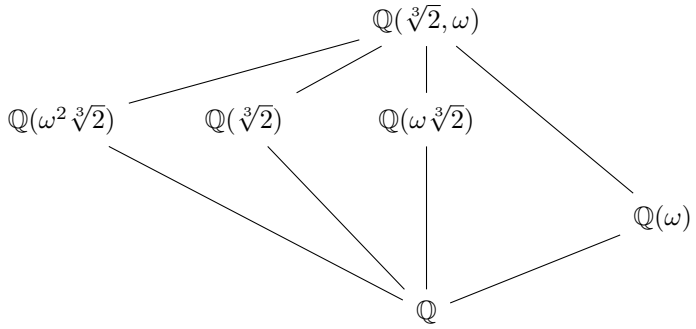
Anyways, here is an example.

Exercise 5.112. We find the intermediate normal extensions of $\mathbb{Q}(\sqrt[3]{2}, \omega)/\mathbb{Q}$.

Proof. Here is our lattice of subgroups.



And here is our lattice of fields, where we have numbered the roots $\{\sqrt[3]{2}, \omega\sqrt[3]{2}, \omega^2\sqrt[3]{2}$ off by $\{1, 2, 3\}$ respectively. We provided the details to this last time.



So, for example, we see that conjugation by (12) takes (23) to $(12)(23)(12) = (13)$ and so the subgroup $\langle (12) \rangle$ to $\langle (13) \rangle$. Isomorphcially, the action by (12) will permute the corresponding fields $\mathbb{Q}(\sqrt[3]{2})$ and $\mathbb{Q}(\omega\sqrt[3]{2})$. In particular, the extensions $\mathbb{Q}(\omega^\bullet\sqrt[3]{2})/\mathbb{Q}$ are not normal.

However, the subgroup $\langle (123) \rangle$ is normal (it is index 2 in S_3), so the extension $\mathbb{Q}(\omega)/\mathbb{Q}$ is a normal extension (it is quadratic), and the Galois group here is

$$\mathrm{Gal}(\mathbb{Q}(\omega)/\mathbb{Q}) \cong \frac{\mathrm{Gal}(\mathbb{Q}(\sqrt[3]{2}, \omega)/\mathbb{Q})}{\mathrm{Gal}(\mathbb{Q}(\sqrt[3]{2}, \omega))} \cong \frac{S_3}{\langle (123) \rangle} \cong \mathbb{Z}/2\mathbb{Z}.$$

This is what we wanted. ∎

5.4.2 Inverse Galois Problem: Cyclic Extensions

Before jumping into the proof, we pick up the following technical lemma.

Lemma 5.113 (Nir). Fix p a prime, and fix $K := \mathbb{Q}(\zeta_p)$ so that $G := \mathrm{Gal}(K/\mathbb{Q}) \cong (\mathbb{Z}/p\mathbb{Z})^\times$. Then, for any subgroup $H \subseteq G$, we have

$$K^H = \mathbb{Q}\left(\sum_{\sigma\in H} \sigma\zeta_p\right).$$

Proof. This is surprisingly technical. Fix $\alpha := \sum_{\sigma\in H} \sigma\zeta_p$. We show that $H = \mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q}(\alpha))$, which will be enough by the Galois correspondence. Well, $\tau \in G$ fixes $\mathbb{Q}(\alpha)$ if and only if τ fixes α (τ already fixes $\mathbb{Q}$) if and only if

$$\sum_{\sigma\in H} (\tau\sigma)\zeta_p = \tau \cdot \sum_{\sigma\in H} \sigma\zeta_p = \tau\alpha = \alpha = \sum_{\sigma\in H} \sigma\zeta_p.$$

Now, if $\tau \in H$, then the map $\sigma \mapsto \tau\sigma$ is a bijection $H \to H$, so τ certainly fixes H.

The converse requires a little more care. The main point is that, because p is prime, we see $\{\sigma\zeta_p\}_{\sigma\in G}$ is a basis for $\mathbb{Q}(\zeta_p)/\mathbb{Q}$. Indeed, by our classification of $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})$, we have

$$\{\sigma\zeta_p\}_{\sigma\in G} = \left\{\zeta_p^k\right\}_{k=1}^{p-1}.$$

There are $p-1$ of these elements, which is indeed the degree $[\mathbb{Q}(\zeta_p) : \mathbb{Q}]$, and these elements are $\mathbb{Q}$-linearly independent because the minimal polynomial of ζ_p has degree $p-1$.

Thus, we see that both sides of the equality

$$\sum_{\sigma\in H} (\tau\sigma)\zeta_p = \sum_{\sigma\in H} \sigma\zeta_p.$$

feature decompositions of the same element under a basis, so they must be permutations of each other. In particular, $\tau\zeta_p$ appears somewhere on the right, so $\tau \in H$. This finishes. ∎

Remark 5.114. In office hours, Professor Borcherds pointed out that this need not be true if we remove the prime condition: for ζ_8, the elements $\{\sigma\zeta_8\}_{\sigma\in\mathrm{Gal}(\mathbb{Q}(\zeta_8)/\mathbb{Q})} = \{\zeta_8, \zeta_8^3, \zeta_8^5, \zeta_8^7\}$ are not linearly independent (e.g., they sum to 0). We can manifest this into a problem as

$$\zeta_8 + \zeta_8^5 = 0 = \zeta_8^3 + \zeta_8^7,$$

and, in particular, the fixed field of $H = \{1, 5\}$ is not $\mathbb{Q}\left(\zeta_8 + \zeta_8^5\right) = \mathbb{Q}$.

With that annoyance out of the way, let's move into examples.

Exercise 5.115. We find a $\mathbb{Q}$-extension with Galois group $\mathbb{Z}/5\mathbb{Z}$.

Proof. By our discussion of normal subgroups and quotients it suffices to find some extension $L/\mathbb{Q}$ such that $\mathrm{Gal}(L/\mathbb{Q})$ has $\mathbb{Z}/5\mathbb{Z}$ as a quotient. Well, we have that

$$\mathrm{Gal}(\mathbb{Q}(\zeta_{11})/\mathbb{Q}) \cong (\mathbb{Z}/11\mathbb{Z})^\times \cong \mathbb{Z}/10\mathbb{Z},$$

where the first isomorphism is by associating the automorphism $\sigma_k : \zeta_{11} \mapsto \zeta_{11}^k$ to $k \in (\mathbb{Z}/11\mathbb{Z})^\times$.

Now, $\mathbb{Z}/10\mathbb{Z}$ does surject onto $\mathbb{Z}/5\mathbb{Z}$, so tracking things backwards, we are looking at the quotient

$$\mathbb{Z}/5\mathbb{Z} \cong \frac{\mathbb{Z}/10\mathbb{Z}}{5\mathbb{Z}/10\mathbb{Z}} \cong \frac{\mathrm{Gal}(\mathbb{Q}(\zeta_{11})/\mathbb{Q})}{\langle \zeta_{11} \mapsto \zeta_{11}^{-1}\rangle}.$$

In other words, we want the elements of $\mathbb{Q}(\zeta_{11})$ which are fixed by the action $\zeta_{11} \mapsto \zeta_{11}^{-1}$, so Lemma 5.113 tells us that this field is

$$\mathbb{Q}\left(\zeta_{11} + \zeta_{11}^{-1}\right) \cong \mathbb{Q}\left(\cos\left(\frac{2\pi}{11}\right)\right).$$

Namely,

$$\boxed{\mathrm{Gal}\left(\mathbb{Q}\left(\cos\left(\frac{2\pi}{11}\right)\right)/\mathbb{Q}\right) \cong \mathbb{Z}/5\mathbb{Z}}.$$

This is what we wanted. ■

Here is the general statement.

Proposition 5.116. We find a $\mathbb{Q}$-extension with Galois group $\mathbb{Z}/n\mathbb{Z}$.

Proof. We again start with $\mathbb{Q}(\zeta_p)$ for some prime p to be chosen later. Then we find that

$$\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q}) \cong (\mathbb{Z}/p\mathbb{Z})^\times \cong \mathbb{Z}/(p-1)\mathbb{Z}.$$

To get this to surject onto $\mathbb{Z}/n\mathbb{Z}$, so that means we want $p \equiv 1 \pmod{n}$, of which there are infinitely many by Dirichlet's theorem on arithmetic progressions.

To finish, we can do the algorithm suggested above. Fix g a generator of $(\mathbb{Z}/p\mathbb{Z})^\times$ so that g^n has order $(p-1)/n$ and so generates a subgroup of order $\frac{p-1}{n}$. This subgroup will be normal because the group $(\mathbb{Z}/p\mathbb{Z})^\times$ is abelian, so accordingly we set

$$\alpha \coloneqq \sum_{k \in \langle g^n \rangle} \zeta_p^k$$

so that

$$\mathrm{Gal}(\mathbb{Q}(\alpha)/\mathbb{Q}) \cong \frac{\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})}{\mathrm{Gal}(\mathbb{Q}(\alpha)/\mathbb{Q})} \cong \frac{\langle g \rangle}{\langle g^n \rangle} \cong \mathbb{Z}/n\mathbb{Z}.$$

This is what we wanted. ■

Let's see this in practice again.

Example 5.117. We find a $\mathbb{Q}$-extension with Galois group $\mathbb{Z}/7\mathbb{Z}$. We fix $p \coloneqq 29 \equiv 1 \pmod{7}$. So we would need to find the subgroup of $(\mathbb{Z}/29\mathbb{Z})^\times$ fixed by $\langle \sqrt{-1} \rangle = \langle 12 \rangle$, so we can find that our subfield is

$$\mathbb{Q}\left(\zeta_{29} + \zeta_{29}^{-1} + \zeta_{29}^{12} + \zeta_{29}^{-12}\right).$$

This is what we wanted.

As an aside, we note that our application of Dirichlet's theorem was a bit unnecessary because there are easier ways to go about this.

Lemma 5.118. Fix n a prime. Then there are infinietly many primes $p \equiv 1 \pmod{n}$.

Proof. The main idea is to look at the prime factors of

$$\Phi_n(m) \coloneqq \frac{m^n - 1}{m - 1}$$

as m varies. Indeed, if p divides $\frac{m^n-1}{m-1}$, then $p \mid m^n - 1$, so $m \pmod{p}$ will have multiplicative order dividing n. Because n is prime, the multiplicative order will thus either be 1 or n. We deal with these cases one at a time.

- If the multiplicative order is 1, then $m \equiv 1 \pmod{p}$, so

$$\frac{m^n - 1}{m - 1} = 1 + m + \cdots + m^{n-1} \equiv \underbrace{1 + 1 + \cdots + 1}_{n} = n \pmod{p}.$$

  So because p divides the left-hand side, $p \mid n$ as well.

- If the multiplicative order is n, then $n \mid \#(\mathbb{Z}/p\mathbb{Z})^\times = p-1$ by Lagrange's theorem on groups, so $n \mid p-1$.

So $p \mid \Phi_n(m)$ implies that $p \mid n$ or $p \equiv 1 \pmod{n}$. We would like to focus on the second kind of prime, so we note that $p \mid \Phi_n(nm)$ has $\Phi_n(nm) \equiv 1 \pmod{n}$, so $p \nmid n$, and $p \equiv 1 \pmod{n}$ is forced.

Now we finish the proof in a Euclidean way. We show that no finite set S contains all $1 \pmod{n}$ primes. Indeed, let the product of the primes in S be P, and we study

$$\Phi_n(knP)$$

as $k \to \infty$. In particular, for sufficiently large k, we can promise[5] $\Phi_n(knP) > 1$ so that it must have a prime factor p. By the argument above, $p \equiv 1 \pmod{p}$, but we can also see that $p \nmid P$, so $p \notin S$. This finishes. ■

Remark 5.119 (Nir)**.** Something like his can be done for more general n, using cyclotomic polynomials in a similar way.

5.4.3 Inverse Galois Problem: Symmetric Groups

We have the following exercise.

Exercise 5.120. We find an extension of $\mathbb{Q}$ with Galois group S_5.

Proof. We take L to be the splitting field of $f(x) \coloneqq x^5 - 4x + 2$, which is irreducible by Eisenstein's criterion at the prime 2. Now we have the following observations.

- Surely $\mathrm{Gal}(L/\mathbb{Q}) \subseteq S_5$ because $\mathrm{Gal}(L/\mathbb{Q})$ acts on the roots of f, and this action determines the rest of the automorphism because L is generated by the roots of f.
- The fact that $\deg f = 5$ is quintic implies that there is a subfield of degree 5, so $5 \mid [L : \mathbb{Q}] = \#\,\mathrm{Gal}(L/\mathbb{Q})$. Thus, $\mathrm{Gal}(L/\mathbb{Q})$ contains a 5-cycle by Cauchy's theorem.
- The polynomial $f(x)$ has exactly thee real roots, which we can check graphically (we won't do this here). In particular, the action of complex conjugation restricted to L induces an automorphism of $L/\mathbb{Q}$, and this automorphism must permute two roots. So $\mathrm{Gal}(L/\mathbb{Q})$ has a transposition.

But now the point is that any 5-cycle and 2-cycle in S_5 will generate all of S_5. Indeed, we have the following lemma.

Lemma 5.121. Fix p a prime. Then the p-cycle $(0, 1, \dots, p-1) \in S_p$ and any transposition $(a, b) \in S_p$ will fully generate S_p.

Proof. Without loss of generality, take $a < b$. Set $\sigma \coloneqq (0, 1, \dots, p-1)$. We see that

$$\sigma^{p-a}(a, b)\sigma^{-(p-a)} = (0, b-a),$$

so we have some transposition of the form $(0, c)$ where $c \neq 0$. Then we see that

$$\sigma^{kc}(0, c)\sigma^{-kc} = (kc, (k+1)c),$$

so we may chain

$$(c, 2c)(0, c)(c, 2c) = (0, 2c), \quad (2c, 3c)(0, 2c)(2c, 3c) = (0, 3c), \quad \cdots.$$

[5] The only reason to introduce this k variable is this end behavior argument. It is surprisingly annoying.

The point is that we can get $(0, ck)$ for any nonnegative integer k, so taking $k \equiv c^{-1} \pmod{p}$ (here we use the fact that p is prime!), we see that we can get $(0, 1)$. Repeating the above chain, we see that we can get $(0, 1)$ and $(0, 2)$ and $(0, 3)$ and so on. Further, we then see we can get

$$(0, k)(0, \ell)(0, k) = (k, \ell)$$

for any k and ℓ, so we can get any transposition. So it follows that we can indeed get all of S_p. ■

Finishing up, we number the roots so that the 5-cycle in $\mathrm{Gal}(L/\mathbb{Q})$ is (12345), and noting that conjugation gives us our transposition as above, we see that these fully generate an S_5. So we do find

$$\mathrm{Gal}(L/\mathbb{Q}) \cong S_5,$$

which is what we wanted. ■

A similar approach will work for any prime p, not just 5. We are restricted to primes to make the final argument about the transposition and 5-cycle to work.

Proposition 5.122. For any prime p, there exists a Galois extension $K/\mathbb{Q}$ with Galois group S_p.

Proof. Rigorizing this is a bit annoying, but here is one sketch: by the argument above provided in the example, we need to find an irreducible polynomial $f \in \mathbb{Q}[x]$ with degree p and exactly 2 complex roots. Well, we can start with

$$g(x) = \left(x^2 + 1\right) \cdot \prod_{k=1}^{p-2} (x - k),$$

which does indeed have exactly 2 complex roots (in particular intersecting the x-axis $p - 2$ times). Because g has no repeated roots and is locally linear, it follows there is an interval $(-\alpha, \alpha)$ such that any $\varepsilon \in (-\alpha, \alpha)$ will still have $g(x) + \varepsilon$ with exactly $p - 2$ real roots.

Because the generic polynomial is irreducible, such an $\varepsilon \in \mathbb{Q}$ should exist to make $g(x) + \varepsilon$ irreducible; for example, for sufficiently large primes q, we have $\varepsilon = \frac{1}{q^{\deg g - 1}} < \alpha$ will make

$$f(x) \coloneqq q^{\deg g} \left(g\left(\frac{x}{q}\right) + \frac{1}{q^{\deg g - 1}} \right) = q^{\deg g} g\left(\frac{x}{q}\right) + q \in \mathbb{Z}[x]$$

Eisenstein at the prime q while still having exactly $p - 2$ real roots. ■

Here is a nice consequence.

Proposition 5.123. Fix G any finite group. Then we can find an extension M/L of $\mathbb{Q}$ such that $\mathrm{Gal}(M/L) \cong G$.

Proof. The point is that, by Proposition 5.122, we may take $M/\mathbb{Q}$ to have Galois group S_p such that $G \subseteq S_p$; for example, if $p > \#G$, then we can embed $G \hookrightarrow S_{\#G}$ (by having G act on itself by left multiplication) and then embed $S_{\#G} \hookrightarrow S_p$ (by fixing the last $p - \#G$ coordinates). Then we take $L = M^G$ so that

$$\mathrm{Gal}(M/L) = \mathrm{Gal}\left(M/M^G\right) \cong G$$

by the Galois correspondence. ■

5.4.4 Cubic Polynomials

Finding the Galois group of specific polynomials is somewhat hard; let's see what we can do with cubic polynomials.

Fix $f \in K[x]$ a cubic irreducible polynomial with L/K its splitting field. As with last time, we start with the following two facts.

- The Galois group $\mathrm{Gal}(L/K)$ is contained in S_3 because S_3 will act on the roots of f, and the action on the roots will uniquely determine an automorphism because L is genreated by these roots.
- Because f is of degree 3, adjoining one root creates a cubic subextension of the splitting field, so $3 \mid [L:K] = \#\,\mathrm{Gal}(L/K)$.

So we find that must be a subgroup of S_3 with at least 3 elements, of which our options are

$$\mathrm{Gal}(L/K) \in \{A_3, S_3\}.$$

We would like to determine between the two. Quickly we verify that both are possible.

Example 5.124. The splitting field of $x^3 - 2$ over $\mathbb{Q}$ has Galois group $\mathrm{Gal}(\mathbb{Q}(\sqrt[3]{2}, \omega)/\mathbb{Q}) \cong S_3$, as we discussed in an earlier example.

Example 5.125. In the case of x^3+x+1 over $\mathbb{F}_2$, we are looking at $\mathrm{Gal}(\mathbb{F}_8/\mathbb{F}_2) \cong \mathbb{Z}/3\mathbb{Z}$, where the Galois group is generated by the Frobenius automorphism.

To describe our algorithm, we fix α, β, γ the roots of f in L. The main idea is to fix

$$\Delta \coloneqq (\alpha - \beta)(\beta - \gamma)(\gamma - \alpha).$$

Indeed, we see that Δ^2 is fully fixed by any permutation of the roots given by $\mathrm{Gal}(L/K)$, so $\Delta^2 \in K$. So the question is if $\Delta \in K$ or $\Delta \notin K$. We have the following cases.

- Take $\Delta \in K$. Then any $\sigma \in \mathrm{Gal}(L/K)$ will fix Δ, so in particular, $\mathrm{Gal}(L/K)$ does not contain the transposition (α, β). So $\mathrm{Gal}(L/K)$ is stricly contained in S_3, so $\mathrm{Gal}(L/K) \cong A_3$.
- Take $\Delta \notin K$. But we do know $\Delta^2 \in K$ and $\Delta \in L$, so the chain

  $$K \subseteq K(\Delta) \subseteq L$$

  provides a quadratic subextension of L. In particular, $2 \mid [L:K] = \#\,[Gal](L/K)$, so $\#\,\mathrm{Gal}(L/K) \geq 6$, so $\mathrm{Gal}(L/K) \cong S_3$.[6]

So we see that $\Delta \in K$ can detect the Galois group.

But now we notice that Δ^2 was just the discriminant all along, so we know how to compute this. In particular, if our cubic polynomial is $x^3 + bx + c$, then we are asking if $-4b^3 - 27c^2$ is a square in K. This gives the following result.

Proposition 5.126. Fix K a field and $f \in K[x]$ a cubic polynomial in the form $f(x) = x^3 + bx + c$. Set $\Delta^2 \coloneqq -4b^3 - 27c^2$, and we have two cases.

- If Δ^2 is a square in K, then the Galois group of f is isomorphic to $A_3 \cong \mathbb{Z}/3\mathbb{Z}$.
- If Δ^2 is not a square in K, then the Galois group of f is isomorphic so S_3.

Proof. This essentially follows from the above discussion. The point is that Δ^2 is a square in K if and only if $\pm\Delta \in K$, so we get to reduce to the casework from earlier. ∎

Let's finish with some examples.

[6] We could also argue as we did before: the fact that $\Delta \notin K$ implies that there must be an element in $\mathrm{Gal}(L/K) \setminus A_3$, so $\mathrm{Gal}(L/K) \cong S_3$.

Example 5.127. Fix $x^3 - 3x - 1 \in \mathbb{Q}[x]$. We can see this is irreducible because it has no roots in $\mathbb{Q}$. Now, we can compute the discriminant as

$$\Delta = -4(-3)^3 - 27(-1)^2 = 81,$$

which is a square in $\mathbb{Q}$, so the Galois group of $x^3 - 3x - 1$ is A_3.

Example 5.128. Fix $x^3 - x - 1 \in \mathbb{Q}[x]$. We can see this is irreducible because it has no roots in $\mathbb{Q}$. Now, we can compute the discriminant as

$$\Delta = -4(-1)^3 - 27(-1)^2 = -23,$$

which is not a square in $\mathbb{Q}$, so the Galois group of $x^3 - x - 1$ is S_3.

We close this subsection with a remark.

Remark 5.129. There is an analogous process for degree-4 polynomial, but it gets very annoying. It can be done by hand, but it requires a lot of invariant computations. In general, degree 2 is easy, 3 is fine, 4 is really annoying, and 5 and up need a computer.

5.4.5 Fundamental Theorem of Algebra

Let's give some proofs of the Fundamental theorem of algebra.

Theorem 5.130 (Fundamental theorem of algebra)**.** We have that $\mathbb{C}$ is algebraically closed.

Proof by complex analysis. We sketch a proof using Louiville's theorem. Given any nonconstant polynomial $p \in \mathbb{C}[z]$, we show that it has a root somewhere.

Suppose that p has no roots in $\mathbb{C}$, and we show that p is constant. Well, because $|p(z)| \to \infty$ as $|z| \to \infty$ (e.g., by the triangle inequality), so it follows that $|p(z)|$ has a well-defined and achieved minimum on $\mathbb{C}$ (by compactness). Set the minimum to be m so that

$$\left| \frac{1}{p(z)} \right| \le m$$

for each $z \in \mathbb{C}$. Note that the left-hand side is always well-defined because p has no roots in $\mathbb{C}$. But this makes $z \mapsto 1/p(z)$ a bounded holomorphic function on $\mathbb{C}$, so Louiville's theorem implies that $1/p(z)$ is constant, so $p(z)$ is constant. ∎

Proof by Galois theory. We pick up the following facts.

(a) Any polynomial of odd degree has a root somewhere. This is by the Intermediate value theorem because the end behavior of any odd-degree polynomial will be different going to $+\infty$ and $-\infty$.

Note that this where we are using topology in our proof; for example, this step does not work for $\mathbb{Q}$, say.

(b) All elements of $\mathbb{C}$ have a square root in $\mathbb{C}$. For example, if we write our complex number as $re^{i\theta}$, then $\sqrt{r}e^{i\theta/2}$ is a square root.

In fact, this can be extended by the quadratic formula (here we are using that the characteristic of $\mathbb{R}$ is not 2) to show that any quadratic has roots.

Now, to show that $\mathbb{C}$ is algebraically closed, we pick any element element α a root of a polynomial in $\mathbb{C}[z]$, and it will generate a splitting field $L/\mathbb{C}$. We would like to show that $\alpha \in \mathbb{C}$, for which we show $L = \mathbb{C}$.

Less specifically, we show that any finite Galois extension $L/\mathbb{C}$ will collapse to $L = \mathbb{C}$. We note that we have the tower

$$\mathbb{R} \subseteq \mathbb{C} \subseteq L,$$

so $L/\mathbb{R}$ is also Galois. Now, fix $G \coloneqq \mathrm{Gal}(L/\mathbb{R})$, and we use Galois theory to turn our theorem into a group theory problem. Then we note the following.

(a) We claim that that G has no proper subgroups of odd index; this follows from (a) earlier.

Indeed, a subgroup $H \subseteq G$ of odd index would induce a field L^H with

$$[L^H : \mathbb{R}] = \frac{[L : \mathbb{R}]}{[L : L^H]} = \frac{\#G}{\#\,\mathrm{Gal}(L/L^H)} = [G : H].$$

But there are no nontrivial extensions of $\mathbb{R}$ of odd degree because all polynomials of odd degree over $\mathbb{R}$ have a root and are not irreducible. So we must have $[L^H : \mathbb{R}] = 1$ so that $[G : H] = 1$, making H not a proper subgroup.

In particular, fixing S to be a Sylow 2-subgroup, we find that S has odd index by construction, so $S = G$. So G is a 2-group, so $\mathrm{Gal}(L/\mathbb{C}) \subseteq G$ is a 2-group.

(b) It remains to show that $\mathrm{Gal}(L/\mathbb{C})$ must be trivial. This follows from (b) earlier.

Indeed, supposing for contradiction that $\#\,\mathrm{Gal}(L/\mathbb{C}) > 1$, we see that, being a 2-group and hence nilpotent, $\mathrm{Gal}(L/\mathbb{C})$ will contain an index-2 subgroup. But this corresponds to a nontrivial quadratic extension of $\mathbb{C}$, which does not exist by (b) above.

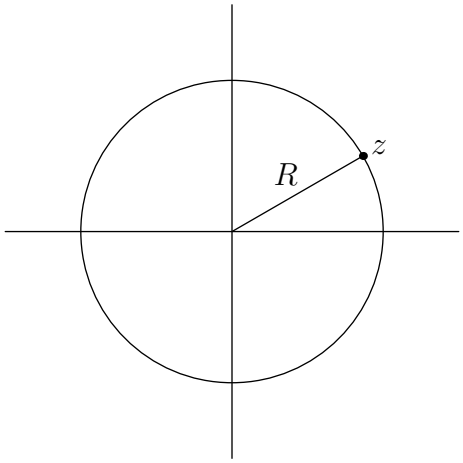
So from the above reasoning we have that $\mathrm{Gal}(L/\mathbb{C})$ must trivial, forcing $L/\mathbb{C}$ to collapse into $L = \mathbb{C}$. ■

Remark 5.131. Essentially what happened in the above proof is that we turned a result into fields into some logic about groups. Namely, the Intermediate value theorem turned into no subgroups of odd index larger than 1; and every element having a square root turned into no quadratic subextensions. The Galois theory bridges these.

Remark 5.132. We used a lot of group theory in the above proof: we used the Sylow theorems and some theory of p-groups/nilpotent groups.

Anyways let's see another proof.

Proof by winding numbers. Let's make the Fundamental theorem of algebra intuitively obvious. Fix $f \in \mathbb{C}[z]$ some polynomial, and we would like to give it a root. Imagine we have our parameter z run around a large circle of radius R.



Now, we watch what happens with f. For R large enough, then $f(z) \approx z^{\deg f}$, so $f(z)$ will loop around the origin $\deg f$ times. But if we contract R to 0, then we will go around the origin 0 times. So by the "continuity" the contraction as $R \to 0$ must send the image of f of this circle to intersect the origin, which is the root we were looking for. ■

Remark 5.133. The hard part of this proof is to verify that the number of times we go around the origin is a well-defined integer, namely $\deg f$. To rigorize this, take an algebraic topology class.

Remark 5.134. This proof actually works for $f(z) = z^n + g(z)$, where $g(z)$ is any continuous function for which $|g(z)| < R^n$ when $|z| = R$, for some given R. So perhaps this proof has little to do with polynomials.

Remark 5.135. One reason this proof is subtle and undiscovered for a while is that thinking topologically is hard.

5.4.6 Separable Extensions

And we continue with our applications.

Proposition 5.136. Fix L/K a finite separable extension. Then there are only finitely many extensions between L/K.

Proof. We extend L to some finite Galois extension M/K, say by taking a splitting field of the polynomials for some finite generating set for L/K.

But now the number of extensions between M/K is finite because they correspond to subgroups of $\mathrm{Gal}(M/K)$, which is finite because its size is $[M : K] < \infty$. Because each intermediate extension between L/K will also be between M/K, we conclude that there are finitely many intermediate extensions between L/K. ■

Importantly, inseparable extensions cannot be embedded into Galois extensions, so this proof does not work for free. Explicitly, if L/K is inseparable, then any extension M/K with L as an intermediate field will still be inseparable. Here is the standard example.

Example 5.137. Fix k an infinite field of characteristic p, and we consider the extension

$$k\left(t^p, u^p\right) \subseteq k(t, u).$$

This is an extension of degree p^2, which we can check by hand. But if x is any element of $k(t, u)$, then $x^p \in k\left(t^p, u^p\right)$ by the Frobenius automorphism, so $[k(x) : k] = p$. This gives us an infinite number of extensions of degree p; in particular no finite number of them can cover all of L because no finite number of proper K-subspaces can fully cover L.

Technically in the above example, we do need to check that proper subspaces cannot fully cover a space (over an infinite field), which requires the following lemma.

Lemma 5.138. If L is a vector space over an infinite field K. Then L is not the union of a finite number of proper subspaces.

Proof. This is surprisingly technical; we take our proof from here. Suppose for the sake of contradiction we can write

$$L = \bigcup_{k=1}^{n} L_k$$

for some proper subspaces $L_k \subsetneq L$. We show that V_1 is contained in $\bigcup_{k=2}^{n} L_k$, from which an induction will collapse the entire union to $L = L_n$, which will be a contradiction.

Well, fix any $\ell \in L_1$ so that we want to show $\ell \in \bigcup_{k=2}^{n} L_k$. We should use the fact that L_1 is proper, so we note that we can also find $\ell' \in L \setminus L_1$. Now, because K is infinite (!), we can look at the family of vectors in the form

$$\ell + \ell' k$$

as $k \in K \setminus \{0\}$ varies. None of these vectors can go into L_1, for this would imply $\ell' \in L_1$, but as they must go into one of the $L_\bullet$s of our union. In particular, because there are infinitely many of these vectors, two of them must fit into some particular L_k. But then

$$\ell + \ell' k_1, \ell + \ell' k_2 \in L_k$$

implies that $\ell \in L_k$. This finishes. ∎

In particular, this tells us that there cannot be finitely many fields $k(x)$, for these fields must generate the full L.

We remark that this gives us another proof of the Primitive element theorem, more or less by "set theory."

Theorem 5.139 (Primitive element)**.** Fix L/K a finite extension with only finitely many intermediate subfields; for example, we can take L/K finite and separable. Then there exists $\alpha \in L$ such that $L = K(\alpha)$.

Proof. For finite fields, proceed as we did before: fix a generator $g \in L^\times$, and we see that $L = K(g)$. (This is technically the only place that we use the fact that L/K is a finite extension.)

For K infinite, we let $\{L_k\}_{k=1}^{n}$ a list of the proper intermediate extensions between L/K. Then we see that

$$\bigcup_{k=1}^{n} L_k$$

is a finite union of proper K-subspaces of L, so because K is infinite, this cannot cover all of L by Lemma 5.138. In particular, fix α in L but not in the above union so that $K(\alpha)$ contains α and hence cannot be a proper intermediate extension. So $L = K(\alpha)$, finishing. ∎

Remark 5.140 (Nir)**.** Technically we may remove the condition that L/K is a finite extension, for this follows from only having finitely many intermediate subfields. Fix L/K an infinite extension, and we show that there are infinitely many intermediate subfields.

- If L/K can be generated by a single element $L = K(\alpha)$, then α must be transcendental, so the various $K(\alpha^\bullet)$ provide our intermediate subfields.
- If L/K cannot be generated by a single element, then fix $K(\alpha, \beta)$ a subextension which cannot be generated by a single element. If K is finite, then α and β cannot be algeraic, for then $K(\alpha, \beta)$ would be finite and generated by a single element; so one of α or β is transcendental, reducing the previous case.

  Otherwise K is infinite. Then we can show that $K(\alpha + k\beta)$ for various $k \in K$ will each give distinct subspaces, for $K(\alpha + k_1\beta) = K(\alpha + k_2\beta)$ for $k_1 \neq k_2$ would imply that $K(\alpha, \beta) = K(\alpha + k_1\beta) = K(\alpha + k_2\beta)$ is generated by a single element.

5.4.7 Kummer Theory Advertisement

We will focus on the following question.

Question 5.141. Suppose a Galois extension L/K has Galois group G. Then what can we say about the extension?

Here are some examples.

Exercise 5.142. Fix $G = \mathbb{Z}/2\mathbb{Z}$, and we discuss the Galois extensions L/K with $\mathrm{Gal}(L/K) \cong G$.

Proof. Our extension is Galois and in particular separable, so L/K will have $L = K(\alpha)$, and α must now be a degree-2 element because $[L : K] = 2$. Namely, we will have

$$\alpha^2 + b\alpha + c = 0$$

for some $b, c \in K$. Assuming our characterisitc is not 2, we can solve for α as in $\frac{-b \pm \sqrt{b^2 - 4c}}{2}$, so $L = K\left(\sqrt{\beta}\right)$ for some $\beta \in K$. Then here, setting $g \in G$ to be the nontrivial element, we see that we must have

$$g\sqrt{\beta} = -\sqrt{\beta}$$

because $g\sqrt{\beta} \in \{\pm\sqrt{\beta}\}$ as these are the roots of $x^2 - \beta = 0$, but g cannot fix $\sqrt{\beta}$ because this would make G fix all of $K(\beta)$.

In particular, we remark that G is acting as not just field automorphisms, but viewing L as a K-vector space, we are getting a lienar representation of G as $G \to \mathrm{Aut}(L)$ by viewing the automorphisms as linear transformation. Under this view, $\sqrt{\beta}$ is an eigenvector with eigenvalue -1. Of course, we have another eigenvector as $1 \in L$ with eigenvalue 1. The point is that

$$G \cong \left\langle \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\rangle$$

by diagonalizing with our eigenbasis $\{1, \sqrt{\beta}\}$.

However, if the characteristic of K is 2, we need to worry a bit more. Here we cannot even hope to get $L = K(\sqrt{\beta})$ because $K(\sqrt{\beta})/K$ is not separable because the minimal polynomial $x^2 - \beta = (x - \sqrt{\beta})^2$ is not a separable polynomial. So we return to our polynomial

$$x^2 + bx + c = 0.$$

Now here we should have $b \neq 0$ to make L/K separable, as just described, so we scale $x \mapsto bx$ and divide out by b^2 to get

$$x^2 + x + c' = 0,$$

for $c' := c/b^2$. The point is that we have somewhat controlled our linear term, which gives us an "Artin-Shreier polynomial." Because $1 = -1$, we may rewrite this as

$$x^2 - x - c' = 0.$$

Now, α is a root implies that $\alpha + 1$ is a root because $(\alpha + 1)^2 - (\alpha + 1) = \alpha^2 - \alpha$, so do indeed have distinct roots. So $\mathbb{Z}/2\mathbb{Z}$-extensions in characteristic 2 are controlled by polynomials in the form $x^2 - x - c = 0$.

Here we have that our nontrivial element $g \in G$ must send $\alpha \mapsto \alpha + 1$ to the other root, so expanding $g \in G$ with the basis $\{1, \alpha\}$, we find that

$$G \cong \left\langle \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right\rangle.$$

This is not diagonalizable, but it does at least have all eigenvalues equal to 1, which is called "unipotent." ∎

Example 5.143. The extension $\mathbb{F}_4/\mathbb{F}_2$ will have $\mathbb{F}_4 = \mathbb{F}_2(\omega)$, where ω is a root of $x^2 + x + 1$. Here the Galois group takes $1 \mapsto 1$ and $\omega \mapsto \omega^2 = \omega + 1$, so it behaves as described.

Let's push harder.

Exercise 5.144. Fix $G = \mathbb{Z}/p\mathbb{Z}$, and we discuss the extensions L/K with $\mathrm{Gal}(L/K) \cong G$.

Proof. As before, we have to quarantine out characteristic p, so let's start with the case where K has characteristic not equal to p. To make our lives easier, we will assume that K contains all pth roots of 1; we did not have to do this for $p = 2$ because ± 1 are always in our field.

Now, the point is that

$$K(\sqrt[p]{a})$$

is the splitting field of $x^p - a$ because the roots are $\zeta_p^{\bullet} \sqrt[p]{a} \in K(\sqrt[p]{a})$. We will continue this discussion next lecture. ■

5.5 November 23

The billboard said "The End is Near."

5.5.1 Kummer Theory in Characteristic Not Dividing n

Last lecture we were trying to describe Galois extensions L/K where $\mathrm{Gal}(L/K) \cong \mathbb{Z}/n\mathbb{Z}$. Quickly, we recall that for $n = 2$, we found the following.

Proposition 5.145. Fix L/K a Galois extension with $\mathrm{Gal}(L/K) \cong \mathbb{Z}/2\mathbb{Z}$.

- If $\mathrm{char}\, K \neq 2$, then $L = K(\sqrt{a})$ for some $a \in K$.
- If $\mathrm{char}\, K = 2$, then $L \cong K[x]/\left(x^2 - x - a\right)$ for some $a \in K$.

Remark 5.146. In particular, in the former case, $\sqrt{a}$ was an eigenvector of the generator of $\mathrm{Gal}(L/K)$. "Eigenvector" will be today's magic word.

We would like to extend this to all positive integers n.

Warning 5.147. In class, Professor Borcherds focused on the case where n is prime, for psychological reasons. For my personal benefit, I have generalized below.

To start, work with $\mathrm{char}\, K \nmid n$. Well, fix σ a generator of $\mathrm{Gal}(L/K)$ so that $\mathrm{Gal}(L/K) = \langle \sigma \rangle$, and our end goal will (roughly speaking) be to diagonalize σ. Namely, we view L as a K-vector space upon which $\mathrm{Gal}(L/K)$ acts.

Well, the eigenvalues of σ are going to be the nth roots of unity because $\sigma^n = \sigma^{\#\mathrm{Gal}(L/K)} = \mathrm{id}$: indeed, if v is an eigenvector with eigenvalue λ, then

$$v = \sigma^n(v) = \lambda^n v,$$

so λ is indeed an nth root of unity. So we will just add the assumption that K contains the pth roots of unity.

Warning 5.148. The problem becomes substantially harder in the case where K does not contain all the pth roots of unity. Namely, over a global field, it is roughly class field theory.

Note that, by the derivative trick, $X^n - 1$ is separable over K (the only root of its derivative is 0), so there are in fact n roots of unity. Because the roots of $X^n - 1$ form a finite multiplicative subgroup of $K^\times$, it is cyclic,[7] so (for convenience) fix ζ some primitive (generating) nth root of unity.

In order to claim an eigenbasis, let's go find our eigenvectors for each eigenvalue. Namely, we fix some eigenvalue ζ^k, and we want to find vectors $v \in L$ such that

$$\sigma v = \zeta^k v.$$

In other words, we want to find vectors fixed by $\sigma\zeta^{-k}$. (Here $\sigma\zeta^{-k} = \zeta^{-k}\sigma$ because σ fixes K; namely, we are using the assumption that $\zeta \in K$.) For this, we do something clever: we apply G-averages to get the eigenvectors, taking some fixed $v \in L$ and setting

$$v_k \coloneqq \sum_{\ell=0}^{n-1} \left(\sigma\zeta^{-k}\right)^\ell v.$$

Indeed, we see that applying $\sigma\zeta^{-k}$ will simply cycle the terms in the sum, so v_k does indeed have $\sigma v_k = \zeta^k v_k$. So $(*)$ gives us access to lots of eigenvectors (one from each $v \in L$), but it is technically possible that any choice $v \in L$ will give $v_k = 0$ so that we are not actually generating a dimension-1 eigenspace.

There are a few ways to finish from here. Here is one finish in the case where n is prime.

Theorem 5.149. Fix n a prime, and fix a field K with $\operatorname{char} K \nmid n$ such that K contains all nth roots of unity. Now, if L/K is a Galois extension such that $\operatorname{Gal}(L/K) \cong \mathbb{Z}/n\mathbb{Z}$, then $L = K(\sqrt[n]{a})$ for some $a \in K$ such that $(\sqrt[n]{a})^m \notin K$ for any $0 < m < n$.

Proof. We continue from the above discussion, trying to show that the claimed eigenspaces do promise an eigenbasis. The trick for this is to add our eigenvectors v_k together. We see that

$$\sum_{k=0}^{n-1} v_k = \sum_{k=0}^{n-1} \left(\sum_{\ell=0}^{n-1} \left(\sigma\zeta^{-k}\right)^\ell v \right) = \left(\sum_{\ell=0}^{n-1} \sigma^\ell \left(\sum_{k=0}^{n-1} \zeta^{-k\ell} \right) \right) v.$$

Now, for $\ell \neq 0$, the inner sum will vanish as $\frac{\zeta^{-n\ell}-1}{\zeta^{-\ell}-1} = 0$, so we only have to worry about the $\ell = 0$ term, which tells us

$$\sum_{k=0}^{n-1} v_k = nv.$$

So because $\operatorname{char} K \nmid n$ (!), we see that v is a sum of elements in each individual eigenspace. To be explicit, if we let $L_k \subseteq L$ be the eigenspace for the eigenvalue ζ^k, we have found that

$$L \subseteq \bigoplus_{k=0}^{n-1} L_k,$$

so the equality follows.

In particular, at least one of the eigenspaces is nonzero, and we cannot just have the eigenspace $L_0 = K$ be nonempty because this would imply $L = L_0 = K$. So there is some nonzero eigenvector v with eigenvalue $\zeta^k \neq 1$, but because n is prime (here we use that n is prime), ζ^k is still a primitive nth root of unity. Now, for any $m \in \mathbb{Z}$,

$$\sigma \cdot v^m = (\sigma v)^m = (\zeta^k v)^m = \zeta^{km} v^m,$$

so $v^m \in K$ if and only if v^m is fixed by $\operatorname{Gal}(L/K)$ if and only if $\zeta^{km} = 1$ if and only if $n \mid km$ if and only if $n \mid m$. So $v^n \in K$, and n is the least such positive integer. So $K(\sqrt[n]{v^n})$ is indeed a degree-n extension and will be equal to L, finishing. ∎

[7] I don't see an easy way to avoid invoking this machinery, so I have invoked it.

And here is one way to finish in the general case.

Theorem 5.150. Fix n a positive integer, and fix a field K with $\operatorname{char} K \nmid n$ such that K contains all nth roots of unity. Now, if L/K is a Galois extension such that $\operatorname{Gal}(L/K) \cong \mathbb{Z}/n\mathbb{Z}$, then $L = K(\sqrt[n]{a})$ for some $a \in K$ such that $(\sqrt[n]{a})^m \notin K$ for any $0 < m < n$.

Proof. We continue where we left off before the previous theorem, attempting to show that the claimed eigenspaces do promise an eigenbasis. That is, we would like to show that, for fixed k,

$$\sum_{\ell=0}^{n-1} \left(\sigma\zeta^{-k}\right)^{\ell} v$$

is not identically 0 for each $v \in L$. Squinting a bit harder at this, we see that we are basically trying to prove that

$$\sum_{\ell=0}^{n-1} \zeta^{-k\ell}\sigma^{\ell} \neq 0.$$

For this, we pick up the following somewhat technical lemma.

Lemma 5.151. Fix L a field. Then a finite set of automorphisms of L are L-linearly independent. In other words, given a finite set of distinct automorphisms $\{\sigma_k\}_{k=1}^n \subseteq \operatorname{Aut}(L)$, we have that

$$\sum_{k=1}^{n} a_k\sigma_k = 0$$

for $\{a_k\}_{k=1}^n \subseteq L$ implies that $a_k = 0$ for each k.

Proof. We proceed by contradiction. Suppose for the sake of contradiction that such a linearly dependent set $\{\sigma_k\}_{k=1}^n$ exists, and find a set with the smallest such n, and we will find a smaller counterexample.

Well, $\sigma_n \neq \sigma_1$, so there exists some $y \in L$ such that $\sigma_n(y) \neq \sigma_1(y)$. Then, for any $x \in L$, we see

$$a_1\sigma_1(xy) + \cdots + a_n\sigma_n(xy) = 0 \qquad \text{and} \qquad \sigma_n(y)\big(a_1\sigma_1(x) + \cdots + a_n\sigma_n(x)\big) = 0.$$

Subtracting these two equations, we find that the $a_n\sigma_n(x)\sigma_n(y)$ term will cancel out, leaving us with

$$a_1(\sigma_1(y) - \sigma_n(y))\sigma_1(x) + \cdots + a_{n-1}(\sigma_{n-1}(y) - \sigma_n(y))\sigma_n(x) = 0$$

for each $x \in L$. Namely, this a nontrivial relation between the $\{\sigma_k\}_{k=1}^{n-1}$ and so is a smaller counterexample. This finishes. ∎

In particular, we find that

$$\sum_{\ell=0}^{n-1} \zeta^{-k\ell}\sigma^{\ell} \neq 0$$

is forced, so the given eigenspace must be nonempty.

We now finish as before. We see that there exists a vector $v \in L$ with eigenvalue ζ, which satisfies, for any $m \in \mathbb{Z}$,

$$\sigma \cdot v^m = (\sigma v)^m = (\zeta^k v)^m = \zeta^{km} v^m,$$

so $v^m \in K$ if and only if v^m is fixed by $\operatorname{Gal}(L/K)$ if and only if $\zeta^{km} = 1$ if and only if $n \mid km$ if and only if $n \mid m$. So $v^n \in K$, and n is the least such positive integer. So $K(\sqrt[n]{v^n})$ is indeed a degree-n extension and will be equal to L, finishing. ∎

Remark 5.152 (Nir). This second finish for the general case was not done in class. Some reader is likely to complain that I am essentially proving Hilbert's Theorem 90 without ever saying that I am proving Hilbert's Theorem 90. My response to such readers is to try to salvage the general case without doing something like this and please tell me how so that I can adjust the above accordingly.

While we're here, we note that the converse of Theorem 5.150 is also true; the argument is in Theorem VI.6.2(ii) of Lang.

Proposition 5.153. Fix n a positive integer, and fix a field K with $\operatorname{char} K \nmid n$ such that K contains all nth roots of unity. Now, if $L = K(\sqrt[n]{a})$ for some $a \in K$ such that $(\sqrt[n]{a})^m \notin K$ for any $0 < m < n$, then L/K is a Galois extension such that $\operatorname{Gal}(L/K) \cong \mathbb{Z}/n\mathbb{Z}$.

Proof. Note L is the splitting field of the separable polynomial $X^n - a$, so L/K is Galois. In particular, fixing $\sigma \in \operatorname{Gal}(L/K)$,

$$\sigma \sqrt[n]{a} = \zeta^\bullet \sqrt[n]{a}$$

for some root $\zeta^\bullet$ of $X^n - 1$. The map $\sigma \mapsto \zeta^\bullet$ is an injective group homomorphism $\operatorname{Gal}(L/K) \to \mathbb{Z}/n\mathbb{Z}$; we would like this map to be an isomorphism.

Well, fixing σ a generator of $\operatorname{Gal}(L/K)$ (which is cyclic because it is a subgroup of $\mathbb{Z}/n\mathbb{Z}$), we see that

$$\sigma(\sqrt[n]{a}^{\#\operatorname{Gal}(L/K)}) = \sigma(\sqrt[n]{a})^{\#\operatorname{Gal}(L/K)} = \sqrt[n]{a}^{\#\operatorname{Gal}(L/K)},$$

so $\sqrt[n]{a}^{\#\operatorname{Gal}(L/K)} \in K$, so $\#\operatorname{Gal}(L/K) \geq n$. It follows $\operatorname{Gal}(L/K) \cong \mathbb{Z}/n\mathbb{Z}$, as needed. ∎

5.5.2 Kummer Theory in Characteristic p

Here we are still interested in Galois extensions L/K wth $\operatorname{Gal}(L/K) \cong \mathbb{Z}/n\mathbb{Z}$, but now we discuss $\operatorname{char} K \mid n$. However, because it matters this time, we will focus on the case where n is prime so that $n = p$.

Again, we fix σ a generator of $\operatorname{Gal}(L/K)$. We would still like to find eigenvectors, but we find that

$$X^p - 1 = (X - 1)^p,$$

so our only eigenvalue is 1. Explicitly, if v is an eigenvector of σ with eigenvalue λ, then

$$v = \sigma^p v = \lambda^p v,$$

so λ is a root of $X^p - 1$, so $\lambda = 1$. So our only eigenvectors have $\sigma v = v$, which is equivalent to $v \in K$. Thus, the entire process we did for characteristic not equal to p (namely, trying to diagonalize σ) is impossible here.

Well, if we cannot get eigenvectors, we for generalized eigenvectors. Explicitly, we see that the ring $\operatorname{End}_K(L)$ is a K-module, so $p\varphi = 0\varphi = 0$ for any $\varphi \in \operatorname{End}_K(L)$. It follows $\sigma - 1$ is nilpotent, for

$$(\sigma - 1)^p = \sigma^p - 1 = 0,$$

so we can be interested in generalized eigenvectors v such that

$$(\sigma - 1)^n v = 0$$

for some fixed n. Namely, we see that we have the increasing sequence of spaces

$$K = \ker(\sigma - 1) \subseteq \ker(\sigma - 1)^2 \subseteq \ker(\sigma - 1)^3 \subseteq \cdots \subseteq \ker(\sigma - 1)^p = L.$$

Anyways, $\ker(\sigma - 1)$ is boring, so let's look at $\ker(\sigma - 1)^2$. We want $(g - 1)^2 v_2 = 0$, and in fact, we claim we can find such a vector $v_2 \in L$ with $(g - 1)v_2 \neq 0$ as well. Indeed, fix n the smallest positive integer such that $(\sigma - 1)^n = 0$, which means that we can find $w \in L$ such that $(\sigma - 1)^{n-1} w \neq 0$. So we see that

$$v_2 \coloneqq (\sigma - 1)^{n-1} v$$

is the vector we want.

Remark 5.154 (Nir)**.** In fact, p is the smallest positive integer n such that $(\sigma-1)^n = 0$. One way to see this is by direct expansion: if we are to have

$$0 = (\sigma-1)^n = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} \sigma^k,$$

then we see that $n < p$ would imply that the above equation is a nontrivial relation of automorphisms, which cannot exist. But of course $(\sigma-1)^p = 0$ works.

The main point of saying this is that, choosing v such that $(\sigma-1)^{p-1}v \neq 0$, we can put σ in Jordan canonical form by using the basis $\{(\sigma-1)^\bullet v\}$. So these generalized eigenvectors are almost diagonalizing. Regardless, we will not need this much power for the argument.

So why does this generalized eigenvector help us? Well, we fix $a \coloneqq (\sigma-1)v_2$ so that $(\sigma-1)a = 0$, so $a \in K$. So we may take the equation

$$\sigma v_2 = v_2 + a$$

and divide through by a (note $a \neq 0$ because $v_2 \notin \ker(\sigma-1)$), giving an element $v \coloneqq v_2/a \in L$ such that

$$\sigma v = v + 1.$$

This is our analogue to finding an element $v \in L$ such that $\sigma v = \zeta v$, as we had in the characteristic not equal to p case. Namely, v seems to have the simplest possible behavior with respect to the Galois action.

Continuing with the analogy, we hope that v generates L/K, so we would like to find the minimum polynomial for our v. Well, we find that

$$\sigma\left(v^p\right) = (\sigma v)^p = (v+1)^p = v^p + 1.$$

In particular, $\sigma\left(v^p - v\right) = v^p + 1 - (v+1) = v^p - v$, so $v^p - v$ is fixed by σ and hence in K, so we find some $b \coloneqq v^p - v \in K$ such that v satisfies

$$X^p - X - b \in K[X].$$

This equation has a name.

Definition 5.155 (Artin–Shreier)**.** Equations of the form $X^p - X - b$ are called Artin–Schreier polynomials.

Anyways, we get the following.

Theorem 5.156. Fix p a prime, and fix a field K with $\operatorname{char} K = p$. Now, if L/K a Galois extension with $\operatorname{Gal}(L/K) \cong \mathbb{Z}/p\mathbb{Z}$, then there exists $b \in K$ such that

$$L \cong \frac{K[X]}{(X^p - X - b)}.$$

Proof. This mostly follows from the above discussion. Namely, we have been promised an element $v \in K$ such that v is the root of some $X^p - X - b$. Further, we see that

$$K(v) \subseteq L$$

is strictly larger than K because $v \notin K$ (else $v \in \ker(\sigma-1)$ which was hypothesized false), so $K(v) = L$ is forced because $[L:K]$ is prime, and $[K(v):K]$ is a nontrivial factor. Here is where we used that $\#\operatorname{Gal}(L/K)$ is prime. ∎

Remark 5.157 (Nir). Note that the above argument did not require K to have the pth roots of unity. This is because 1 is the only pth root of unity in characteristic p, so K already had them.

The converse is almost true.

Proposition 5.158. Fix p a prime, and fix a field K with $\operatorname{char} K = p$. Then, given $b \in K$, either $X^p - X - b$ will fully split in K or, fixing any root α of $X^p - X - b$, we have that $L \coloneqq K(\alpha)$ makes L/K a Galois extension such that $\operatorname{Gal}(L/K) \cong \mathbb{Z}/p\mathbb{Z}$.

Proof. Note that, if α is a root of $X^p - X - b$ (say, in $\overline{K}$), then $\alpha + 1$ is also a root by the Frobenius automorphism, so continuing this process gives p roots

$$\alpha, \quad \alpha + 1, \quad \alpha + 2, \quad \dots, \quad \alpha + (p-1),$$

and all these roots must be distinct, so they must all of them by Lagrange's theorem on polynomials. In particular, if an extension L/K contains any roots of $X^p - X - b$, then L will contain all of them.

We now look closer at $L \coloneqq K(\alpha)$, which forcibly contains one root and hence all of them. Now, L contains all of the roots of $X^p - X - 1$ while being generated by such a root, so L/K is normal. Further, $X^p - X - 1$ has all distinct roots, so L/K is also separable and hence Galois.

So now we may study $\operatorname{Gal}(L/K)$ more closely. Any automorphism $\sigma \in \operatorname{Gal}(L/K)$ must send α to one of the other roots $\alpha + k_\sigma$, for some $k_\sigma \in \mathbb{Z}/p\mathbb{Z}$. Because the action of an automorphism is fully defined by the action on α, we see that

$$\sigma \mapsto k_\sigma$$

gives an injective group homomorphism $\operatorname{Gal}(L/K) \to \mathbb{Z}/p\mathbb{Z}$. We have two cases.

- If $L = K$ so that $\operatorname{Gal}(L/K)$ is trivial, then $X^p - X - b$ fully splits because $L = K$ contains all the roots of $X^p - X - b$.
- Otherwise, $\operatorname{Gal}(L/K)$ is nontrivial, so because $\mathbb{Z}/p\mathbb{Z}$ has no nontrivial proper subgroups, we must have $\operatorname{Gal}(L/K) \cong \mathbb{Z}/p\mathbb{Z}$. ∎

5.5.3 Applications of Kummer Theory

Let's do an application, for fun.

Exercise 5.159. We construct $\mathbb{F}_{p^p}$ as an extension of $\mathbb{F}_p$.

Proof. By previous work with finite fields, we know that $\operatorname{Gal}(\mathbb{F}_{p^p}/\mathbb{F}_q)$ is cyclic of order p (generated by $x \mapsto x^p$). So by our work with $\mathbb{Z}/p\mathbb{Z}$-extensions in characteristic p, it suffices to note that

$$X^p - X - 1$$

is irreducible over $\mathbb{F}_q$: namely, $X^p - X - 1$ has no roots over $\mathbb{F}_p$ because $X^p - X$ fully vanishes on $\mathbb{F}_p$, so instead we must have $X^p - X - 1$ irreducible of degree p. ∎

Remark 5.160. Professor Borcherds said that this can be extended to explicitly construct $\mathbb{F}_{p^\bullet}/\mathbb{F}_p$, though I am not sure how to do this. Lang roughly asserts that the correct machinery comes from Witt vectors.

Our work above also gives the following statement.

Theorem 5.161. A polynomial $f \in K[X]$ can be solved via radicals or Artin–Schreier equations (of degree $\operatorname{char} K$) if and only if the Galois group over K is solvable.

Proof. We show the directions one at a time. Set L the splitting field of f over K.

- Suppose that the polynomial is solvable, and we want to show that the Galois group is solvable. We start by taking
$$K_1 \coloneqq K(\zeta_{(\deg f)!})$$
and work over K_1 so that it suffices to show that the Galois group of f over K_1 is solvable because K_1/K is an abelian extension. The main point of introducing K_1 is to get the mth roots of unity for any intermediate extension between K_1 and the splitting field L' of f (over K_1) because this intermediate extension will have degree less than or equal to $\deg f$.

  Now, if our equation is solvable by radicals and Artin–Schreier equations of degree $\operatorname{char} K$, then we can build it up by one such extension at a time, so there is a chain of fields
$$K \subseteq K_1 \subseteq \underbrace{K_1(\alpha_1)}_{K_2 \coloneqq} \subseteq \underbrace{K_2(\alpha_2)}_{K_3 \coloneqq} \subseteq \cdots \subseteq K_n(\alpha_n) = L',$$
where each $K_\bullet(\alpha_\bullet)/K_\bullet$ is defined so that $\alpha_\bullet$ is either a radical or the root of an Artin–Schreier equation of degree $\operatorname{char} K$. For convenience, set $n_\bullet \coloneqq [K_\bullet(\alpha_\bullet) : K_\bullet] \leq \deg f$.

  Because $K_\bullet(\alpha_\bullet)/K_\bullet$ has $K_\bullet \supseteq K_1$ containing all the $n_\bullet$th roots of unity because $n_\bullet \leq \deg f$. So by the converse to our Kummer theory work above, we see
$$\operatorname{Gal}(K_\bullet(\alpha_\bullet)/K_\bullet) \cong \mathbb{Z}/n_\bullet\mathbb{Z},$$
so taking $G_\bullet \coloneqq \operatorname{Gal}(L'/K_\bullet)$, we get the sequence of subgroups
$$\operatorname{Gal}(L'/K) \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq \langle e \rangle$$
where $\operatorname{Gal}(L/K')/G_1 \cong \operatorname{Gal}(K_1/K)$ is abelian, and $G_{k+1}/G_k \cong \operatorname{Gal}(K_{k+1}/K_k) \cong \mathbb{Z}/n_k\mathbb{Z}$ is cyclic. So indeed this sequence witnesses that $\operatorname{Gal}(L/K)$ is solvable.

  Technically, we are actually interested in $\operatorname{Gal}(L/K)$ and not $\operatorname{Gal}(L'/K)$, but we do know
$$\operatorname{Gal}(L'/K) \twoheadrightarrow \operatorname{Gal}(L/K)$$
by restriction, so the solvability of $\operatorname{Gal}(L'/K)$ implies the solvability of $\operatorname{Gal}(L/K)$.[8]

- Conversely, suppose that $\operatorname{Gal}(L/K)$ is a solvable group. Lifting up to L'/K' where $K' \coloneqq K(\zeta_{(\deg f)!})$ and $L' \coloneqq LK_1$. It is still true that L'/K is a solvable extension because $\operatorname{Gal}(L'/K)/\operatorname{Gal}(L/K) \cong \operatorname{Gal}(L'/K)$ is a cyclotomic extension and hence abelian. So taking subgroups, we get that $\operatorname{GaL}(L'/K')$ is solvable as well.

  So it suffices to show that f is solvable by radicals and Artin–Schreier equations over K', where we assume $\operatorname{Gal}(L'/K')$ is solvable. (In particular, the roots of unity we added to K_1 are legal because they are "radicals" of a sort.) Well, because $\operatorname{Gal}(L'/K_1)$ is solvable, we may build a chain of subgroups
$$G_0 \coloneqq \operatorname{Gal}(L'/K') \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq \langle e \rangle =: G_n$$
so that $G_{k+1}/G_k \cong \mathbb{Z}/p_k\mathbb{Z}$ for some prime p_k. Looking at the corresponding fields, we set $K_\bullet \coloneqq (L')^{G_\bullet}$ so that
$$K' = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n = L'$$
has $\operatorname{Gal}(K_{k+1}/K_k) \cong G_{k+1}/G_k \cong \mathbb{Z}/p_k\mathbb{Z}$. But each of these field extensions contains the p_kth roots of unity ($p_k \leq [L' : K'] \leq \deg f$), so our classification of these extensions promises that
$$K_{k+1} = K_k(\alpha_k),$$

[8] I should probably say something about solvability in short exact sequences, but I can't be bothered.

where α_k is the root of some element of K_k or the root of an Artin–Schreier equation of degree $\operatorname{char} K$. In particular, any element of L'—namely, the roots of f—can be built from radicals and solutions to Artin–Schreier equations. ■

Remark 5.162. This is where the term "solvable" comes from.

Example 5.163. Any polynomial of degree at most 4 can be solved by radicals or Artin–Schreier equations. Indeed, the groups S_1, S_2, S_3, S_4 are all solvable, and the Galois group of any polynomial of degree at most 4 is a subgroup of S_4.

Example 5.164. The polynomial x^5-4x+2 has Galois group S_5 as we showed earlier, so it is not solvable by radicals.

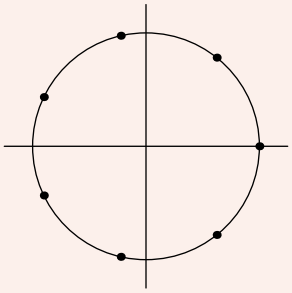
Example 5.165. Of course, $x^5 - 2$ is solvable by radicals.

In general, it is difficult to tell if an equation is solvable by radicals because finding the Galois group is difficult.

5.5.4 Cyclotomic Polynomials: Examples

We saw that roots of unity were somewhat important for our discussion, so let's study them.

Remark 5.166. "Cyclotomic" means cutting up the circle, which comes from their picture in the complex plane. For example, here are the 7th roots of unity cutting up the circle.



Cyclotomic extension are essentially the only higher-degree extensions we can control. The next easiest are Artin–Schreier extensions or adjoining nth roots, but aside from these, it is difficult to control other extensions.

In particular, we study cyclotomic extensions of $\mathbb{Q}$. To start, we need to find the minimal polynomial for some primitive nth root of unity, which we name ζ_n. We don't know it yet, but the following definition will become the minimal polynomial of ζ_n.

Definition 5.167 (Cyclotomic polynomial)**.** Given a positive integer n, we define *nth cyclotomic polynomial* as

$$\Phi_n(X) \coloneqq \prod_{\substack{1\le k\le n\\ \gcd(k,n)=1}} \left(X-\zeta_n^k\right).$$

In other words, $\Phi_n(X)$ is constructed to have roots which are the primitive nth roots of unity.

Note that, because all roots of $\Phi_n(X)$ are nth roots of unity, we will have

$$\Phi_n(X) = \prod_{\substack{1\le k\le n\\ \gcd(k,n)=1}} \left(X-\zeta_n^k\right) \;\Bigg|\; \prod_{1\le k\le n}\left(X-\zeta_n^k\right) = X^n-1.$$

However, X^n-1 will typically reduce (e.g., it is divisible by $X-1$), and it will turn out that our minimal polynomial for ζ_n will be $\Phi_n(X)$.

Example 5.168. If $n=p$ is prime, then $f(X)=\frac{X^p-1}{X-1}$ is irreducible because

$$\frac{(X+1)^p-1}{(X+1)-1} = \frac{1}{X}\sum_{k=1}^{p}\binom{p}{k}X^k = \sum_{k=1}^{p}\binom{p}{k}X^{k-1}$$

is Eisenstein at the prime p. So f is the minimal polynomial for ζ_p.

Example 5.169. If $n=p^k$ is a prime power, then

$$f(X) \coloneqq \frac{X^{p^n}-1}{X^{p^{n-1}}-1} = \sum_{k=0}^{p-1} X^{np^k}$$

is irreducible again by taking $X\mapsto X+1$ and applying Eisenstein's criterion at p. To be explicit, reducing to $\mathbb{F}_p(X)$, we have

$$f(X+1) = \frac{(X+1)^{p^n}-1}{(X+1)^{p^{n-1}}-1} = \frac{X^{p^n}+1-1}{X^{p^{n-1}}+1-1} = X^{p^n-p^{n-1}}.$$

Then to evaluate the constant term, we evaluate $f(0+1)=f(1)=p$, which is indeed not divisible by p^2. So f is the minimal polynomial for ζ_{p^k}.

However, for numbers which are not prime powers, this becomes harder. In fact, we see that if $m\mid n$, then mth roots of unity are nth roots of unity, so these need to be thrown out by hand if we want to focus on primitive nth roots of unity. For prime-powers, this is not so bad because we have relative control over divisors of prime-powers.

Exercise 5.170. We compute lots of cyclotomic polynomials.

Proof. We have the following list. We remark some properties as we go down the list, which we will rigorize in the next subsection.

- For $n=1$, our minimal polynomial is $\Phi_1(X)=X-1$.
- For $n=2$, our minimal polynomial is $\Phi_2(X)=X+1$ by dividing X^2-1 by $X-1$.
- For $n=3$, our minimal polynomial is $\Phi_3(X)=X^2+X+1$ by dividing X^3-1 by $X-1$.
- For $n=4$, our minimal polynomial is $\Phi_4(X)=X^2+1$ by dividing X^4-1 by X^2-1.
- For $n=5$, our minimal polynomial is $\Phi_5(X)=X^4+X^3+X^2+X+1$ by dividing X^5-1 by $X-1$.
- For $n=6$, we are looking at the roots of X^6-1, but we need to kill the third roots of unity as well as the square roots roots of unity. So we get

  $$X^6-1=(X-1)\left(X^2+X+1\right)(X+1)\left(X^2-X-1\right),$$

  so the one that we want is $\Phi_6(X)=X^2-X-1$.

- For $n = 7$, we get $X^6 + X^5 + X^4 + X^3 + X^2 + X + 1$ by primality.
- For $n = 8$, we need to divide $X^8 - 1$ by the first, second, and fourth roots of unity, so we get $X^4 + 1$.
- For $n = 9$, we need to divide $X^9 - 1$ by the cube roots of unity, so we get $X^6 + X^3 + 1$. (Visually, we can write down the ninth roots of unity and kill the third roots of unity by hand.)
- For $n = 10$, we need to divide $X^{10} - 1$ out by the square roots of unity and the fifth roots of unity, so we get

$$X^{10} - 1 = (X-1)(X+1)\left(X^4 + X^3 + X^2 + X + 1\right)\left(X^4 - X^3 + X^2 - X + 1\right).$$

  In particular, $\Phi_{10}(X) = \Phi_5(-X)$, which makes sense because any negative fifth root of unity will artificially gain a factor of 2 in its order, so the roots are negatives.
- For $n = 15$, we need to divide $X^{15} - 1$ out by the third roots of unity and fifth roots of unity, so we find that we want

$$\Phi_{15}(X) = \frac{\left(X^{15} - 1\right)(X-1)}{\left(X^3 - 1\right)\left(X^5 - 1\right)} = X^8 - X^7 + X^5 - X^4 + X^3 - X + 1.$$

We remark that all of the above coefficients were in $\{-1, 0, 1\}$. This is not true in general, and the smallest counterexample is 105. We will discuss this more shortly. ■

5.5.5 Cyclotomic Polynomials: Theory

Let's list some basic properties of Φ_n.

Proposition 5.171. We have the following.

(a) We have that

$$X^n - 1 = \prod_{d \mid n} \Phi_d(X).$$

(b) We have that

$$\Phi_n(X) = \prod_{d \mid n} \left(X^d - 1\right)^{\mu(n/d)},$$

where μ is the Möbius function.

Proof. We take these one at a time.

(a) This is saying that all nth roots of unity are a primitive dth root of unity for some $d \mid n$. Rigorously, we write

$$\begin{aligned}
\prod_{d \mid n} \Phi_d(X) &= \prod_{d \mid n} \left(\prod_{\substack{1 \le k \le d \\ \gcd(k,d)=1}} \left(X - e^{2\pi i k/d}\right) \right) \\
&= \prod_{d \mid n} \left(\prod_{\substack{1 \le kn/d \le n \\ \gcd(kn/d,n)=n/d}} \left(X - e^{2\pi i (k/(n/d))/d}\right) \right) \\
&= \prod_{d \mid n} \left(\prod_{\substack{1 \le k \le n \\ \gcd(k,n)=n/d}} \left(X - e^{2\pi i k/n}\right) \right).
\end{aligned}$$

Now, every k between 1 and n will have exactly one greatest common divisor with n, and this greatest common divisor will be a divisor n/d of n for some $d \mid n$. So in fact the above product is over all the k with $1 \leq k \leq n$, so we find

$$\prod_{d|n} \Phi_d(X) = \prod_{1 \leq k \leq n} \left(X - e^{2\pi i k/n}\right) = X^n - 1,$$

which is what we wanted.

(b) This comes from applying Möbius inversion to (a), in a multiplicative form. Doing this formally is somewhat annoying (we essentially have to reprove Möbius inversion), but one can see what we are supposed to do by noting we want to prove something like

$$\log \Phi_n(X) = \sum_{d|n} \log\left(X^d - 1\right) \mu\left(\frac{n}{d}\right)$$

given that

$$\log\left(X^n - 1\right) = \sum_{d|n} \log \Phi_n(X),$$

which looks more immediately like Möbius inversion. (Formalizing this would require a rigorously defined $\log$ function, but it is easier to just show the inversion by hand.) ■

The above two formulae give us a recursive way to compute cyclotomic polynomials, which will focus more on in the next subsection.

Remark 5.172 (Nir)**.** The recursion is probably the most direct way to show that $\Phi_n(X) \in \mathbb{Z}[X]$. For example, we see that

$$\Phi_n(X) = \prod_{d|n} \left(X^d - 1\right)^{\mu(n/d)} = \frac{\prod_{d|n, \mu(n/d)=1} \left(X^d - 1\right)}{\prod_{d|n, \mu(n/d)=-1} \left(X^d - 1\right)} =: \frac{f(X)}{g(X)} \in \mathbb{Q}(X).$$

But by definition, $\Phi_n(X) \in \mathbb{C}[X]$, so $\Phi_n(X) \in \mathbb{Q}(X) \cap \mathbb{C}[X] = \mathbb{Q}[X]$.

To get $\Phi_n(X) \in \mathbb{Z}[X]$, finer study is required. We see that $f(X), g(X) \in \mathbb{Z}[X]$ with $c(f) = c(g) = 1$, and $f(X) = \Phi_n(X) g(X)$, for some $\Phi_n(X) \in \mathbb{Q}[X]$. It follows from Gauss's lemma that $c(\Phi_n) = 1$, so $\Phi_n(X) = \Phi_n(X)/c(\Phi_n) \in \mathbb{Z}[X]$.

While we're here, we should probably show that Φ_n is actually an irreducible polynomial, completing the proof that Φ_n is the monic irreducible polynomial for ζ_n over $\mathbb{Q}$. Namely, we have just remarked that $\Phi_n \in \mathbb{Q}[X]$.

Proposition 5.173. We have that $\Phi_n(x)$ is irreducible (in characteristic 0).

Proof. We have done this in the case where n is prime or a prime-power using Eisenstein's criterion (see Example 5.169). Technically we did not know that those polynomials were $\Phi_{p^r}(X)$ at the time, but we an see it via the recursion now, for

$$\Phi_1(X) = X - 1 \qquad \text{and} \qquad \Phi_{p^r}(X) \stackrel{?}{=} \frac{X^{p^r} - 1}{X^{p^{r-1}} - 1} \text{ for } r \geq 1$$

satisfies

$$\prod_{d|p^r} \Phi_d(X) = \prod_{k=0}^{r} \Phi_{p^k}(X) = (X-1) \prod_{k=1}^{r} \frac{X^{p^r} - 1}{X^{p^{r-1}} - 1} = X^{p^r} - 1,$$

so get the equality by an induction.

To get that Φ_n is irreducible in general, we lift from the prime case. We have the following technical lemma.

Lemma 5.174. Fix n a positive integer and p a prime with $p \nmid n$. Further, suppose that $f(X) \in \mathbb{Z}[X]$ divides $\Phi_n(X)$ and has ζ_n as a root. Then the roots ζ of f are preserved under the mapping $\zeta \mapsto \zeta^p$.

Proof. The trick is to reduce $\pmod p$ and carry the roots of f with us. We start by fixing the root ζ_n of f. Now, we set $f_n \mid f$ to be the minimal polynomial for ζ_n in $\mathbb{Z}[X]$, and we reduce f_n to $\overline{f_n} \in \mathbb{F}_p[X]$. Picking up an irreducible factor $\overline{g}$ of $\overline{f_n}$, we see that we can induce a map

$$\mathbb{Z}[\zeta] \cong \frac{\mathbb{Z}[X]}{(f_n)} \to \frac{\mathbb{F}_p[X]}{(\overline{g})} \cong \mathbb{F}_p[\overline{\zeta_n}]$$

lifting $\mathbb{Z} \to \mathbb{F}_p[\zeta_n]$ by sending $\zeta_n \mapsto \overline{\zeta_n}$. In particular, all the work we did above was to guarantee that f_n is in the kernel of this induced map so that we may safely mod it out as we did above.

We now attempt to map the roots of f from $\mathbb{Z}[\zeta_n]$ to $\mathbb{F}_p[\overline{\zeta_p}]$. We understand the roots of f in $\mathbb{Z}[\zeta_n]$ will be a subset of the roots of $X^n - 1$, which is $\langle \zeta_n \rangle$, so we need to understand the roots of $X^n - 1$ in $\mathbb{F}_p[\overline{\zeta_n}]$.

Well, each $\overline{\zeta_n}^\bullet$ will be a root because $\overline{\zeta_n}$ is a root of $\overline{g} \mid \overline{f} \mid \Phi_n(X) \mid X^n - 1$. We claim that these are all of the roots of $X^n - 1$, for which it suffices to show that there are n distinct powers of $\overline{\zeta_n}$. This is surprisingly technical because we need to use the condition that $p \nmid n$ here.

Let m be the least positive integer such that $\overline{\zeta_n}^m = 1$, and we need to show that $m = n$; because $\overline{\zeta_n}^n = 1$, we know $m \mid n$. Observe that $\overline{\zeta_n}$ will also be a root of

$$\Phi_m(X) = \prod_{d \mid m} \left(X^d - 1\right)^{\mu(m/d)}$$

because the $X^m - 1$ factor will vanish while none of the smaller factors will. Thus, supposing for the sake of contradiction that $m < n$, we see

$$X^n - 1 = \prod_{d \mid n} \Phi_d(X)$$

has at least a double root at $\overline{\zeta_n}$—one root coming from Φ_m and one root coming from $\overline{g}(X) \mid \Phi_n(X)$. But this is impossible because $X^n - 1$ has no double roots by the derivative trick (here we use the fact $p \nmid n$)!

From all of our hard work, we see that the map

$$\zeta_n^\bullet \mapsto \overline{\zeta_n}^\bullet$$

is injective and in fact a group isomorphism $\langle \zeta_n \rangle \to \langle \overline{\zeta_n} \rangle$. In particular, we may restrict this to an injective map

$$\{\zeta \in \mathbb{Z}[\zeta_n] : f(\zeta) = 0\} \to \left\{\overline{\zeta} \in \mathbb{F}_p[\overline{\zeta_n}] : \overline{f}(\overline{\zeta}) = 0\right\},$$

which is well-defined because $f(\zeta) = 0$ implies $\overline{f}(\overline{\zeta}) = 0$. In fact, the set on the left has $\deg f$ elements, and the set on the right has at most $\deg \overline{f} = \deg f$ elements, so they both have $\deg f$ elements, so this injection is a bijection.

To finish, we see that, by the Frobenius automorphism, the roots on the right-hand side are fixed by the map $\overline{\zeta} \mapsto \overline{\zeta}^p$, so the roots on the left-hand side are fixed by $\zeta \mapsto \zeta^p$ as well. To be explicit, if ζ_n^k is a root on the left-hand side, then $\overline{\zeta_n}^k$ is a root on the right-hand side, then $\overline{\zeta_n}^{pk}$ is a root on the right-hand side, so ζ_n^{pk} is a root on the left-hand side. This finishes. ∎

The point of the lemma is to show that the Galois group of $\Phi_n(X)$ (which is $\mathbb{Q}(\zeta_n)$) is equal to $(\mathbb{Z}/n\mathbb{Z})^\times$. Certainly it is a subgroup because any $\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ must map ζ_n to some $\zeta_n^{k_\sigma}$ for $k_\sigma \in (\mathbb{Z}/n\mathbb{Z})^\times$ and is uniquely determined by this action. So

$$\sigma \mapsto k_\sigma$$

gives an injective homomorphism $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \to (\mathbb{Z}/n\mathbb{Z})^\times$.

But now, by Lemma 5.174, the Galois group of $\Phi_n(X)$ contains

$$\zeta_n \mapsto \zeta_n^p$$

for each p coprime to n, but these elements actually generate $(\mathbb{Z}/n\mathbb{Z})^\times$ by simply prime-factoring a number in each of the various equivalence classes of $(\mathbb{Z}/n\mathbb{Z})^\times$.

So to finish, we see that

$$\deg \Phi_n = \varphi(n) = \#(\mathbb{Z}/n\mathbb{Z})^\times = [\mathbb{Q}(\zeta_n) : \mathbb{Q}],$$

so in fact Φ_n must be the minimal irreducible polynomial of ζ_n. ■

5.5.6 Cyclotomic Polynomials: Computation

Warning 5.175. This subsection covers more explicit computation of cyclotomic polynomials, which was not covered in class. The main point here is to find the smallest n for which $\Phi_n(X)$ has a coefficient outside of $\{-1, 0, 1\}$.

As a quick example before doing any theory, we evaluate cyclotomic polynomials for semiprimes.

Exercise 5.176. Fix p and q distinct primes. Then $\Phi_{pq}(X)$ only has coefficients in $\{-1, 0, 1\}$.

Proof. Because p and q are distinct, we see that

$$\Phi_{pq}(X) = \prod_{d|n} \left(X^d - 1\right)^{\mu(pq/d)} = \frac{(X^{pq}-1)(X-1)}{(X^p-1)(X^q-1)} = \frac{X-1}{X^q-1}\sum_{n=0}^{q-1} X^{pn}.$$

Expanding this out, we see that

$$\Phi_{pq}(X) = \sum_{n=0}^{q-1} \frac{X^{pn+1} - X^{pn}}{X^q - 1} = \sum_{n=0}^{q-1} \frac{X^{pn+1}}{X^q-1} - \sum_{n=0}^{q-1} \frac{X^{pn}}{X^q-1}.$$

Now, we see that, by polynomial division, we get

$$\frac{X^N}{X^q-1} = \sum_{k=1}^{\lfloor n/q \rfloor} X^{N-qk} + \frac{X^{N \pmod q}}{X^q-1},$$

where $N \pmod q$ is referring specifically to the smallest nonnegative integer in the residue class. Summing over all of our N, we get

$$\Phi_{pq}(X) = \left(\sum_{n=0}^{q-1} \sum_{k=1}^{\lfloor (pn+1)/q \rfloor} X^{pn+1-qk} - \sum_{n=0}^{q-1} \sum_{k=1}^{\lfloor pn/q \rfloor} X^{pn-qk}\right) + \left(\sum_{n=0}^{q-1} \frac{X^{pn \pmod q}}{X^q-1} - \sum_{n=0}^{q-1} \frac{X^{pn+1 \pmod q}}{X^q-1}\right).$$

We note that the last two sums on the right-hand side will cancel out because $pn \pmod q$ and $pn+1 \pmod q$ will both loop over all possible residue classes $\pmod q$ because p and q are coprime. Thus,

$$\Phi_{pq}(X) = \sum_{n=0}^{q-1} \sum_{k=1}^{\lfloor (pn+1)/q \rfloor} X^{pn+1-qk} - \sum_{n=0}^{q-1} \sum_{k=1}^{\lfloor pn/q \rfloor} X^{pn-qk}. \tag{$*$}$$

By the Chinese remainder theorem, we see that $(n, k) \mapsto pn - qk$ from $\mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/pq\mathbb{Z}$ is a bijection, so $(n, k) \mapsto pn - qk + 1$ is also bijection. So because the inner sums have k range over at most $[1, q]$, we see that each outer sum above will have no repeated $X^\bullet$ terms.

So when we collect $\Phi_{pq}(X)$, we see that any coefficient $X^\bullet$ gets at most $+1$ from the left sum of $(*)$ and gets at most -1 from the right sum. So each $X^\bullet$ will have coefficient contained in $\{-1, 0, 1\}$. ■

Now let's see some small things we can do with our recursion.

Lemma 5.177. Fix n a positive integer and p a prime.

(a) If $p \nmid n$, then $\Phi_{np}(X) = \Phi_n\left(X^p\right)/\Phi_n(X)$.

(b) If $p \mid n$, then $\Phi_{np}(X) = \Phi_n\left(X^p\right)$.

Proof. The main idea is to use the Möbius inversion formula, from which we find

$$\Phi_{np}(X) = \prod_{d \mid np} \left(X^d - 1\right)^{\mu(np/d)}.$$

We now split into cases.

(a) Take $p \nmid n$. Here we see that $d \mid np$ has two cases: either $p \mid d$ or $p \nmid d$, but either way, $d/p^{\nu_p(d)}$ divides n. In particular, divisors $d \mid np$ such that $p \mid d$ will only have one power of p, so we can parameterize these by the divisor d/p of n. Namely,

$$\Phi_{np}(X) = \underbrace{\prod_{d \mid n} \left(X^d - 1\right)^{\mu(np/d)}}_{p \nmid d} \cdot \underbrace{\prod_{d \mid n} \left(X^{dp} - 1\right)^{\mu(np/(dp))}}_{p \mid dp}.$$

The first factor is

$$\prod_{d \mid n} \left(X^d - 1\right)^{\mu(np/d)} = \prod_{d \mid n} \left(X^d - 1\right)^{-\mu(n/d)} = \frac{1}{\Phi_n(X)}$$

where the Möbius function got a sign because of the extra prime $p \nmid n$. The second factor is

$$\prod_{d \mid n} \left(X^{dp} - 1\right)^{\mu(np/(dp))} = \prod_{d \mid n} \left(\left(X^p\right)^d - 1\right)^{\mu(n/d)} = \Phi_n\left(X^p\right).$$

So indeed, $\Phi_{np}(X) = \Phi_n\left(X^p\right)/\Phi_n(X)$.

(b) Take $p \mid n$. Then we see that each $d \mid np$ giving $\mu(np/d) \neq 0$ had better have $p \mid d$, for otherwise np is divisible by p^2 so that $\mu(np/d) = 0$. So we only care about divisors $d \mid np$ such that $p \mid np$, which again we can parameterize by the underlying divisor $d/p \mid n$. So we see that

$$\Phi_{np}(X) = \prod_{d \mid np} \left(X^d - 1\right)^{\mu(np/d)} = \prod_{d \mid n} \left(X^{pd} - 1\right)^{\mu(np/(dp))} = \prod_{d \mid n} \left(\left(X^p\right)^d - 1\right)^{\mu(n/d)} = \Phi_n\left(X^p\right),$$

which is what we wanted. ∎

Example 5.178. If n is odd, then we claim $\Phi_{2n}(X) = \Phi_n(-X)$. We could see this directly by studying the primitive $2n$th roots of unity and finding they are all $-\zeta_n^\bullet$. Alternatively, we see that any divisor $d \mid n$ will be odd, so

$$\Phi_n(X)\Phi_n(-X) = \prod_{d \mid n} \left(X^d - 1\right)^{\mu(n/d)} \left((-X)^d - 1\right)^{\mu(n/d)} = \prod_{d \mid n} (-1)^{\mu(n/d)} \left(X^{2d} - 1\right)^{\mu(n/d)}.$$

Because $\deg \Phi_n(X) = \varphi(n)$ is even, we know in advance that $\Phi_n(X)\Phi_(-X)$ is the product of monic polynomials, so the above becomes an equality. So $\Phi_{2n}(X) = \Phi_n\left(X^2\right)/\Phi_n(X) = \Phi_n(-X)$, finishing.

The point of these results above is that it implies that many $\Phi_n(X)$ will have coefficients in $\{-1, 0, 1\}$.

Exercise 5.179. Fix p and q distinct odd primes and suppose $n := 2^a p^b q^c$ for nonnegative integers a, b, c; in other words, n has at most two distinct odd prime divisors. Then $\Phi_n(X)$ only has coefficients in $\{-1, 0, 1\}$.

Proof. We have the following cases.

- If $\{a, b, c\} = \{0\}$, then we are looking at $\Phi_1(X) = X - 1$.
- If two of $\{a, b, c\}$ are zero while the third is nonzero, then we are evaluating
$$\Phi_{r^d}(X) = \sum_{k=0}^{d-1} X^{r^k}$$
for some prime-power r^k, so indeed, our coefficients are $\{-1, 0, 1\}$.
- If one of $\{a, b, c\}$ is zero while the other two are nonzero, then we are rename p and q so that $n = p^a q^b$ for primes p, q where $a, b > 0$. Now, by inductively applying Lemma 5.177 part (b), we see that
$$\Phi_{p^a q^b}(X) = \Phi_{pq}\left(X^{p^{a-1}q^{b-1}}\right).$$
So because $\Phi_{pq}(X)$ has coefficients in $\{-1, 0, 1\}$, we see that $\Phi_{p^a q^b}(X)$ will also have coefficients in $\{-1, 0, 1\}$.
- Lastly, take all of $\{a, b, c\}$ nonzero. Again inductively applying Lemma 5.177, we see that
$$\Phi_{2^a p^b q^c}(X) = \Phi_{2pq}\left(X^{2^{a-1}p^{b-1}q^{c-1}}\right).$$
Now, pq is odd, so $\Phi_{2^a p^b q^c}(X) = \Phi_{pq}\left(-X^{2^{a-1}p^{b-1}q^{c-1}}\right)$ will still have all coefficients in $\{-1, 0, 1\}$ again from $\Phi_{pq}(X)$. ■

In particular, if we are to have coefficients outside of $\{-1, 0, 1\}$, we must have at least three distinct odd prime divisors, for which $\Phi_{3\cdot5\cdot7}(X) = \Phi_{105}(X)$ is the first candidate. This polynomial has degree 48, so actually computing it by hand would be quite annoying, but indeed it does have a
$$-2X^{41}$$
term. So $\boxed{\Phi_{105}(X)}$ is the first cyclotomic polynomial with a coefficient outside of $\{-1, 0, 1\}$.

5.5.7 Cyclotomic Polynomials: Application

As an application, we show a special case of Dirichlet's theorem on arithmetic progressions.

Exercise 5.180. We show that there are infinitely many primes $1 \pmod n$ for each positive integer n.

Proof. The main point is to take primes $p \mid \Phi_n(b)$ for some b. We have the following lemma.

Lemma 5.181. Fix n a positive integer and $p \nmid n$ a prime factor of $\Phi_n(b)$ for some integer b. Then the order of $b \pmod p$ is n.

Proof. Now, we see that

$$b^n \equiv 1 \pmod{p}$$

because $\Phi_n(b) \mid b^n - 1$, so the order of $b \pmod{p}$ divides into n. Because the roots of $\Phi_n(X) \mid X^n - 1$ are distinct $\pmod{p}$ when $p \nmid b$, we can be sure that $b \pmod{p}$ has exactly the order n. To be explicit, we proceed as in Lemma 5.174: if the order is $m \mid n$, with $m < n$, then b will be a root of

$$\Phi_m(X) = \prod_{d \mid m} \left(X^d - 1\right)^{\mu(m/d)},$$

which forces

$$X^n - 1 = \prod_{d \mid n} \Phi_d(X)$$

to have a double root, which is a contradiction. ∎

In particular, given $p \mid \Phi_n(b)$ where $p \nmid n$, we see that applying Lagrange's theorem to $(\mathbb{Z}/p\mathbb{Z})^\times$, we find that our element of order n witnesses $p \equiv 1 \pmod{n}$.

So to finish, we suppose for the sake of contraction that there are finitely many primes which are $1 \pmod{n}$. Then let their (finite) product be P, and we look at the polynomial

$$\Phi_n(PnX).$$

Any prime p dividing into this will be coprime to n and P (because $\Phi_n(PnX) \equiv \Phi_n(0) \equiv \pm 1 \pmod{nP}$), forcing $p \equiv 1 \pmod{n}$ by the argument above.

But now, sending $X \to \infty$ will make $\Phi_n(PnX)$ number large enough to be at least 1 and hence have a prime factor, so we have a prime $p \equiv 1 \pmod{n}$ not dividing P, which is our contradiction. ∎

Remark 5.182. This is not a good way to find $1 \pmod{n}$ primes because $\Phi_n(PnX)$ will get quite large quite quickly. For example, to find $1 \pmod{10}$ primes, we want divisors of

$$\Phi_{10}(X) = X^4 - X^3 + X^2 - X + 1,$$

but as soon as we have one prime 11, we want to compute $\Phi_{10}(110x)$, which is huge.

Example 5.183. In the case of $\Phi_4(X) = X^2 + 1$, we see that all prime factors of $X^2 + 1$ for X even will be $1 \pmod{4}$.

Next lecture will be on the coming Tuesday because there is some sort of holiday or something.

5.6 November 30

A dying man can do nothing easy.

5.6.1 Inverse Galois Problem: Abelian Groups

Let's continue or discussion of cyclotomic fields. For example, last time we showed that there are infinitely many primes $1 \pmod{n}$ for any positive integer n, and we will use this fact to solve the inverse Galois problem in the abelian case.

Proposition 5.184. Fix G a finite abelian group. Then we can find a Galois extension $K/\mathbb{Q}$ such that $\mathrm{Gal}(K/\mathbb{Q}) \cong G$.

Proof. As usual, our approach is to find some extension $L/\mathbb{Q}$ with a homomorphism $\mathrm{Gal}(L/\mathbb{Q}) \twoheadrightarrow G$ and take quotients to finish. The main idea is to see that G, being finite, can be written as

$$G \cong \prod_{k=1}^{n} \mathbb{Z}/m_k\mathbb{Z}$$

for some positive integers $\{m_k\}_{k=1}^n$, not necessarily coprime. Setting $N = m_1 \cdot m_2 \cdot \ldots \cdot m_n$ so that we see

$$G \cong \prod_{k=1}^{n} \frac{\mathbb{Z}/N\mathbb{Z}}{m_k\mathbb{Z}/N\mathbb{Z}},$$

so there is a surjection $(\mathbb{Z}/N\mathbb{Z})^n \twoheadrightarrow G$ given above, so we focus on creating an extension $L/\mathbb{Q}$ with Galois group such that $\mathrm{Gal}(L/\mathbb{Q}) \twoheadrightarrow (\mathbb{Z}/N\mathbb{Z})^n$. For this, we select n primes $\{q_k\}_{k=1}^n$ such that $q_k \equiv 1 \pmod N$ and set

$$M \coloneqq q_1 \cdot q_2 \cdot \ldots \cdot q_n$$

so that

$$\mathrm{Gal}(\mathbb{Q}(\zeta_M)/\mathbb{Q}) \cong (\mathbb{Z}/M\mathbb{Z})^\times \cong \prod_{k=1}^{n} (\mathbb{Z}/q_k\mathbb{Z})^\times \cong \prod_{k=1}^{n} \mathbb{Z}/(q_k-1)\mathbb{Z}$$

which has a surjection

$$\prod_{k=1}^{n} \mathbb{Z}/(q_k-1)\mathbb{Z} \twoheadrightarrow \prod_{k=1}^{n} \frac{\mathbb{Z}/(q_k-1)\mathbb{Z}}{N\mathbb{Z}/(q_k-1)\mathbb{Z}} \cong (\mathbb{Z}/N\mathbb{Z})^n \twoheadrightarrow G.$$

So, to finish, set H to be the kernel of this surjection $\mathrm{Gal}(L/\mathbb{Q}) \twoheadrightarrow G$ and define $K \coloneqq L^H$. Then we see that H is the kernel, so it is a normal subgroup of $\mathrm{Gal}(L/\mathbb{Q})$, and we compute

$$\mathrm{Gal}(K/\mathbb{Q}) \cong \frac{\mathrm{Gal}(L/\mathbb{Q})}{\mathrm{Gal}(L/K)} = \frac{\mathrm{Gal}(L/\mathbb{Q})}{H} \cong G,$$

where we are using our discussion of normal subgroups to work this out. This finishes. ∎

The above statement even has a partial converse.

Theorem 5.185 (Kronecker–Weber)**.** Any abelian extension $K/\mathbb{Q}$ is contained in a cyclotomic extension.

Proof. This is a pretty difficult theorem in number theory, more or less a primer on class field theory. For example, it would make a good capstone for a first course on algebraic number theory. Anyways, we will not prove this here. ∎

Verifying Kronecker–Weber is not even easy for quadratic extensions, but it is not out of reach. We will do this, for fun.

Exercise 5.186 (Nir)**.** Fix an integer m. Then there exists an integer n such that $\mathbb{Q}(\sqrt{m}) \subseteq \mathbb{Q}(\zeta_n)$.

Proof. The approach will be to focus on the case where m is prime and build up from there, so fix $p \coloneqq m$ a prime. As a first guess, we check the quadratic subextension of $\mathbb{Q}(\zeta_p)$, which will almost work. The main idea, now, is to use "Gauss sums" as we did Exercise 5.98. As motivation, we fix

$$\alpha = \sum_{k=1}^{p-1} \zeta_p^{k^2},$$

which by Lemma 5.113 will generate the subfield of $\mathbb{Q}(\zeta_p)$ fixed by the squares of $(\mathbb{Z}/p\mathbb{Z})^\times$, which is the index-2 subgroup of $(\mathbb{Z}/p\mathbb{Z})^\times$, so α will generate the quadratic subfield of $\mathbb{Q}(\zeta_p)$.

To make α better-behaved, we fix

$$G \coloneqq \frac{1}{2}\alpha - (-1) = \frac{1}{2}\alpha - \sum_{k=1}^{p-1} \zeta_p^k = \sum_{k \in \mathbb{F}_p^\times} \left(\frac{k}{p}\right) \zeta_p^k,$$

where $\left(\frac{k}{p}\right)$ is 1 when $k \pmod p$ is a nonzero square, -1 when $k \pmod p$ is not a square, and 0 when $k \equiv 0$. (This last equality above is checked by casework on the various k.) Now, G will still generate the quadratic subfield of $\mathbb{Q}(\zeta_p)$, and G is better-behaved because

$$G^2 = \left(\sum_{k \in \mathbb{F}_p^\times} \left(\frac{k}{p}\right) \zeta_p^k\right) \left(\sum_{\ell \in \mathbb{F}_p^\times} \left(\frac{\ell}{p}\right) \zeta_p^\ell\right) = \sum_{k,\ell \in \mathbb{F}_p^\times} \left(\frac{k\ell}{p}\right) \zeta_p^{k+\ell}.$$

Setting x so that $\ell = kx$, we see

$$G^2 = \sum_{k,x \in \mathbb{F}_p^\times} \left(\frac{k^2 x}{p}\right) \left(\zeta_p^k\right)^{1+x} = \sum_{x=1}^{p-1} \left(\frac{x}{p}\right) \left[\sum_{k=1}^{p-1} \left(\zeta_p^k\right)^{1+x}\right]$$

When $x \neq p-1$, the inner sum will cycle as the sum of all the primitive pth roots of unity and produce -1. When $x = p-1$, we simply accumulate $p-1$, which in total gives

$$G^2 = \sum_{x=1}^{p-2} \left(\frac{x}{p}\right)(-1) + \left(\frac{-1}{p}\right)(p-1).$$

Adding back in the $x = p-1$ term to the sum, we see $\sum_{x=1}^{p-1} \left(\frac{x}{p}\right) = 0$ because $\left(\frac{\cdot}{p}\right) : \mathbb{F}_p^\times \to \mathbb{C}$ is a nontrivial character. So this term will cancel, leaving us with $G^2 = \left(\frac{-1}{p}\right) p$.

It's not too hard to show that $G^2 = (-1)^{(p-1)/2} p$, in fact, but we will not do this. The point is that $G = \pm\sqrt{\pm p}$, for some particular choice of signs. So we see that, the quadratic subfield of $\mathbb{Q}(\zeta_p)$ is either $\mathbb{Q}(\sqrt{p})$ or $\mathbb{Q}(\sqrt{-p}) = \mathbb{Q}(i\sqrt{p})$, so in either case, $\mathbb{Q}(i, \zeta_p)$ will surely contain $\sqrt{p}$.

So to finish, fix m a general integer. Then, m has a prime factorization, and let $p_1, p_2, \ldots, p_n$ be the prime factors of m. Then, adding in a sign $\varepsilon \in \{\pm 1\}$ for m, we see

$$\sqrt{m} = \sqrt{\varepsilon} \prod_{k=1}^{n} (\sqrt{p_k})^{\nu_{p_k}(m)} \in \mathbb{Q}(i, \zeta_{p_1}, \zeta_{p_2}, \ldots, \zeta p_n) \subseteq \mathbb{Q}(\zeta_{4p_1p_2\cdots p_n}),$$

finishing. ∎

5.6.2 Inverse Galois Problem: Solvable Groups

Here is an extension.

Theorem 5.187 (Shafarevich)**.** Any finite, solvable group is the Galois group of some Galois extension $K/\mathbb{Q}$.

This theorem is hard to prove, and let's give some reasoning why.

Not a proof. Here is one attempt: suppose that we are given a short exact sequence

$$1 \to A \to B \to C \to 1$$

such that A and C are Galois groups, and we want to show it for B as well. For this, we might want to take an extension $K/\mathbb{Q}$ with Galois group C and then try to extend this up to an extension $L \supseteq K$ with Galois group $\mathrm{Gal}(L/\mathbb{Q}) = B$. ∎

However, this approach might not even be possible. Namely, if we have a short exact sequence

$$1 \to A \to B \to C \to 1$$

such that C is the Galois group of $L/\mathbb{Q}$, there need not exist an extension $M \supseteq L$ with $M/\mathbb{Q}$ having group B. For example, consider the short exact sequence

$$0 \to \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/4\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0.$$

Here, $\mathbb{Z}/2\mathbb{Z}$ is a quadratic extension, say $\mathbb{Q}(\sqrt{n})$, and we would like to find L such that $\mathrm{Gal}(L/\mathbb{Q}) \cong \mathbb{Z}/4\mathbb{Z}$. The following lemma provides a physical obstruction for this.

Lemma 5.188. Fix n a squarefree integer. Then, if there exists a Galois extension $L/\mathbb{Q}$ containing $\mathbb{Q}(\sqrt{n})$ such that $\mathrm{Gal}(L/\mathbb{Q}) \cong \mathbb{Z}/4\mathbb{Z}$, then there exist rationals $b, c \in \mathbb{Q}$ such that $b^2 - nc^2 = -1$.

Proof. Well, suppose that we can do this so that $\mathrm{Gal}(L/\mathbb{Q}) = \langle\sigma\rangle \cong \mathbb{Z}/4\mathbb{Z}$. Then $\mathbb{Q}(\sqrt{n})$ is going to be the field fixed by $\langle\sigma^2\rangle$ because $\langle\sigma^2\rangle$ is the only index-2 subgroup of $\langle\sigma\rangle$.

The main idea, now, is that $L/\mathbb{Q}$ is a cyclic extension, and even though $\mathbb{Q}$ does not contain all fourth roots of unity, we can try to imitate our Kummer theory.

Much of Kummer theory was concerned with verifying the existence of the correct generating element, but ours will be somewhat easy to find: the extension $L/\mathbb{Q}(\sqrt{n})$ must have some $a \in \mathbb{Q}(\sqrt{n})$ such that $\alpha \coloneqq \sqrt{a}$ has $L = \mathbb{Q}(\sqrt{n})(\alpha)$.

Now, our magic word continues to be "eigenvalue." We do still have $\sigma^4\alpha = \alpha$, but this no longer helps us because, when we write $\lambda \coloneqq \sigma\alpha/\alpha$ so that

$$\sigma\alpha = \lambda\alpha,$$

we might not have $\lambda \in \mathbb{Q}$. However, looking at $L/\mathbb{Q}(\sqrt{n})$, we see that

$$\sigma^2\alpha = -\alpha$$

because α was chosen an element with eigenvalue -1, and -1 will certainly be fixed by σ. So to salvage our approach, we notice

$$\sigma^2\left(\frac{\sigma\alpha}{\alpha}\right) = \frac{-\sigma\alpha}{-\alpha} = \frac{\sigma\alpha}{\alpha},$$

so $\lambda \in \mathbb{Q}(\sqrt{n})$, which is the next best thing.

Continuing to salvage our Kummer theory, instead of using $\sigma\alpha = \lambda\alpha$ to pin down λ, we notice that $\sigma^2\alpha = -\alpha$ will give

$$-\alpha = \sigma^2\alpha = \sigma(\lambda\alpha) = \sigma\lambda \cdot \sigma\alpha = (\sigma\lambda \cdot \lambda)\alpha.$$

Thus, $\lambda \cdot \sigma\lambda = -1$.

To finish, we set $\lambda \coloneqq b + c\sqrt{n}$ with $b, c \in \mathbb{Q}$. Then $\sigma\sqrt{n} \neq \sqrt{n}$ (because $\sigma \notin \mathrm{Gal}(L/\mathbb{Q}(\sqrt{n})) = \langle\sigma^2\rangle$), so $\sigma\sqrt{n} = -\sqrt{n}$ instead, implying

$$\lambda \cdot \sigma\lambda = \left(b + c\sqrt{n}\right)\left(b - c\sqrt{n}\right) = b^2 - nc^2.$$

Thus, $b^2 - nc^2 = -1$. This finishes. ∎

Remark 5.189 (Nir)**.** The condition $b^2 - nc^2 = -1$ is actually effective: take $\alpha \coloneqq \sqrt{b - c\sqrt{n}}$, which has minimal polynomial $f(X) = \left(X^2 - b\right)^2 - nc^2$. We can see that $(b + c\sqrt{n})\sqrt{b - c\sqrt{n}} = \pm\sqrt{-b + c\sqrt{n}}$, which is another root of $f(X)$, so there is indeed an isomorphism σ defined as the composite

$$L = \mathbb{Q}(\sqrt{b - c\sqrt{n}}) \cong \frac{\mathbb{Q}[X]}{(f(X))} \cong \mathbb{Q}(\pm\sqrt{-b + c\sqrt{n}}) = \mathbb{Q}\left((b + c\sqrt{n})\sqrt{b - c\sqrt{n}}\right) = L,$$

sending $\sigma\alpha = (b + c\sqrt{n})\alpha$. We can check by hand σ has order four, which finishes verifying that $L/\mathbb{Q}$ has $\#\,\mathrm{Aut}(L/\mathbb{Q}) \geq 4$, so $L/\mathbb{Q}$ is Galois with $\mathrm{Gal}(L/\mathbb{Q}) = \langle\sigma\rangle \cong \mathbb{Z}/4\mathbb{Z}$.

And let's see this in action.

Example 5.190. If $n < 0$, then $b^2 - cn^2 \geq b^2 \geq 0 > -1$, so there exists no Galois extension $L/\mathbb{Q}$ containing $\mathbb{Q}(\sqrt{n})$ such that $\mathrm{Gal}(L/\mathbb{Q}) \cong \mathbb{Z}/4\mathbb{Z}$. To be explicit, $\mathbb{Q}(i)$ cannot be embedded into a $\mathbb{Z}/4\mathbb{Z}$-extension.

Example 5.191. Our work with Gauss sums in Exercise 5.186 shows that $\sqrt{5} = \pm G \in \mathbb{Q}(\zeta_5)$, so we can embed $\mathbb{Q}(\sqrt{5}) \subseteq \mathbb{Q}(\zeta_5)$. And indeed, we see that $2^2 - 5 \cdot 1^2 = -1$ is a solution, as needed.

Example 5.192. Taking $n = 3$, we are trying to solve $b^2 - 3c^2 = -1$ for $b, c \in \mathbb{Q}$, which becomes

$$x^2 + z^2 = 3y^2 \tag{$*$}$$

for some $x, y, z \in \mathbb{Z}$ and $z \neq 0$ after setting z to be the least common multiple of the denominators of b and c. However, by checking $\pmod 3$, we can see $z^2 \equiv -x^2$ forces z and x to be divisible by 3, which forces y to be divisible by 3, meaning that solutions to $(*)$ can be transformed under

$$(x, y, z) \mapsto (x/3, y/3, z/3).$$

However, this implies that z is divisible by infinitely many powers of 3, which does not make sense.

Remark 5.193. The actual proof of Shafarevich's theorem involves a lot of back-tracking to attempt to find bigger extensions, finding that sometimes we cannot do this, and then going backwards.

5.6.3 Division Rings

Let's continue with our applications of cyclotomic polynomials. Here is the object we will focus on.

Definition 5.194 (Division ring)**.** A *division ring* K is a ring K with identity (but not necessarily commutative) such that every element has a left and right multiplicative inverse.

Example 5.195. Any field is a division ring.

Example 5.196. The quaternions $\mathbb{H}$ make a noncommutative division ring.

Here is our theorem.

Theorem 5.197 (Wedderburn)**.** Any finite division ring is a field.

Proof. The proof is in two steps: write down the conjugacy class equation and then apply some theory of cyclotomic polynomials. We set

$$\mathbb{F}_q \coloneqq \{a \in K : ab = ba \text{ for each } b \in K\}.$$

It is not too hard to see that $\mathbb{F}_q$ contains 1 and 0, is closed under addition ($(a_1 + a_2)b = a_1b + a_2b = ba_1 + ba_2 = b(a_1 + a_2)$), and closed is closed under multiplication ($a_1a_2b = ba_1a_2$), so in fact $\mathbb{F}_q$ is a subring of K where multiplication commutes and hence is a field. To finish our set-up, we note that K is an abelian group containing $\mathbb{F}_q$ and hence will behave like an $\mathbb{F}_q$-vector space. In particular, $\#K = q^{[K:\mathbb{F}_q]}$ is a power of q; set $n \coloneqq [K : \mathbb{F}_q]$.

Our end goal is to show that $K = \mathbb{F}_q$. We now apply our first trick, writing down the conjugacy class equation of $K^\times$ as

$$\#K^\times = \sum_{\mathcal{C} \subseteq K^\times} \#\mathcal{C} = \#\mathbb{F}_q^\times + \sum_{\substack{[x] \subseteq K^\times \\ \#[x]>1}} \#[x],$$

where our sums are over distinct conjugacy classes. Now, by the Orbit-stabilizer theorem, the size of $[x]$ for some $a \in K$ is equal to $\#K^\times / \operatorname{Stab}(x)$, where

$$\operatorname{Stab}(x) = \left\{a \in K^\times : ax = xa\right\}.$$

Essentially the same checks as before verify that $C(x) \coloneqq \{a \in K : ax = xa\}$ contains 1 and 0 and is closed under addition and multiplication, so we won't write them out. So $C(x)$ is a subring of K and in particular also a $\mathbb{F}_q$-vector space, so it will have size $q^{[C(x):\mathbb{F}_1]}$. So, throwing out 0 as appropriate,

$$\#K^\times = (q-1) + \sum_{\substack{[x] \subseteq K^\times \\ \#[x]>1}} \frac{q^n - 1}{q^{[C(x):\mathbb{F}_q]} - 1}. \tag{$*$}$$

Now, $\#[x] > 1$ becomes $[C(a) : \mathbb{F}_q] < [K : \mathbb{F}_q] = n$.

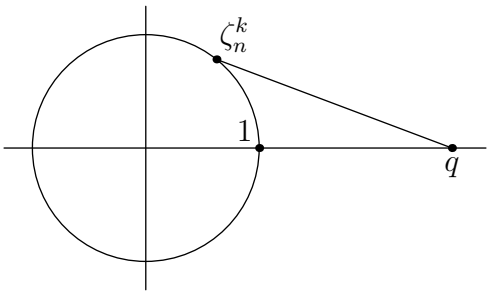
We now apply our second trick, bringing in cyclotomic polynomials. Namely, look at $\Phi_n(q)$. This will certainly divide $q^n - 1$, and it will certainly divide $\frac{q^n-1}{q^k-1}$ for each $k = [C(x) : \mathbb{F}_q] < n$ from the conjugacy classes, so $(*)$ implies that

$$\Phi_n(q) \mid q - 1.$$

But this will force $n = 1$ because $|\Phi_n(q)| > |q-1|$ for $n \neq 1$. Namely,

$$|\Phi_n(q)| = \left| \prod_{k \in (\mathbb{Z}/n\mathbb{Z})^\times} (q - \zeta_n^k) \right| = \prod_{k \in (\mathbb{Z}/n\mathbb{Z})^\times} \left|q - \zeta_n^k\right| \geq |q-1|^{\varphi(n)} \geq |q-1|.$$

Here, the bound $\left|q - \zeta_n^k\right| \geq |q-1|$ comes essentially because $q \geq 1$, coming from the following picture.



Getting this bound rigorously would be annoying, but the main point is that $\left|q - e^{i\theta}\right|^2 = (q - \cos\theta)^2 + (\sin\theta)^2$ achieves its minimum when we simultaneously minimize $(q - \cos\theta)^2$ to $(q-1)^2$ and $(\sin\theta)^2$ to 0.

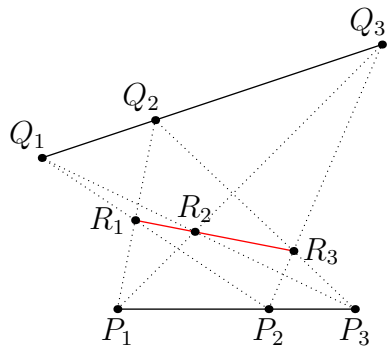
Anyways, if we are to have $|\Phi_n(q)| \leq |q-1|$, then we must be hitting all of our equality cases, so in particular $\zeta_n^k = 1$ for each $k \in (\mathbb{Z}/n\mathbb{Z})^\times$ so that $n = 1$ is forced. Thus, K is one-dimensional over $\mathbb{F}_q$, so $K = \mathbb{F}_q$. This finishes. ∎

As an application of Wedderburn's theorem to projective geometry, we state Pappus's theorem.

Theorem 5.198 (Pappus)**.** Fix $X \coloneqq \mathbb{P}^n(F)$ some n-dimensional projective space over a field F. Then given three collinear points $P_1, P_2, P_3 \in X$ and three more collinear points $Q_1, Q_2, Q_3 \in X$, define the point R_k as being the intersection of $\overline{P_{k+1}Q_{k+2}}$ and $\overline{P_{k+2}Q_{k+1}}$, where the indices are taken $\pmod 3$. Then R_1, R_2, R_3 are collinear.

Proof. As a bad proof, one can give coordinates to everything and solve for R_1, R_2, R_3 explicitly in terms of the coordinates of everything else and then find the line explicitly containing all three points. Roughly speaking, because the statement is true, this approach must work. ■

Here is the mandatory image for this theorem.



To see the importance of Pappus's theorem, we note that the following is true (without proof).

Theorem 5.199. Fix $X \coloneqq \mathbb{P}^n(R)$ some n-dimensional projective space over a division ring R. Then Pappus's theorem holds if and only if R is a field.

Proof. The backwards direction was given above. As for the forwards direction, one can imagine choosing our six points in a particular convenient way and then writing out R_1, R_2, R_3 so that being collinear depends on a particular application of the commutativity of multiplication. I'm not sure how to write this out, but I also don't care very much. ■

The point of bringing in Wedderburn's theorem is that it tells us Pappus's theorem will hold whenever X is finite because this forces R to be finite, hence forcing R to be a field.

Remark 5.200. It feels as if there ought to be a geometric/combinatorial proof that Pappus's theorem holds whenever X is finite, but I think Professor Borcherds said that no such easy proof is known.

As an aside, we note that the set of finite-dimensional division algebras over a field forms a group.

Definition 5.201 (Brauer)**.** Roughly speaking, the *Brauer group* of a field k consists of equivalence classes of finite-dimensional division algebras $[D]$ over k, where the group law is given by

$$[D] * [E] = [F],$$

where $D \otimes E \cong F^{n \times n}$.

This group law is strange, but it works.

5.6.4 Determinants

We are going to quickly talk about the determinant and trace of a linear transformation to later talk about the norm and trace of an element.

Definition 5.202 (Determinants for $\mathbb{R}$)**.** Fix V a finite-dimensional $\mathbb{R}$-vector space with a linear transformation $T : V \to V$. The amount that T "multiplies volumes" is $\det V$, up to sign.

Observe that the map taking T to the amount T scales volumes by induces a map

$$d : \mathrm{Hom}(V, V) \to \mathbb{R}_{\geq 0}.$$

As some examples of what this can do, we see that applying one transformation T_1 and then another T_2 will cause the scaling to compound, so

$$d(T_1 T_2) = d(T_1) d(T_2).$$

Additionally, it is not too hard to show that

$$d\begin{bmatrix} 1 & \bullet & \bullet & \cdots & \bullet & \bullet \\ & 1 & \bullet & \cdots & \bullet & \bullet \\ & & 1 & \cdots & \bullet & \bullet \\ & & & \ddots & \vdots & \vdots \\ & & & & 1 & \bullet \\ & & & & & 1 \end{bmatrix} = 1 \tag{1}$$

by more or less using "base times height"-type arguments. (Here blank spaces are 0, and $\bullet$ are generic elements.) Additionally, we can see that

$$d\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} = \prod_{k=1}^{n} |\lambda_k|$$

because the linear transformation sends the unit cube to a $|\lambda_1| \times \cdots \times |\lambda_n|$ box. However, we know that the actual determinant has

$$\det\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} = \prod_{k=1}^{n} \lambda_k. \tag{2}$$

So the difference between d and $\det$ is a possible sign in our volume. We might want to consider signed volumes or something to fix this, but oftentimes we do actually want to consider volume changes, in which case we do need to keep track of the absolute value.

Remark 5.203. The sign of $\det T$ is whether T preserves "parity." Intuitively, $\det T = 1$ means that the action of T requires some kind of reflection. Rigorously, it is not a bad idea to define "rotations" as linear transformations T with $\det T = 1$, especially in more esoteric spaces.

We would like to extend this definition to general fields. Here are a few ways we can do this.

Definition 5.204 (Determinants, I)**.** Fix k a field. Then we simply define the *determinant* of a matrix in $k^{n \times n}$ by

$$\det\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \coloneqq \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) a_{1,\sigma 1} \cdot a_{2,\sigma 2} \cdot \ldots \cdot a_{n,\sigma n}.$$

Then from this definition we can show that $\det$ is multiplicative and that $\det$ satisfies (1) and (2).

Warning 5.205. The above sum is a really terrible way to evaluate determinants because the number of terms is $n!$. In practice, one should use Gaussian elimination, which requires merely n^3 operations.

Here is another definition, which is more coordinate-free.

Definition 5.206 (Determinants, II)**.** Fix k a field and V an n-dimensional vector space. Then, given $T : V \to V$ a linear transformation, $\det T$ is the amount T multiplies the one-dimensional vector space $\Lambda^n(V)$.

What is $\Lambda^n(V)$? As a start, we define the symmetric algebra $S(V)$.

Definition 5.207 (Symmetric algebra)**.** Fix V a k-vector space. Given a positive integer n, we define the *nth symmetric algebra* $S^n(V)$ as $V^{\otimes n}$ modded out by the relation

$$\cdots \otimes a \otimes b \otimes \cdots = \cdots \otimes b \otimes a \otimes \cdots .$$

Formally, $V^{\otimes n}$ has an S_n-action $S_n \to \operatorname{Aut} V^{\otimes n}$ by permuting the coordinates, so we define $S^n(V)$ as $V^{\otimes n}$ modded out by this action. Then the *symmetric algebra* is defined as

$$S(V) \coloneqq \bigoplus_{n=1}^{\infty} S^n(V).$$

We will not actually check that $S^n(V)$ is a k-algebra, but it is. The main point is that "modding out by the S_n-action" is really modding out by the subspace generated by the elements

$$\left\{ \sigma v - \tau v : \sigma, \tau \in S_n \text{ and } v \in V^{\otimes n} \right\}.$$

Anyways, this definition/intuition can be moved to the slightly more complicated $\Lambda^n(V)$.

Definition 5.208 (Exterior algebra)**.** Fix V a k-vector space. Given a positive integer n, we define the *nth symmetric algebra* $\Lambda^n(V)$ as $V^{\otimes n}$ modded out by the relation

$$\cdots \otimes a \otimes b \otimes \cdots = -(\cdots \otimes b \otimes a \otimes \cdots).$$

Formally, $V^{\otimes n}$ has an S_n-action $S_n \to \operatorname{Aut} V^{\otimes n}$ by

$$\sigma(v_1 \otimes \cdots \otimes v_n) = (\operatorname{sgn} \sigma)(v_{\sigma 1} \otimes \cdots \otimes v_{\sigma n})m$$

so we define $S^n(V)$ as $V^{\otimes n}$ modded out by this action. Then the *exterior algebra* is defined as

$$\Lambda(V) \coloneqq \bigoplus_{n=1}^{\infty} \Lambda^n(V).$$

Elements of $\Lambda^n(V)$ are usually denoted by $v_1 \wedge \cdots v_n$ for $\{v_k\}_{k=1}^n \subseteq V$. Again, we won't actually check that $\Lambda^n(V)$ is a k-algebra, mostly because I don't see a way to do this which avoids pain.

To talk about $\Lambda^n(V)$ more concretely, let's give it a basis. Well, give V a basis $\{b_k\}_{k=1}^n$, and we claim that the set of elements

$$b_{k_1} \wedge \cdots \wedge b_{k_n}$$

such that $k_1 < \cdots < k_n$. To see that these elements span, we see that fully expanding some generic element $v_1 \wedge \cdots \wedge v_n$ along the basis and fully distributing along the tensor product, we see that at least the elements

$$b_{k_1} \wedge \cdots \wedge b_{k_n},$$

with no extra constraint on the $k_\bullet$, will span. However, some permutation $\sigma \in S_n$ will be able to force $\sigma(k_1) \le \cdots \le \sigma(k_n)$, and we see

$$b_{k_1} \wedge \cdots \wedge b_{k_n} = (\operatorname{sgn} \sigma) b_{\sigma k_1} \wedge \cdots \wedge b_{\sigma k_n},$$

so we are allowed to force our basis elements to have $k_1 \le \cdots \le k_n$. Further, we note that, if $k_i = k_j$f for $i \ne j$, then applying the transposition (i, j) will preserve $b_{k_1} \wedge \cdots \wedge b_{k_n}$ while adding a sign, forcing

$$b_{k_1} \wedge \cdots \wedge b_{k_n} = 0.$$

Well, we can throw these elements out too, so we find that we can also force $k_i \neq k_j$ for $i \neq j$.

Remark 5.209 (Nir)**.** I am not recording here a proof that this basis set is actually linearly independent because I don't think one was given in class, and it seems somewhat removed from the class. The sufficiently inclined can read the check in this MathExchange post.

Anyways, we see that if V is also n-dimensional, then $\Lambda^n(V)$ has basis consisting of the single element

$$b_1 \wedge \cdots \wedge b_n,$$

and surely we can track how much a linear transformation scales a one-dimensional subspace. Explicitly, we are defining $\det T$ by

$$(Tb_1 \wedge \cdots \wedge Tb_n) = (\det T)(b_1 \wedge \cdots \wedge b_n).$$

It is somewhat believable that this is indeed equal to the symmetric sum definition: if we have

$$T = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \qquad \text{so that} \qquad Tb_j = \sum_{i=1}^{n} a_{ij} b_i.$$

In particular, fully expanding $Tb_1 \wedge \cdots \wedge Tb_n$ gives terms of the form

$$a_{i_1 1} b_{i_1} \wedge \cdots \wedge a_{i_n n} b_{i_n} = (a_{i_1 1} \cdot \ldots \cdot a_{i_n n})(b_{i_1} \wedge \cdots \wedge b_{i_n}),$$

where the $i_\bullet$ range over any function $\{1, \ldots, n\} \to \{1, \ldots, n\}$. Note that we can still force the $i_\bullet$ to a permutation of $\{1, \ldots, n\}$ because $i_x = i_y$ would cause the entire term to vanish. But then rearranging the $i_\bullet$ into $b_1 \wedge \cdots \wedge b_n$ adds a sign corresponding to the permutation $i_\bullet$, which gives exactly the sum we want.

5.6.5 Trace

Let's quickly review the trace. This is defined as follows.

Definition 5.210 (Trace)**.** Fix k a field. Then we define the trace of a matrix in $k^{n \times n}$ by

$$\operatorname{tr} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} := \sum_{i=1}^{n} a_{ii}.$$

Remark 5.211 (Nir)**.** As an example of some results we get immediately from the definition are that

$$\operatorname{tr}(cM) = c \operatorname{tr} M \qquad \text{and} \qquad \operatorname{tr}(M_1 + M_2) = \operatorname{tr} M_1 + \operatorname{tr} M_2$$

for $c \in k$ and matrices $M, M_1, M_2 \in k^{n \times n}$ (or linear transformations in $\operatorname{End}(V)$ for some k-vector space V). Indeed, the left result comes from writing out cM, and the right result comes from writing out $M_1 + M_2$.

Idea 5.212. The trace is, more or less, the derivative of the determinant.

More precisely, we can show the following.

Exercise 5.213. Fix k a field and $A \in k^{n \times n}$. Then we have that, for $\varepsilon > 0$ small,

$$\det(I + \varepsilon A) = 1 + \varepsilon \operatorname{tr} A + O\left(\varepsilon^2\right).$$

Proof. In practice, we set ε to be a transcendental element so that $\det(I+\varepsilon A)$ is a polynomial in ε. Set

$$A \coloneqq \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

so that $(I+\varepsilon A)_{ij} = 1_{i=j} + \varepsilon a_{ij}$. Namely, we find that

$$\det(I+\varepsilon A) = \sum_{\sigma \in S_n} (\operatorname{sgn}\sigma) \prod_{i=1}^{n} (1_{i=\sigma i} + \varepsilon a_{i,\sigma i}) \in k[\varepsilon].$$

The constant term of this polynomial in ε will come from setting $\varepsilon = 0$, which we can see gives $\det(I+0A) = 1$.

So it remains to study the linear term. Fixing some σ for now, we see that the term

$$\prod_{i=1}^{n} (1_{i=\sigma i} + \varepsilon a_{i,\sigma i})$$

will be able to give us a linear term if we pick $(n-1)$ of the $1_{i=\sigma i}$ terms and one of the $\varepsilon a_{i,\sigma i}$.

But we see now that if $i = \sigma i$ is triggered for $(n-1)$ values of i, then we must have $\sigma = \mathrm{id}$, so this occurs only once, and our linear term is

$$\sum_{i=1}^{n} \varepsilon a_{i,\sigma i} = \varepsilon \operatorname{tr} A,$$

which gives us what we wanted. ∎

We also have the following almost "almost homomorphic" law.

Proposition 5.214. Fix k a field and $A, B \in k^{n\times n}$. Then $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Proof. We do this by direct computation. Namely, set

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix}.$$

Then we see that, for indices x, z we have

$$(AB)_{xz} = \sum_{y=1}^{n} a_{xy} b_{yz} \quad \text{and} \quad (BA)_{xz} = \sum_{y=1}^{n} b_{xy} a_{yz}$$

so that

$$(AB)_{xx} = \sum_{y=1}^{n} a_{xy} b_{yx} \quad \text{and} \quad (BA)_{xx} = \sum_{y=1}^{n} b_{xy} a_{yx}.$$

Namely, we find that

$$\operatorname{tr}(AB) = \sum_{x=1}^{n} (AB)_{xx} = \sum_{x,y=1}^{n} a_{xy} b_{yx} = \sum_{x,y=1}^{n} b_{yx} a_{xy} = \sum_{y=1}^{n} (BA)_{yy} = \operatorname{tr}(BA),$$

which is what we wanted. ∎

Remark 5.215 (Nir)**.** The above two results give a coordinate-free way to define the trace. On one hand, Proposition 5.214 implies that, under a coordinate change matrix S, we have

$$\operatorname{tr}\left(SAS^{-1}\right) = \operatorname{tr}\left(AS^{-1}S\right) = \operatorname{tr}(A),$$

so the trace is invariant under change of basis. To fully remove the coordinates, Exercise 5.213 implies that, given any finite-dimensional k-vector space V, we can define $\operatorname{tr} A$ for $A \in \operatorname{Hom}(V,V)$ as the linear term of the polynomial $\det(I + \varepsilon A) \in k[\varepsilon]$.

Another more coordinate-free view of the trace and determinant is by eigenvalues.

Proposition 5.216. Fix V a finite-dimensional k-vector space and $A \in \operatorname{Hom}(V,V)$ a diagonalizable matrix with eigenvalues $\{\lambda_k\}_{k=1}^n$. Then

$$\det(A) = \prod_{k=1}^n \lambda_k \qquad \text{and} \qquad \operatorname{tr}(A) = \sum_{k=1}^n \lambda_k.$$

Proof. Using an eigenbasis for A, we may write

$$A = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

From here we can directly compute $\operatorname{tr}(A)$ and $\det(A)$ to get the result. ∎

Remark 5.217. One might imagine that we could look at other elementary symmetric polynomials of the eigenvalues, but they are not homomorphisms in general.

Let's give a quick application of the trace.

Theorem 5.218 (Heisenberg commutation relations)**.** Fix V a finite-dimensional k-vector space, where k is a field of characteristic 0. Then if there are linear transformations $A, B \in \operatorname{Hom}(V,V)$ such that

$$AB - BA = I,$$

then $V = \{0\}$.

Proof. Taking the trace of both sides of our equation, we find that

$$0 = \operatorname{tr}(AB) - \operatorname{tr}(BA) = \operatorname{tr} I = \dim V.$$

We have to be somewhat careful because $\dim V \in \mathbb{N}$, but the above equation takes place in k, so $\dim V = 0$ will really only assert that $\operatorname{char} k \mid \dim V$. But in the case where $\operatorname{char} k = 0$, this does force $\dim V = 0$, so V is the zero space. ∎

Example 5.219. The requirement that V be finite-dimensional is necessary: if we take $V = \mathbb{C}[X]$ as a $\mathbb{C}$-vector space, then we can consider the linear transformations $D : f \mapsto \frac{d}{dX} f$ and $\mu_X : f \mapsto Xf$. There are infinite-dimensional vector spaces; e.g., take $V = \mathbb{C}[x]$ and A to be the derivative and B to be multiplication by x. Then, for any $f(X) \in V$, we have

$$(D\mu_X - \mu_X D)(f(X)) = \frac{d}{dX}(Xf(X)) - X\frac{d}{dX}f(X) = f(X) + Xf'(X) - Xf'(X) = f(X),$$

so indeed, $D\mu_X - \mu_X D = \mathrm{id}$.

Example 5.220 (Nir)**.** The requirement that $\operatorname{char} k = 0$ is also necessary. Otherwise, the trace condition merely gives $\operatorname{char} k \mid \dim V$, for which there are examples of A and B. For example, in $k = \mathbb{F}_2$,

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

where we have used the fact that $1 = -1$ in $\mathbb{F}_2$.

Remark 5.221. Apparently p-adic string theory exists.

5.6.6 Norm and Trace

We have the following definition.

Definition 5.222 (Norm and trace)**.** Fix L/K a finite extension of fields. Fix $\alpha \in L$ and view $\mu_\alpha : x \mapsto \alpha x$ as a linear transformation $L \to L$, where we view L as a K-vector space.

(a) The *norm* of α is $\mathrm{N}_K^L(\alpha) \coloneqq \det(x \mapsto \alpha x)$.

(b) The *trace* of α is $\mathrm{T}_K^L(\alpha) \coloneqq \operatorname{tr}(x \mapsto \alpha x)$.

When the extension L/K is clear, we will abbreviate N_K^L to N and T_K^L to T.

Warning 5.223. Professor Borcherds would like you to ignore Lang's definition of the norm and trace because it is somewhat complicated, for example doing different cases based on separability.

And of course, here is a good example to keep track of.

Example 5.224. Fix $\mathbb{C}/\mathbb{R}$ or extension with $\alpha \coloneqq x + yi \in \mathbb{C}$. Using $\{1, i\}$ as our basis of $\mathbb{C}/\mathbb{R}$, we see that our multiplication by α sends $(1, 0) \mapsto (x, y)$ and $(0, 1) \mapsto (-y, x)$, so we get the matrix

$$\begin{bmatrix} x & -y \\ y & x \end{bmatrix}.$$

In particular, we can compute $\mathrm{T}(\alpha) = 2x$ and $\mathrm{N}(\alpha) = x^2 + y^2$. To connect this to more familiar functions, we see $\mathrm{T}(\alpha) = 2\operatorname{Re}\alpha$ and $\mathrm{N}(\alpha) = |\alpha|^2$.

Remark 5.225 (Nir)**.** This doesn't show up anywhere below, so we state it now: N is multiplicative, and T is additive. Indeed, fixing our finite extension L/K and $\alpha_1, \alpha_2 \in L$, we find that

$$\mathrm{T}(\alpha_1+\alpha_2) = \mathrm{tr}(x \mapsto (\alpha_1+\alpha_2)x) = \mathrm{tr}\left((x\mapsto \alpha_1 x) + (x \mapsto \alpha_2 x)\right) = \mathrm{tr}(x\mapsto \alpha_1 x) + \mathrm{tr}(x\mapsto \alpha_2 x),$$

which is $\mathrm{T}(\alpha_1)+\mathrm{T}(\alpha_2)$. Similarly,

$$\mathrm{N}(\alpha_1\alpha_2) = \det(x\mapsto \alpha_1\alpha_2 x) = \det\left((x\mapsto \alpha_1 x)\circ(x\mapsto \alpha_2 x)\right) = \det(x\mapsto \alpha_1 x)\det(x\mapsto \alpha_2 x),$$

which is again the needed $\mathrm{N}(\alpha_1)\mathrm{N}(\alpha_2)$.

Here is a useful way to compute the trace and norm, and it will be a precursor to the rest of our discussion.

Proposition 5.226. Suppose that $L = K(\alpha)$ is a finite extension of degree $n := [L:K]$ such that α has minimal polynomial

$$f(X) = \prod_{k=1}^{n}(X-\alpha_k) \in \overline{K}[X].$$

Then

$$\mathrm{T}(\alpha) = \sum_{k=1}^{n}\alpha_k \qquad \text{and} \qquad \mathrm{N}(\alpha) = \prod_{k=1}^{n}\alpha_k.$$

In other words, $\mathrm{T}(\alpha)$ is the sum of the conjugates of α, and $\mathrm{N}(\alpha)$ is the product of the conjugates.

Proof. The main idea here is that $L = K(\alpha)$ lets us write our linear transformation $x \mapsto \alpha x$ in our power basis

$$\left\{1, \alpha, \alpha^2, \dots, \alpha^{n-1}\right\}.$$

Letting our minimal polynomial $f(X) \in K[X]$ be

$$f(X) = \sum_{k=0}^{n} a_k X^k \in K[X],$$

where we force $a_n = 1$, we see that $x \mapsto \alpha x$ can be defined by sending the basis vectors $\alpha^k \mapsto \alpha^{k+1}$ for $0 \le k \le n-2$ and

$$\alpha^{n-1} \mapsto \alpha^n = \sum_{k=0}^{n-1}(-a_k)\alpha^k.$$

Namely, $x \mapsto \alpha x$ looks like the matrix

$$\begin{bmatrix} 0 & & & & -a_0 \\ 1 & 0 & & & -a_1 \\ & 1 & 0 & & -a_2 \\ & & \ddots & \ddots & \vdots \\ & & & 1 & -a_{n-1} \end{bmatrix}.$$

We immediately see that the trace is $-a_{n-1}$, which is

$$\mathrm{T}(\alpha) = -a_{n-1} = \sum_{k=1}^{n}\alpha_k$$

by Vieta's formulae. Similarly, for the norm, we see that we can bubble-sort the top row of this matrix to the bottom with $(n-1)$ swaps and thus introducing $(n-1)$ signs, meaning we want the determinant of the

matrix

$$(-1)^{n-1}\begin{bmatrix} 1 & & & & -a_1 \\ & 1 & & & -a_2 \\ & & \ddots & & \vdots \\ & & & 1 & -a_{n-1} \\ & & & & -a_0. \end{bmatrix}.$$

But now this matrix is upper-triangular and hence the determinant we want is $(-1)^{n-1}(-a_0) = (-1)^n a_n$, which is

$$\mathrm{N}(\alpha) = (-1)^n a_n = \prod_{k=1}^{n} \alpha_k,$$

again using Vieta's formulae. ■

Remark 5.227 (Nir). A more coordinate-free to get this result is to actually go compute the characteristic polynomial of $\mu_\alpha : x \mapsto \alpha x$ and find that it is actually $f(X)$. We outline this. Note $f(X)$ is irreducible over $K[X]$ and has the right degree, so it suffices to check $f(\mu_\alpha) = 0$ by the Cayley–Hamilton theorem. But $\mu_\bullet : K \to \mathrm{End}_K(K(\alpha))$ is actually a ring homomorphism, so we see $f(\mu_\alpha) = \mu_{f(\alpha)} = \mu_0 = 0$.

5.6.7 Norms and Traces in Towers

More generally, we have the following.

Proposition 5.228. Fix L/K a finite extension with $\alpha \in L$. Then fix α with minimal polynomial

$$f(X) = \prod_{k=1}^{n} (X - \alpha_k) \in \overline{K}[X],$$

where $n = [K(\alpha) : K]$. Then

$$\mathrm{T}_K^L(\alpha) = [L : K(\alpha)] \sum_{k=1}^{n} \alpha_k \qquad \text{and} \qquad \mathrm{N}_K^L(\alpha) = \left(\prod_{k=1}^{n} \alpha_k \right)^{[L:K(\alpha)]}.$$

Proof. Once we note that we have the chain $K \subseteq K(\alpha) \subseteq L$, it suffices to note the previous proposition gives

$$\mathrm{T}_K^{K(\alpha)}(\alpha) = \sum_{k=1}^{n} \alpha_k \qquad \text{and} \qquad \mathrm{N}_K^{K(\alpha)}(\alpha) = \prod_{k=1}^{n} \alpha_k$$

and then finish by applying the tower law in the next proposition. ■

Remark 5.229 (Nir). We have had to be quite careful above, both in statement and proof because we are not assuming that the $\alpha_\bullet$ are distinct, as might not be the case in inseparable extensions.

Here is the aforementioned tower law.

Proposition 5.230. Give a chain of fields $K \subseteq L \subseteq M$ and $\alpha \in L$, we have the tower law

$$\mathrm{N}_K^M(\alpha) = \left(\mathrm{N}_K^L \alpha\right)^{[M:L]} \qquad \text{and} \qquad \mathrm{T}_K^M(\alpha) = [M : L]\, \mathrm{T}_K^L(\alpha).$$

Proof. We essentially work with the following tower of fields.

$$\begin{array}{c} M \\ | \\ L \\ | \\ K \end{array}$$

Essentially what is going on, now, is that $\mu_\alpha : x \mapsto \alpha x$ is rather controlled on L because μ_α is a multiplication by a constant in the ground field. To be explicit, fix $\{v_\ell\}_{\ell=0}^{m-1}$ a basis of M as a L-vector space. Then we can write

$$M = \bigoplus_{\ell=0}^{m-1} Lv_\ell.$$

Now, the linear transformation $\mu_\alpha : x \mapsto \alpha x$ preserves each of these subspaces Lv_ℓ because $\alpha \in L$, so we can decompose $\mu_\alpha : x \mapsto \alpha x$ as a direct sum of its action on each of these subspaces Lv_ℓ.

To finish, we note that $Lv_\ell \cong L$ by division by v_ℓ, and the action of μ_α on Lv_ℓ commutes with this isomorphism; in other words, the following diagram commutes because multiplication commutes.[9]

$$\begin{array}{ccc} Lv_\ell & \cong & L \\ {\scriptstyle \mu_\alpha|_{Lv_\ell}}\downarrow & & \downarrow{\scriptstyle \mu_\alpha|_L} \\ Lv_\ell & \cong & L \end{array}$$

So we can compute the determinant and trace of μ_α as it behaves on L instead of Lv_ℓ, which we have already studied in the previous proposition. Indeed, we find

$$\mathrm{T}_K^M(\alpha) = \operatorname{tr} \mu_\alpha = \sum_{\ell=0}^{m-1} \operatorname{tr}(\mu_\alpha|_{Lv_\ell}) = \sum_{\ell=0}^{m-1} \operatorname{tr}(\mu_\alpha|_L) = [M:L]\,\mathrm{T}_K^L(\alpha),$$

where we used the previous proposition in the last inequality. Similarly,

$$\mathrm{N}_K^M(\alpha) = \det \mu_\alpha = \prod_{\ell=0}^{m-1} \det(\mu_\alpha|_{Lv_\ell}) = \prod_{\ell=0}^{m-1} \det(\mu_\alpha|_L) = \left(\mathrm{N}_K^L\,\alpha\right)^{[M:L]},$$

which is what we wanted. ∎

This more or less lets us define an "absolute" trace and norm.

Definition 5.231 (Reduced trace and norm)**.** Fix L/K a finite extension with $\alpha \in L$. Then we define the *reduced trace* as $\frac{1}{[L:K]}\,\mathrm{T}_K^L(\alpha)$ and the *reduced norm* is $\mathrm{N}_K^L(\alpha)^{1/[L:K]}$.

Both of these definitions have problems: if K has characteristic dividing $[L : K]$, then the reduced trace doesn't exist. If K does not have enough roots of unity, we might not be able to take the $1/[L : K]$ powers.

But given that the reduced trace and norm always actually exist and are well-defined, we can show that there are independent of the field L. Indeed, suppose that $\alpha \in L_1, L_2$ where L_1, L_2 are finite extensions of K. Then we know that

$$\frac{1}{[L_1:K]}\,\mathrm{T}_K^{L_1}(\alpha) = \frac{1}{[L_1:K][L_2L_1:L_1]}\,\mathrm{T}_K^{L_1L_2}(\alpha) = \frac{1}{[L_1L_2:K]}\,\mathrm{T}_K^{L_1L_2}(\alpha)$$

[9] More rigorously, what is going on is that we are expanding our μ_α along the basis $\{v_\ell w_\bullet\}$, where $\{w_\bullet\}$ is some fixed basis of L, and we find that the matrix is the same as if we had just acted on $\{w_\bullet\}$ to begin with.

where we have applied the tower law. But now $\frac{1}{[L_1L_2:K]}\,\mathrm{T}_K^{L_1L_2}(\alpha)$ is symmetric in L_1 and L_2, so the above must also be equal to $\frac{1}{[L_2:K]}\,\mathrm{T}_K^{L_2}(\alpha)$, as needed.

Similarly,[10]

$$(\mathrm{N}_K^{L_1}(\alpha))^{1/[L_1:K]} = \mathrm{N}_K^{L_1L_2}(\alpha)^{1/([L_1:K][L_1L_2:L_1])} = \mathrm{N}_K^{L_1L_2}(\alpha)^{1/[L_1L_2:K]}$$

where we again have applied the tower law. But now $\mathrm{N}_K^{L_1L_2}(\alpha)^{1/[L_1L_2:K]}$ is symmetric in L_1 and L_2, so the above must also be equal to $\mathrm{N}_K^{L_2}(\alpha)^{1/[L_2:K]}$, as needed.

Anyways, for Galois extensions, much of our story with the minimal polynomial collapses nicely.

Proposition 5.232. Fix L/K a separable extension and let G be the set of embeddings $L \hookrightarrow \overline{K}$. Then, for $\alpha \in L$, we have that

$$\mathrm{T}(\alpha) = \sum_{\sigma\in G} \sigma(\alpha) \qquad \text{and} \qquad \mathrm{N}(\alpha) = \prod_{\sigma\in G} \sigma(\alpha).$$

In particular, when L/K is Galois, $G = \mathrm{Gal}(L/K)$.

Proof. Let $f(X)$ be the monic minimal polynomial $\alpha \in K[X]$ so that

$$f(X) = \prod_{k=1}^{n} (X - \alpha_k)$$

for some elements $\alpha_1, \dots, \alpha_n \in \overline{K}$ and $n = [K(\alpha) : K]$. But L/K is separable, so these elements are distinct, and L/K is normal, so all of these elements live in L because α is one root of $f(X)$.

The main point is to understand the multiset $\{\sigma\alpha\}_{\sigma\in G}$ in order to compute the sum and product in the statement. To start, we notice that the embedding $K(\alpha) \hookrightarrow \overline{K}$ by

$$K(\alpha) \cong \frac{K[X]}{f(X)} \cong K(\alpha_\bullet) \subseteq \overline{K}$$

can be extended to a full embedding in G with the property that $\alpha \mapsto \alpha_\bullet$. In particular, for each $\sigma \in G$, the fact σ is an embedding forces $\sigma\alpha \in \{\alpha_k\}_{k=1}^n$, and we can indeed hit every $\alpha_\bullet$ as described above, so the orbit of α under G is the entire $\{\alpha_k\}_{k=1}^n$.

We now directly focus on the multiset

$$\{\sigma\alpha\}_{\sigma\in G}.$$

Using the bijection $G/\operatorname{Stab}(\alpha)$ to the set $\{\sigma\alpha\}_{\sigma\in G}$, we see that each $\sigma\alpha$ in the above multiset will be hit $\#\operatorname{Stab}(\alpha)$ times, which by the Orbit-stabilizer theorem is $\#G/n$. Namely, when we write

$$\sum_{\sigma\in G} \sigma\alpha,$$

this sum is hitting each $\alpha_\bullet$ exactly $\#G/n = [L : K]/[K(\alpha) : K] = [L : K(\alpha)]$ times (note $\#G = [L : K]$ because L/K is separable!), so it is

$$\sum_{\sigma\in G} \sigma\alpha = [L : K(\alpha)] \sum_{k=1}^{n} \alpha_k = \mathrm{T}_K^L(\alpha).$$

Similarly, we find that

$$\prod_{\sigma\in G} \sigma\alpha = \left(\prod_{k=1}^{n} \alpha_k\right)^{[L:K(\alpha)]} = \mathrm{N}_K^L(\alpha),$$

which is what we wanted. ∎

[10] The roots do not necessarily make sense, but as long as they are defined in some way which is compatible with all of the other roots, we should be safe. I am not going to write this out.

5.6.8 Algebraic Integers

Here is something algebraic number theorists care about.

Definition 5.233 (Algebraic integers)**.** Given a finite extension $K/\mathbb{Q}$, the *algebraic integers* $\mathcal{O}_K \subseteq K$ are those which are the roots of some monic polynomial.

We will take on faith that the sum and product of two algebraic integers is another algebraic integer; showing this is approximately the same as showing that the sum and product of two algebraic numbers is an algebraic number while keeping track of the integral condition. But this is surprisingly technical and somewhat removed from the course, so we will not say more.

Anyways, the point of is that the set of algebraic integers in K forms a ring, once we add in the fact that 0 and 1 are algebraic integers. Here are some examples.

Example 5.234. The algebraic integers of $\mathbb{Q}(\sqrt{-3})$ is not $\mathbb{Z}[\sqrt{-3}]$. The main point is that

$$\frac{1+\sqrt{-3}}{2}$$

is an algebraic integer because it is the root of the polynomial $x^2 + x + 1 = 0$.

Example 5.235. The algebraic integers of $\mathbb{Q}(\sqrt{-2})$ are $\mathbb{Z}[\sqrt{-2}]$. We will show this shortly.

The following example justifies the name "integer."

Exercise 5.236. The set of algebraic integers in $\mathbb{Q}$ is exactly $\mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$.

Proof. Certainly each $n \in \mathbb{Z}$ is an algebraic integer because n is the root of the monic polynomial $X - n \in \mathbb{Z}[X]$. So in one direction, $\mathbb{Z} \subseteq \mathcal{O}_{\mathbb{Q}}$.

Conversely, we show $\frac{p}{q} \in \mathbb{Q}$ with $\gcd(p,q) = 1$ is an algebraic integer forces $q = \pm 1$ and thus $\frac{p}{q} \in \mathbb{Z}$. Indeed, if p/q is an algebraic integer, then find our monic polynomial

$$f(X) \coloneqq X^n + \sum_{k=0}^{n-1} a_k X^k \in \mathbb{Z}[X]$$

such that $f(p/q) = 0$ so that

$$0 = q^n f(p/q) = p^n + \sum_{k=0}^{n-1} a_k p^k q^{n-k}.$$

Now, we see q divides the left-hand side as well as big sum, so q divides p^n as well. But then q divides $\gcd(p^n, q) = 1$, forcing $q = \pm 1$. ∎

As promised, let's compute the algebraic integers of a quadratic extension.

Exercise 5.237. Fix m a squarefree integer not equal to 1, and set $K \coloneqq \mathbb{Q}(\sqrt{m})$. Then

$$\mathcal{O}_K = \begin{cases} \mathbb{Z}[\sqrt{m}] & m \equiv 2,3 \pmod 4, \\ \mathbb{Z}\left[\frac{1+\sqrt{m}}{2}\right] & m \equiv 1 \pmod 4. \end{cases}$$

Note $m \equiv 0 \pmod 4$ never occurs because m is squarefree.

Proof. Note that all integers will be algebraic integers because $n \in \mathbb{Z}$ is the root of the monic polynomial $X - n \in \mathbb{Z}[X]$. Additionally, we see that $\sqrt{m}$ will be a root of $X^2 - m \in \mathbb{Z}[X]$, so $\sqrt{m} \in \mathcal{O}_K$. When $m \equiv 1 \pmod 4$, we also have that $\frac{1+\sqrt{m}}{2}$ is a root of

$$X^2 - X + \frac{1-m}{4} \in \mathbb{Z}[X],$$

so $\frac{1+\sqrt{m}}{2}$ is also an algebraic integer. All this is to say that $\mathbb{Z}[\sqrt{m}] \subseteq \mathcal{O}_K$ always, and when $m \equiv 1 \pmod 4$, we also have $\mathbb{Z}\left[\frac{1+\sqrt{m}}{2}\right] \subseteq \mathcal{O}_K$.

It remains to show the equalities. Suppose $a+b\sqrt{m} \in \mathbb{Q}(\sqrt{m})$ is an algebraic integer. Well, note that the Galois conjugate $a - b\sqrt{m}$ will be a root of the same polynomial as $a + b\sqrt{m}$, so in particular it will be monic with integer coefficients, so $a - b\sqrt{m}$ will be an algebraic integer.

So the key trick is that we know

$$\mathrm{T}(a + b\sqrt{m}) = 2a \qquad \text{and} \qquad \mathrm{N}(a + b\sqrt{m}) = a^2 - bm^2.$$

will also be algebraic integers. But then $2a \in \mathbb{Q}$ is an algebraic integer, so $2a, a^2 - mb^2 \in \mathbb{Z}$ is forced. Further, we notice that

$$4\left(a^2 - mb^2\right) - (2a)^2 = -m(2b)^2$$

must also be an integer, so the same logic in the previous case forces $2b \in \mathbb{Z}$. So, setting $a = 2c$ and $b = 2d$, we see that $\frac{c+d\sqrt{m}}{2}$ is our algebraic integer, and checking its norm now, we see that

$$\frac{c^2 - md^2}{4} \in \mathbb{Z},$$

so $c^2 \equiv md^2 \pmod 4$. We now have two cases.

- If $m \equiv 1 \pmod 4$, then we notice that $c^2 \equiv d^2 \pmod 4$. Reducing to $\pmod 2$ somewhat brazenly, we see that
$$c \equiv c^2 \equiv d^2 \equiv d \pmod 2,$$
so $c \equiv d \pmod 2$ is forced. However, this means that we can write
$$\frac{c+d\sqrt{m}}{2} = \frac{c-d}{2} + d \cdot \frac{1+\sqrt{m}}{2} \in \mathbb{Z}\left[\frac{1+\sqrt{m}}{2}\right],$$
which is what we wanted.
- Otherwise, $m \equiv 2, 3 \pmod 4$, and here $c^2 \equiv md^2 \pmod 4$ forces $c, d \equiv 0 \pmod 2$. Explicitly, if d is even, then $c^2 \equiv 0 \pmod 2$ forces c even. And if d is odd, then we have $c^2 \equiv m \pmod 4$, which has no solutions.
Thus,
$$\frac{c+d\sqrt{m}}{2} = \frac{c}{2} + \frac{d}{2}\sqrt{m} \in \mathbb{Z}[\sqrt{m}],$$
which is again what we wanted. ∎

5.6.9 Trace Form

To close off, we note that the trace of an extension L/K induces a symmetric bilinear form

$$\langle \alpha, \beta \rangle \coloneqq \mathrm{T}(\alpha\beta).$$

To be explicit, we see that $\langle \alpha, \beta \rangle = \langle \beta, \alpha \rangle$ because multiplication commutes; the additivity of the trace gives $\langle \alpha_1 + \alpha_2, \beta \rangle = \langle \alpha_1, \beta \rangle + \langle \alpha_2, \beta \rangle$; and the fact

$$\mathrm{T}(c\alpha) = \mathrm{tr}\left(c \cdot (x \mapsto \alpha x\right) = c\,\mathrm{tr}(x \mapsto \alpha x) = c\,\mathrm{T}(\alpha),$$

for any $c \in K$, gives $\langle c\alpha, \beta \rangle = c\langle \alpha, \beta \rangle$.

Here is our favorite example.

Example 5.238. Fix $\mathbb{C}/\mathbb{R}$ to be our finite extension. Then

$$\langle a+bi, c+di\rangle = \operatorname{tr}((a+bi)(c+di)) = 2(ab-cd),$$

where the minus sign is somewhat important.

We hope that this bilinear form is non-degenerate. We will talk about this next lecture.

5.7 December 2

Overhead, without any fuss, the stars were going out.

5.7.1 Trace Form

Last lecture we had brought up the symmetric bilinear form

$$\langle \alpha, \beta\rangle \coloneqq \mathrm{T}(\alpha\beta)$$

for any finite field extension L/K. In particular, last time we quickly checked that $\langle\cdot,\cdot\rangle$ is in fact a symmetric bilinear form.

While we're here, we bring up the following warning.

Warning 5.239. If L/K has degree 2, then we have the two natural quadratic forms $\alpha \mapsto \mathrm{N}(\alpha)$ and $\alpha \mapsto \mathrm{T}\left(\alpha^2\right)$.

We have not defined what a quadratic form,[11] but for those who do know, it might be somewhat concerning that there need not be an obviously "best" quadratic form for a field extension.

Anyways, we are interested in when our trace form is non-degenerate.

Definition 5.240 (Nondegenerate)**.** A bilinear form $\langle\cdot,\cdot\rangle$ is *nondegenerate* if and only if $\langle x, y\rangle = 0$ for all y implies $x = 0$. In other words, for each $x \neq 0$, there exists $y \neq 0$ such that $\langle x, y\rangle \neq 0$.

We note that our trace form $\langle\cdot,\cdot\rangle$ will be nondegenerate if and only if there exists $\alpha \in L$ such that $\mathrm{T}(\alpha) \neq 0$. Certainly if such an $\langle\cdot,\cdot\rangle$ is nondegenerate, then $1 \neq 0$ promises there is some α such that

$$\mathrm{T}(\alpha) = \langle 1, \alpha\rangle \neq 0.$$

Conversely, if there is some α with $\mathrm{T}(\alpha) \neq 0$, then for any $x \neq 0$, we see that

$$\langle x, \alpha/x\rangle = \mathrm{T}(x\cdot\alpha/x) = \mathrm{T}(\alpha) \neq 0.$$

So this verifies that $\langle\cdot,\cdot\rangle$ is nondegenerate.

It is tempting to believe that $\langle\cdot,\cdot\rangle$ is always nondegenerate because

$$\mathrm{T}_K^L(1) = [L:K]\cdot 1 = [L:K]$$

by Proposition 5.228 because $1 \in K$. However, T_K^L outputs into K, so we still need to check if $[L:K]$ is nonzero in K, so 1 will work to prove that $\langle\cdot,\cdot\rangle$ is nondegenerate only when $\operatorname{char} K \nmid [L:K]$.

Regardless, it looks like $\langle\cdot,\cdot\rangle$ is nondegenerate most of the time. However, it is not always.

[11] One possible definition is that a function $q : V \to k$ is a quadratic form if and only if $q(cv) = c^2 q(v)$ for $c \in k$ and $v \in V$ and $\langle v, w\rangle \coloneqq q(v+w) - q(v) - q(w)$ is a symmetric bilinear form.

Example 5.241. Fix p a prime and $L = \mathbb{F}_p(t^p)$ with $K = \mathbb{F}_p(t)$ so that L/K is not a separable extension. Then the trace is 0 on L. The reasoning for this example will generalize to arbitrary inseparable extensions, so we will just show the general case below.

As alluded to above, we do have the following criteria for being nondegenerate.

Theorem 5.242. Fix L/K a finite field extension. Then there exists $\alpha \in L$ such that $\mathrm{T}(\alpha) \neq 0$ if and only if L/K is separable. In particular, $\langle \cdot, \cdot \rangle$ is nondegenerate if and only if L/K is separable.

We divide this proof into two pieces.

Proof of the backwards direction in Theorem 5.242. Fix L/K separable. Most of the work in the proof will be done assuming that L/K is Galois, and we will go back at the end and do this for general separable extensions. So for now fix $G \coloneqq \mathrm{Gal}(L/K)$.

Namely, we want to find some $\alpha \in L$ such that $\mathrm{T}(\alpha) \neq 0$. But by Proposition 5.232, we know that

$$\mathrm{T}(\alpha) = \sum_{\sigma \in G} \sigma\alpha.$$

Having $\mathrm{T}(\alpha) = 0$ always would imply that $\sum_\sigma \sigma = 0$, which would violate Lemma 5.151 because this provides a nontrivial relation among distinct automorphisms, finishing immediately. However, while we're here, we note that we can generalize Lemma 5.151 in the following way.

Lemma 5.243 (Artin)**.** Fix L a field and M a monoid. Further, pick up some finite set of homomorphisms $S \subseteq \mathrm{Hom}(M, L^\times)$. Then the set S is L-linearly independent.

Proof. We essentially redo the proof from Lemma 5.151. Suppose for the sake of contradiction that there is a nontrivial relation involving the elements of S. This means we can find a nontrivial relation

$$\sum_{k=1}^{m} a_k \sigma_k = 0, \tag{$*$}$$

where m is chosen to be minimal and $\{\sigma_k\}_{k=1}^m \subseteq S$. For example, this implies that $a_k \neq 0$ for each k because then we could remove the term $a_k\sigma_k$ to get a smaller relation.

By dividing out $(*)$ by a_1, we may assume that $a_1 = 1$. Now, $\sigma_1 \neq \sigma_2$, so there exists some h such that $\sigma_1(h) \neq \sigma_2(h)$, so plugging in gh into $(*)$ gives

$$\sigma_1(h) \cdot \sigma_1(g) + \sum_{k=2}^{m} a_k \sigma_k(h) \cdot \sigma_k(g) = 0$$

for each $g \in M$. But multiplying $(*)$ through by $\sigma_1(h)$ gives

$$\sigma_1(h) \cdot \sigma_1(g) + \sum_{k=2}^{m} a_k \sigma_1(h) \cdot \sigma_k(g) = 0$$

for each $g \in M$. Subtracting our two equations, we see that

$$\sum_{k=2}^{m} a_k(\sigma_k(h) - \sigma_1(h))\sigma_k = 0,$$

which is a nontrivial relation because $a_2(\sigma_2(h) - \sigma_1(h)) \neq 0$. But this is a strictly smaller nontrivial relation than our supposed smallest one, so we have our contradiction. ∎

We now apply the lemma to our case. Here we find that the automorphisms of G are homomorphisms $L^\times \to L^\times$, so they are L-linearly independent, so we must have

$$\mathrm{T} = \sum_{\sigma \in G} \sigma \neq 0.$$

In particular, there must exist some α such that $\mathrm{T}(\alpha) \neq 0$, which is what we wanted.

We now turn to the general case. To reduce to the Galois case, we have the following lemma.

Lemma 5.244. Fix L/K a separable field extension. Then there exists a field $M \supseteq L$ such that M/K is a Galois extension. If L/K is finite, then we may assume M/K is finite.

Proof. Fix L/K generated by some separable elements $S \subseteq L$. Then, for each $\alpha \in S$, define $f_\alpha \in K[X]$ to be the monic irreducible polynomial for α. Then we define M to be the splitting field of all these polynomials

$$\{f_\alpha : \alpha \in S\}.$$

We see that M is the splitting field of some set of polynomials, so M/K is normal. Additionally, M will be generated by the roots of these f_α, which will all be separable elements because f_α is separable. So M/K is a separable extension, so M/K is Galois.

Now, when L/K is a finite extension, we may assume that S is finite (for example, take a basis for L as a K-vector space), so there are only finitely many polynomials, so M/K will be finite because each polynomial can only add finitely many degrees. ∎

So we may extend L/K to a Galois extension M/K, reducing to the Galois case. Namely, our work above promises some $\alpha \in M$ such that $\mathrm{T}_K^M(\alpha) \neq 0$. To finish, we pick up the following tower law, which generalizes Proposition 5.230.

Lemma 5.245 (Trace tower law)**.** Fix $K \subseteq L \subseteq M$ a chain of finite separable field extensions. Then $\mathrm{T}_K^M = \mathrm{T}_K^L \circ \mathrm{T}_L^M$.

Proof. Extend M/K to a finite Galois extension N/K. Further, let $\{\sigma_k\}_{k=1}^m$ be the embeddings $L \hookrightarrow \overline{K}$ fixing L and $\{\tau_\ell\}_{\ell=1}^n$ be the embeddings $M \hookrightarrow \overline{L}$ fixing L. Note $m = [L : K]$ and $n = [M : L]$.

Note that we can extend each embedding $\sigma_\bullet : L \hookrightarrow \overline{K}$ to some fixed embedding $\sigma_\bullet : N \hookrightarrow \overline{K}$, but now because N/K is normal, we see that $\sigma_\bullet \in \mathrm{Gal}(N/K)$. In the same way we can extend each embedding $\tau_\bullet : M \hookrightarrow \overline{L}$ to an automorphism $\tau_\bullet \in \mathrm{Gal}(N/L)$.

Now, the main technical claim is that

$$(k, \ell) \mapsto \sigma_k \tau_\ell|_M$$

is an injection; here the restriction to M makes $\sigma_k\tau_\ell$ an embedding $M \hookrightarrow \overline{K}$ fixing K. (The composition here is legal because we lifted these to elements of $\mathrm{Gal}(N/K)$, which is the only reason we need N at all.) Indeed, suppose that $\sigma_{k_1}\tau_{\ell_1}|_M = \sigma_{k_2}\tau_{\ell_2}|_M$. Then we see that

$$\sigma_{k_2}^{-1}\sigma_{k_1}|_M = \tau_{\ell_2}\tau_{\ell_1}^{-1}|_M.$$

Now, the right-hand side fixes L, so $\sigma_{k_2}^{-1}\sigma_{k_1}|_M$ will also have to fix L. So it follows that

$$\sigma_{k_1}|_L = \sigma_{k_2}|_L,$$

so because we lifted the $\sigma_\bullet$ from embeddings $L \hookrightarrow \overline{K}$, it follows that $\sigma_{k_1} = \sigma_{k_2}$. From this we get that $\tau_{\ell_1} = \tau_{\ell_2}$ as well because we lifted these from embeddings $M \hookrightarrow \overline{L}$.

So because $(k, \ell) \mapsto \sigma_k\tau_\ell$ is an injection, the fact that there are $[L : K]$ of the $\sigma_\bullet$ and $[M : L]$ of the $\tau_\bullet$ implies that we have found $[M : K]$ distinct embeddings $M \hookrightarrow \overline{K}$ fixing K, so this must be all of them.

Thus, we find, for any $\alpha \in M$,

$$\mathrm{T}_K^M(\alpha) = \sum_{k=1}^m \sum_{\ell=1}^n (\sigma_k\tau_\ell)(\alpha).$$

But this summation is also equal to

$$\mathrm{T}_K^L\left(\mathrm{T}_L^M(\alpha)\right) = \sum_{k=1}^{m} \sigma\left(\sum_{\ell=1}^{n} \tau_\ell(\alpha)\right)$$

after distributing. So indeed, $T_K^M = \mathrm{T}_K^L \circ \mathrm{T}_L^M$. ∎

Remark 5.246 (Nir)**.** The tower law holds for the norm by using the same argument but replacing the sums at the end with products. There is also an analogous statement for inseparable extensions, but I would rather avoid inseparable extensions as much as possible.

The point of the tower law is that we see $\mathrm{T}_L^M(\alpha) \in L$ satisfies

$$\mathrm{T}_K^L\left(\mathrm{T}_L^M(\alpha)\right) = \mathrm{T}_K^M(\alpha) \neq 0,$$

so we have indeed found an element of L with nonzero trace. This finishes the proof of the backwards direction. ∎

Proof of the forwards direction in Theorem 5.242. We show the other direction by contraposition: take L/K inseparable, and we show that $\mathrm{T}(\alpha) = 0$ for each $\alpha \in L$. We need to know something about inseparable extensions, so we show the following.

Lemma 5.247. Fix K a field and $f(X) \in K[X]$ an inseparable, irreducible polynomial. Then there exists $g(X) \in \overline{K}[X]$ such that $f(X) = g(X)^p$, where $p := \operatorname{char} K$.

We remark that we do have $p > 0$ above because all extensions are separable in characteristic 0 by the derivative trick.

Proof. We see that $f(X)$ is irreducible and has a double root at some $\alpha \in \overline{K}$. This means that $f(\alpha) = 0$ and $f'(\alpha) = 0$, so

$$(X - \alpha) \mid \gcd(f(X), f'(X)).$$

If $f'(X) \neq 0$, then we see that $1 = \deg(X - \alpha) \leq \deg \gcd(f(X), f'(X)) < \deg f(X)$ while $\gcd(f(X), f'(X)) \mid f(X)$, which violates f being irreducible. So we must have $f'(X) = 0$.

But this implies that each nonzero monomial $a_k X^k$ of $f(X)$ must have $k a_k X^k = 0$ in K, so $k a_k$ is 0 in K, so k is 0 in K, so $\operatorname{char} K \mid k$. In other words, the only nonzero monomials of $f(X)$ will have degree divisible by p, so we may write

$$f(X) = \sum_{k=0}^{n} a_k X^{kp}$$

for some coefficients $a_k \in K$. Finding some root $a_k^{1/p} \in \overline{K}$, we see that

$$f(X) = \sum_{k=0}^{n} \left(a_k^{1/p} X^k\right)^p = \left(\sum_{k=0}^{n} a_k^{1/p} X^k\right)^p,$$

which gives what we wanted. ∎

We now attack the result directly. Fix some $\alpha \in L$. Then we are interested in studying

$$\mathrm{T}_K^L(\alpha) = [L : K(\alpha)] \cdot \mathrm{T}_K^{K(\alpha)}(\alpha).$$

Namely, we are working with the following tower of fields.

$$\begin{array}{c} L \\ | \\ K(\alpha) \\ | \\ K \end{array}$$

Note that if both $L/K(\alpha)$ and $K(\alpha)/K$ are separable extensions, then it follows that L/K is separable,[12] which cannot be. So we can do casework on which extension in inseparable.

- If $L/K(\alpha)$ is inseparable, then fix β some inseparable element with $f(X) \in K(\alpha)[X]$ its minimal polynomial. We note that $[K(\alpha)(\beta) : K(\alpha)] = \deg f$, but by Lemma 5.247, we see $p \mid \deg f$, so p divides $[K(\alpha)(\beta) : K(\alpha)]$ and hence $[L : K(\alpha)]$. So it follows
$$[L : K(\alpha)] \cdot \mathrm{T}_K^{K(\alpha)}(\alpha) = 0.$$
- If $K(\alpha)/K$ is inseparable, then we note α must be inseparable. So fix $f(X)$ its minimal polynomial and actually find $g(X) \in \overline{K}[X]$ such that $f = g^p$ by Lemma 5.247. Factoring g in $\overline{K}[X]$, we can write
$$g(X) = \prod_{k=1}^{n} (X - \alpha_k)$$
for some elements $\alpha_\bullet \in \overline{K}$, so it follows that
$$f(X) = \prod_{k=1}^{n} (X - \alpha_k)^p.$$
Thus, Proposition 5.226 implies
$$\mathrm{T}_K^{K(\alpha)}(\alpha) = \sum_{k=1}^{n} p\alpha_k = 0,$$
so we still have $\mathrm{T}_K^{L}(\alpha) = 0$.

Combining the above two cases finishes. ∎

5.7.2 Discriminant: Theory

The trace form gives rise the following invariant.

Definition 5.248 (Discriminant)**.** The *discriminant* of a bilinear form $\langle \cdot, \cdot \rangle$ on a finite-dimensional k-vector space V is defined by taking a basis $\{v_k\}_{k=1}^n$ for V and computing
$$\det \begin{bmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_n \rangle \\ \vdots & \ddots & \vdots \\ \langle v_n, v_1 \rangle & \cdots & \langle v_n, v_n \rangle \end{bmatrix}$$

Of course, changing the basis by some change-of-basis matrix A will change the discriminant, but only by a controlled factor of $(\det A)^2$. Rigorously, we have the following.

[12] The main point is that separability is equivalent to any embedding $K \hookrightarrow \overline{K}$ having $[L : K]$ extensions to $L \hookrightarrow \overline{K}$. So L/K and M/L separable lets us extend each of the $[L : K]$ embeddings $L \hookrightarrow \overline{K} = \overline{L}$ to $[M : L][L : K] = [M : K]$ embeddings $M \hookrightarrow \overline{L} = \overline{K}$. So M/K is separable.

Lemma 5.249. The discriminant of a bilinear form $\langle \cdot, \cdot \rangle$ on a finite-dimensional k-vector space V is a well-defined element of $k/k^{\times 2}$.

Note that we are allowing the discriminant to be zero, but zero belongs to its own coset.

Proof. Fix two bases $\{v_i\}_{i=1}^n$ and $\{w_j\}_{j=1}^n$. We need to show that

$$\det \begin{bmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_n \rangle \\ \vdots & \ddots & \vdots \\ \langle v_n, v_1 \rangle & \cdots & \langle v_n, v_n \rangle \end{bmatrix} \quad \text{and} \quad \det \begin{bmatrix} \langle w_1, w_1 \rangle & \cdots & \langle w_1, w_n \rangle \\ \vdots & \ddots & \vdots \\ \langle w_n, w_1 \rangle & \cdots & \langle w_n, w_n \rangle \end{bmatrix}$$

belong to the same coset of $k/k^{\times 2}$. Well, expanding the basis w_j along the basis v_i, we are promised constants $a_{ij} \in k$ such that

$$w_j = \sum_{i=1}^n a_{ij} v_i.$$

Accordingly, we define the matrix

$$A \coloneqq \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}.$$

This lets us expand

$$\langle w_{j_1}, w_{j_2} \rangle = \left\langle \sum_{i_1=1}^n a_{i_1 j_1} v_{i_1}, \sum_{i_2=1}^n a_{i_2 j_2} v_{i_2} \right\rangle = \sum_{i_1, i_2 = 1}^n a_{i_1 j_1} \langle v_{i_1}, v_{i_2} \rangle a_{i_2 j_2} = \sum_{i_1, i_2} (A^\intercal)_{j_1 i_1} \langle v_{i_1}, v_{i_2} \rangle A_{i_2 j_2}.$$

It follows that

$$\begin{bmatrix} \langle w_1, w_1 \rangle & \cdots & \langle w_1, w_n \rangle \\ \vdots & \ddots & \vdots \\ \langle w_n, w_1 \rangle & \cdots & \langle w_n, w_n \rangle \end{bmatrix} = A^\intercal \begin{bmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_n \rangle \\ \vdots & \ddots & \vdots \\ \langle v_n, v_1 \rangle & \cdots & \langle v_n, v_n \rangle \end{bmatrix} A,$$

so

$$\det \begin{bmatrix} \langle w_1, w_1 \rangle & \cdots & \langle w_1, w_n \rangle \\ \vdots & \ddots & \vdots \\ \langle w_n, w_1 \rangle & \cdots & \langle w_n, w_n \rangle \end{bmatrix} = (\det A)^2 \det \begin{bmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_n \rangle \\ \vdots & \ddots & \vdots \\ \langle v_n, v_1 \rangle & \cdots & \langle v_n, v_n \rangle \end{bmatrix}.$$

Noting that $\det A \neq 0$ because it is a change of basis matrix (the columns are linearly independent because the $w_\bullet$ are linearly independent), we are done. ∎

We also note that all of our work showing that the trace form is nondegenerate is not in vain.

Lemma 5.250. A bilinear form $\langle \cdot, \cdot \rangle$ on a finite-dimensional k-vector space V is nondegenerate if and only if its discriminant is nonzero.

Proof. Fix a basis $\{v_i\}_{i=1}^n$ of V. Then we see that

$$\det \begin{bmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_n \rangle \\ \vdots & \ddots & \vdots \\ \langle v_n, v_1 \rangle & \cdots & \langle v_n, v_n \rangle \end{bmatrix} = 0$$

if and only if there is a linear relation among the columns. Namely, the discriminant is zero if and only if there are constants not all zero $\{a_i\}_{i=1}^n$ such that

$$\sum_{i=1}^n a_i \langle v_i, v_j \rangle = 0$$

for each v_j. This is equivalent to having constants not all zero $\{a_i\}_{i=1}^n$ such that

$$\left\langle \sum_{i=1}^n a_i v_i, v_j \right\rangle = 0$$

for each v_j. But because the $v_\bullet$ form a basis, this is equivalent to having some vector $v \neq 0$ such that

$$\langle v, v_j \rangle = 0$$

for each v_j. (Having the constants $a_\bullet$ not all zero of course gives some $v \neq 0$, and conversely, some $v \neq 0$ can be expanded along the basis $v_\bullet$ to give the constants $a_\bullet$ which cannot be all zero.) Continuing, having v such that $\langle v, v_j \rangle = 0$ for each v_j implies that any vector $w = \sum_{i=1}^n b_i v_i \in V$ has

$$\langle v, w \rangle = \sum_{i=1}^n b_i \langle v, v_i \rangle = 0.$$

And conversely, if $\langle v, w \rangle = 0$ for each $w \in V$, then $\langle v, v_j \rangle = 0$ for each v_j.

Thus, the discriminant vanishes if and only if there is some $v \neq 0$ such that $\langle v, w \rangle = 0$ for each $w \in V$, which is exactly the condition for $\langle \cdot, \cdot \rangle$ being degenerate. ■

In particular, we showed that the trace form on L/K is nondegenerate as long as L/K is inseparable, so in these cases, we can compute the discriminant of the trace form and know that is is nonzero.

5.7.3 Discriminant: Computation

In practice, here is one way to compute the discriminant of a field extension.

Proposition 5.251. Fix L/K a finite extension where $L = K(\alpha)$ for some $\alpha \in L$. Then the discriminant of the trace form of L/K is the discriminant of the monic minimal polynomial $f(X) \in K[X]$ for α.

Proof. This won't actually matter for the proof, but for psychological reasons, we note that L/K is inseparable if and only if the trace form is degenerate if and only if the discriminant of L/K is 0. And on the other side, L/K is inseparable if and only if α is inseparable if and only if f has a double root if and only if the discriminant of f vanishes.

So now take L/K separable, which will make the notation a bit easier later. Fix

$$f(X) = X^n + \sum_{k=0}^{n-1} a_k X^k \in K[X]$$

our monic irreducible polynomial for α over K. The key to compute the discriminant will be to use the power basis

$$\left\{1, \alpha, \ldots, \alpha^{n-1}\right\}.$$

Namely, we want to compute the determinant of

$$\begin{bmatrix} \mathrm{T}(\alpha^0) & \mathrm{T}(\alpha^1) & \cdots & \mathrm{T}\left(\alpha^{n-1}\right) \\ \mathrm{T}(\alpha^1) & \mathrm{T}(\alpha^2) & \cdots & \mathrm{T}\left(\alpha^{n}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{T}\left(\alpha^{n-1}\right) & \mathrm{T}\left(\alpha^{n}\right) & \cdots & \mathrm{T}\left(\alpha^{2n-2}\right) \end{bmatrix}.$$

For this, we set $\{\sigma_k\}_{k=1}^n$ our embeddings $L \hookrightarrow \overline{K}$ fixing K. Because L/K is separable, we are reassured $n = [L : K]$. In particular, fixing indices i and k,

$$\mathrm{T}\left(\alpha^{i+k}\right) = \sum_{j=1}^{n-1} \sigma_j(\alpha)^i \sigma_j(\alpha)^k.$$

But we can view this expansion as the matrix multiplication

$$\underbrace{\begin{bmatrix} \sigma_1\alpha^0 & \cdots & \sigma_n\alpha^0 \\ \vdots & \ddots & \vdots \\ \sigma_1\alpha^{n-1} & \cdots & \sigma_n\alpha^{n-1} \end{bmatrix}}_{M^\intercal} \underbrace{\begin{bmatrix} \sigma_1\alpha^0 & \cdots & \sigma_1\alpha^{n-1} \\ \vdots & \ddots & \vdots \\ \sigma_n\alpha^0 & \cdots & \sigma_n\alpha^{n-1} \end{bmatrix}}_{M},$$

so we are interested in

$$\begin{bmatrix} \mathrm{T}(\alpha^0) & \mathrm{T}(\alpha^1) & \cdots & \mathrm{T}\left(\alpha^{n-1}\right) \\ \mathrm{T}(\alpha^1) & \mathrm{T}(\alpha^2) & \cdots & \mathrm{T}\left(\alpha^{n}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{T}\left(\alpha^{n-1}\right) & \mathrm{T}\left(\alpha^{n}\right) & \cdots & \mathrm{T}\left(\alpha^{2n-2}\right) \end{bmatrix} = \det M^\intercal \cdot \det M = (\det M)^2.$$

But now we see that M is a "Vandermonde determinant," for which we have the following theory.

Lemma 5.252. Working in the polynomial ring $\mathbb{Z}[X_0, \dots, X_{n-1}]$, we have that

$$\det \begin{bmatrix} X_0^0 & \cdots & X_0^{n-1} \\ \vdots & \ddots & \vdots \\ X_{n-1}^0 & \cdots & X_{n-1}^{n-1} \end{bmatrix} = \prod_{0 \le \ell < k < n} (X_k - X_\ell).$$

Proof. For brevity, let the determinant be $D \in \mathbb{Z}[X_0, \dots, X_{n-1}]$.

Choosing distinct indices X_k, X_ℓ, we note that there is an evaluation sending X_k to X_ℓ. But this makes the X_k row equal to X_ℓ, so properties of the determinant implies that the entire determinant must vanish after doing this.

So viewing the determinant a some giant polynomial where the variable X_k, evaluating at $X_k = X_\ell$ for each $k \neq \ell$ gives the polynomial a root. It follows that[13]

$$(X_k - X_\ell) \mid D$$

for each $k \neq \ell$. But now we see that $(X_k - X_\ell)$ generates a prime ideal because these are the elements which vanish on setting $X_k = X_\ell$, which is prime because polynomial rings over $\mathbb{Z}$ are integral domains.

In particular, $X_k - X_\ell$ is irreducible, and because each of these are distinct irreducibles, we see that

$$\prod_{0 \le \ell < k < n} (X_k - X_\ell) \,\Big|\, D.$$

We now compare the total degrees and leading coefficients of both sides.

- By direct expansion, we see that

$$D = \sum_{\sigma \in \mathrm{Sym}(\mathbb{Z}/n\mathbb{Z})} (\operatorname{sgn} \sigma) \prod_{k=0}^{n-1} X_k^{\sigma k}.$$

  In particular, by nature of the sum on permutation, only $\sigma = \mathrm{id}$ will add to the

$$X_0^0 X_1^1 \cdots X_{n-1}^{n-1}$$

  term, so this term will have coefficient $+1$. Additionally, we note that the degree of any term in D will be

$$\deg \prod_{k=0}^{n-1} X_k^{\sigma k} = \sum_{k=0}^{n-1} \sigma(k) = \sum_{k=0}^{n-1} k = \frac{n(n-1)}{2},$$

  for any $\sigma \in S_n$.

[13] Formally, it is true that $f \in R[X]$ has $f(a) = 0$ if and only if $X - a \mid f(X)$.

- On the other hand, we note that the massive polynomial
$$\prod_{0\le \ell<k<n} (X_k - X_\ell)$$
is a product of $\binom{n}{2} = \frac{n(n-1)}{2}$ terms of degree 1, so the entire polynomial has degree $\frac{n(n-1)}{2}$. Additionally, we claim that the coefficient of
$$X_0^0 X_1^1 \cdots X_{n-1}^{n-1}$$
is $+1$. Indeed, the leading coefficient (under the lexicographic ordering) of the product
$$\prod_{0\le \ell<k<n} (X_k - X_\ell)$$
can be found by taking the leading coefficient of each of the factors Lemma 4.91. So we see that our leading coefficient is
$$\prod_{0\le \ell<k<n} X_k.$$
Here, each X_k will appear k times, so our leading coefficient is precisely $+1X_0^0 X_1^1 \cdots X_{n-1}^{n-1}$.

So to finish, we note that D and the product have the same degree of $\frac{n(n-1)}{2}$ and have the same nonzero coefficient for $X_0^0 X_1^1 \cdots X_{n-1}^{n-1}$, so they must be the same polynomial because we already know that the product divides D. ■

Thus, we see that
$$(\det M)^2 = \prod_{1\le k<\ell\le n} (\sigma_k\alpha - \sigma_\ell\alpha)^2.$$
We now note that the $\sigma_\bullet\alpha$ will all be roots of f, and they must all be distinct because $\sigma_k\alpha = \sigma_\ell\alpha$ implies that $\sigma_k = \sigma_\ell$. So the n values $\sigma_k\alpha$ will have to loop through all $\deg f$ distinct roots of f. So the above product is indeed the discriminant of f. ■

Noting that the discriminant L/K only depends on data internal to the field extension, we see that it provides a fairly useful invariant, arguably the second most important after the degree.

Example 5.253. Set $L_1 = \mathbb{Q}[X]/\left(X^3 - X + 1\right)$ and $L_2 = \mathbb{Q}[X]/\left(X^3 + X + 1\right)$. We can compute
$$\operatorname{disc} L_1 = \operatorname{disc}\left(X^3 - X + 1\right) = -4(-1)^3 - 27 = -23,$$
and
$$\operatorname{disc} L_2 = \operatorname{disc}\left(X^3 + X + 1\right) = -4 \cdot 1^3 - 27 = -31.$$
Now we see that $\operatorname{disc} L_1/\operatorname{disc} L_2$ is not a square in $\mathbb{Q}$, so these fields are not isomorphic.

In the case of algebraic number theory, the discriminant is even more important. We have been working with the discriminant up to a square in K, but one can do better than this in the case of number fields.

Definition 5.254 (Discriminant, number fields)**.** Fix $K/\mathbb{Q}$ a number field. Then the *discriminant of* $K/\mathbb{Q}$ is the least possible (in terms of magnitude) discriminant of the trace form when we specifically choose a basis of algebraic integers.

In particular, we note that the determinant computation for a basis $\{v_k\}_{k=1}^n$ of $K/\mathbb{Q}$ will be
$$\det \begin{bmatrix} \mathrm{T}(v_1v_1) & \cdots & \mathrm{T}(v_1v_n) \\ \vdots & \ddots & \vdots \\ \mathrm{T}(v_nv_1) & \cdots & \mathrm{T}(v_nv_n) \end{bmatrix} \in \mathbb{Q},$$
and when the v_k are algebraic integers, we will have that the discriminant is an integer as well, so the discriminant is a rational integer and hence an integer. So it makes sense to define the discriminant as the least possible.

Remark 5.255. There's a theorem due to Hermite which says that there are only finitely many algebraic number fields with a given discriminant, so the discriminant is a pretty good invariant.

5.7.4 Image of the Norm

Fix L/K a finite extension. Number theorists are interested in the image of the norm map $\mathrm{N} : L^\times \to K^\times$.

Example 5.256. The image of $\mathrm{N}_{\mathbb{R}}^{\mathbb{C}} : \mathbb{C}^\times \to \mathbb{R}^\times$ is $\mathbb{R}_{>0}$, and in particular, $\mathbb{R}^\times / \operatorname{im} \mathrm{N} \cong \mathbb{Z}/2\mathbb{Z}$. It is somewhat surprising that we are getting a finite quotient.

Example 5.257. The image of $\mathrm{N} : \mathbb{Q}(i) \to \mathbb{Q}$ consists of all rationals which are the sum of two squares, which is a bit hard to classify. A little bit of elementary number theory is able to show that this image consists of all positive rationals x such that $\nu_p(x)$ is even for each prime $p \equiv 3 \pmod 4$.

However, we can at least gain some control in the easiest cases.

Exercise 5.258. The map $\mathrm{N} : L^\times \to K^\times$ is onto when L and K are finite.

Proof. Because K is finite, define $K = \mathbb{F}_q$, where q is some prime-power, and then we see that $L = \mathbb{F}_{q^n}$ for some positive integer n. Now, $\mathrm{Gal}(L/K)$ is cyclic of order n generated by the Frobenius automorphism $\sigma : \alpha \mapsto \alpha^q$. So we find that

$$\mathrm{N}(\alpha) = \prod_{\sigma \in \mathrm{Gal}(L/K)} \sigma\alpha = \prod_{k=0}^{n-1} \sigma^k(\alpha) = \prod_{k=0}^{n-1} \alpha^{q^k} = \alpha^{(q^n-1)/(q-1)}.$$

We now finish with a size argument. Notice that $\ker \mathrm{N}$ has at most $\frac{q^n-1}{q-1}$ elements because each element of the kernel will be a root of the polynomial

$$X^{(q^n-1)/(q-1)} - 1 = 0.$$

But now we notice that L has $q^n - 1$ elements, and $K^\times$ has $q - 1$ elements, so we can simply size bound by

$$\#L^\times = \#\operatorname{im} \mathrm{N} \cdot \#\ker \mathrm{N} \le \#K^\times \cdot \#\ker \mathrm{N} \le (q-1) \cdot \frac{q^n-1}{q-1} = q^n - 1 = \#L^\times,$$

so equalities are forced. Namely, $\#\operatorname{im} \mathrm{N} = \#K^\times$, so $\operatorname{im} \mathrm{N} = K^\times$, finishing. ∎

5.7.5 Solving a Cubic

Let's solve a cubic by radicals, for fun, though this is somewhat useless because we can well-approximate it other ways.

Exercise 5.259. We solve the cubic equation $f(X) \coloneqq X^3 + bX + c = X^3 + X + 1$ by radicals.

Proof. We can check that the Galois group of f is S_3 because we just showed that the discriminant of f is $\Delta^2 \coloneqq -4b^3 - 27c^2 = -31 \notin \mathbb{Q}^{\times 2}$ above, where here we are using Proposition 5.126. Regardless, we will most of the solution agnostic to b and c.

Instead of solving f over $\mathbb{Q}$, we start by throwing in all the roots of unity we could want. Because we will be making a chain from a group of order 6, the only possible quotients in the chain will have order 2 or 3, so

it suffices to include the square and cube roots of unity. So we set ω to be a primitive third root of unity and solve f over $\mathbb{Q}(\omega)$.

As in our discussion of solving polynomials by radicals, we need a chain witnessing that S_3 is solvable, so we use

$$S_3 \supseteq A_3 \supseteq \langle \mathrm{id} \rangle.$$

By Galois theory, this will correspond to the chain of normal extensions

$$\mathbb{Q}(\omega) \subseteq K \subseteq L,$$

where

$$\mathrm{Gal}(K/\mathbb{Q}(\omega)) \cong S_3/A_3 \cong \mathbb{Z}/2\mathbb{Z} \qquad \text{and} \qquad \mathrm{Gal}(L/K) \cong A_2/\langle \mathrm{id} \rangle \cong \mathbb{Z}/3\mathbb{Z}.$$

In particular, these are all cyclic extensions, by our Kummer theory work, they will be generated by radicals, and in fact we can find these radicals by finding eigenvectors of the Galois groups.

We work these out one at a time.

(i) To start, we need to talk about K, so we need to find an element fixed by $A_3 = \langle (123) \rangle$ but not S_3. Well, suppose that the roots of $f(X)$ are α, β, γ, and we would like an expression fixed by A_3 but not S_3. For this, we use

$$\Delta \coloneqq (\alpha - \beta)(\beta - \gamma)(\gamma - \delta).$$

We see that the action of A_3 on Δ will consist of an even number of transpositions and hence an even number of signs, so Δ will be fixed by A_3. But on the other hand, the transposition (α, β) sends $\Delta \mapsto -\Delta$, so Δ is not fixed by S_3.

Now, we see that Δ^2 is the discriminant of f, which we worked out as $\Delta^2 = -31$ above. So we find that K has the nontrivial element $\sqrt{-31}$, and this must fully generate because $K/\mathbb{Q}(\omega)$ is quadratic. So

$$K = \mathbb{Q}(\omega, \sqrt{-31}).$$

(ii) Next we need to find a generator of L, where we know that L/K is a cyclic extension of order 3. Namely, we want eigenvectors of our A_3-action of L. Back in our work in Kummer theory, we found the eigenvectors

$$v + \omega^{-1}\sigma v + \omega^{-2}\sigma^2 v,$$

for some $v \in L$ and $\sigma \in A_3$. These will work for our purposes provided that we get out a nonzero eigenvector. With this in mind, we choose $v = \alpha$ and hope we get lucky. So we set

$$\begin{cases} x \coloneqq \alpha + \beta + \gamma, \\ y \coloneqq \alpha + \omega^{-1}\beta + \omega^{-2}\gamma, \\ z \coloneqq \alpha + \omega^{-2}\beta + \omega^{-1}\gamma, \end{cases}$$

which are eigenvectors of the given type because A_3 cycles α, β, γ, meaning that there is $\sigma \in A_3$ with $\sigma = (\alpha, \beta, \gamma)$. Namely, we can see that $\sigma x = x$ and $\sigma y = \omega y$ and $\sigma z = \omega^2 z$.

We now have enough tools to finish. We know that $x = 0$ (by Vieta's formulae) and would like to find y and z explicitly. Because y and z are eigenvectors with eigenvalue a power of ω (by construction), we know $y^3, z^3 \in L^{A_3} = K$, so we will find y^3 and z^3.

Now, we see that (β, γ) sends y to z and so will send y^3 to z^3. So the orbit of y^3 under S_3 is at least $\{y^3, z^3\}$, but because y^3 is fixed by A_3, the orbit has size at most 2. So we see that $\{y^3, z^3\}$ is an orbit in $K/\mathbb{Q}(\omega)$, so

$$y^3 + z^3, \quad y^3 z^3 \in \mathbb{Q}(\omega)$$

because these will both be fixed by all of S_3. Using the fact that $\alpha + \beta + \gamma = 0$ and the help of SageMath, we can compute

$$y^3 + z^3 = 27\alpha\beta\gamma = -27c \qquad \text{and} \qquad yz = -3(\alpha\beta + \beta\gamma + \gamma\alpha) = -3b.$$

Thus, $y^3z^3 = -27b^3$, so y^3 and z^3 are the roots of $T^2 + 27cT - 27b^3 = 0$, which are

$$y^3, z^3 = \frac{-27c \pm \sqrt{(-27c)^2 - 4(-27b^3)}}{2} = \frac{-27c \pm 3\Delta i\sqrt{3}}{2}.$$

As a sanity-check, we do see that $y^3, z^3 \in K = \mathbb{Q}(\omega, \Delta)$, as predicted. Anyways, we can set (up to ordering of α, β, γ),

$$y = \sqrt[3]{\frac{-27c + 3\Delta\sqrt{-3}}{2}}, \qquad \text{and} \qquad z = \sqrt[3]{\frac{-27c - 3\Delta\sqrt{-3}}{2}}.$$

There are three possible cube roots for y and z, but we can choose compatible y and z by ensuring $yz = -3b$. Being off by a factor of ω induces the transformation $y \mapsto y\omega^k$ and $z \mapsto z\omega^{-k}$, which will merely permute the roots $\{\alpha, \beta, \gamma\}$ in the definition of y and z, which is safe.

Anyways, we see that a linear-algebra inspired computation shows

$$\alpha = \frac{x + y + z}{3}, \qquad \beta = \frac{x + \omega y + \omega^2 z}{3}, \qquad \gamma = \frac{x + \omega^2 y + \omega z}{3},$$

which upon seeing $x = 0$ gives

$$\begin{aligned}
\alpha &= \frac{1}{3}\left(\sqrt[3]{\frac{-27c + 3\Delta\sqrt{-3}}{2}} + \sqrt[3]{\frac{-27c - 3\Delta\sqrt{-3}}{2}}\right) \\
\beta &= \frac{1}{3}\left(\omega\sqrt[3]{\frac{-27c + 3\Delta\sqrt{-3}}{2}} + \omega^2\sqrt[3]{\frac{-27c - 3\Delta\sqrt{-3}}{2}}\right) \\
\gamma &= \frac{1}{3}\left(\omega^2\sqrt[3]{\frac{-27c + 3\Delta\sqrt{-3}}{2}} + \omega\sqrt[3]{\frac{-27c - 3\Delta\sqrt{-3}}{2}}\right).
\end{aligned}$$

Choosing the right cube roots for y and z does let us recover $\alpha \approx -0.682327$ and $\beta \approx 0.341163 + 1.161541i$ and $\gamma \approx 0.341163 - 1.161541i$. So we are done. ■

5.7.6 Solving a Quartic

We can also do this for degree-4. The solution in Wolfram Alpha took about a page, so we won't bother doing this fully explicitly, but we will sketch.

Exercise 5.260. We sketch how to solve the quartic

$$X^4 + bX^2 + cX + d = 0.$$

Proof. Our Galois group is at worst S_4 and certainly a subgroup of it, so we don't lose anything by forcing the Galois group to actually be S_4. Well, we know that S_4 is solvable, for which we use the chain

$$S_4 \supseteq A_4 \supseteq (\mathbb{Z}/2\mathbb{Z})^2 \supseteq \langle \mathrm{id} \rangle.$$

Namely, if the Galois group is actually $G \subseteq S_4$, then we merely have to intersect each of the above subgroups with G to still have a chain witnessing the solvability of G. To be explicit,

$$(\mathbb{Z}/2\mathbb{Z})^2 = \{\mathrm{id}, (12)(34), (13)(24), (14)(23)\}$$

is our subgroup of A_4 with index 3.

But we note that $S_4/(\mathbb{Z}/2\mathbb{Z})^2 \cong S_3$, so we should be able to reduce to the cubic case.[14] Using our Kummer theory, we start by adding in all the roots of unity we could ever want—which are the roots of unity dividing the order of our quotient groups, namely $\{1,2,3,4\}$. So we are looking at a chain of extensions

$$\mathbb{Q}(i,\omega) \subseteq K \subseteq L \subseteq M,$$

where $\mathrm{Gal}(L/\mathbb{Q}(i,\omega)) \cong S_3$ and $\mathrm{Gal}(M/L) \cong (\mathbb{Z}/2\mathbb{Z})^2$. So we see that, indeed, $L/\mathbb{Q}(i,\omega)$ should be the splitting field of a cubic by, say, finding a normal basis element. This is possible but quite painful.

Remark 5.261. One can try to do a similar thing to solve quintics, but the finest chain we can make is

$$S_5 \supseteq A_5 \supseteq \langle \mathrm{id} \rangle,$$

and undoing A_5 essentially requires a quintic.

To find L, we want to find expressions involving our roots $\alpha_1,\alpha_2,\alpha_3,\alpha_4$ which are fixed by $(\mathbb{Z}/2\mathbb{Z})^2 \subseteq S_4$. Here, we let S_4 act on the roots by acting on the indices.

Well, motivated by our Kummer theory, we look for eigenvectors of our elements in $(\mathbb{Z}/2\mathbb{Z})^2$, so we see that

$$z \coloneqq \alpha_1 + \alpha_2 - \alpha_3 - \alpha_4$$

is fixed by $(12)(34)$ and is an eigenvector with the correct eigenvalue (of -1) by the $(13)(24)$ and $(14)(23)$, so this element will generate one of the quadratic subfields M/L, and in particular its square will be in L. Looking at the orbit of z^2 under S_3, we find the elements

$$\begin{cases} y_1 \coloneqq (\alpha_1+\alpha_2-\alpha_3-\alpha_4)^2, \\ y_2 \coloneqq (\alpha_1-\alpha_2+\alpha_3-\alpha_4)^2, \\ y_3 \coloneqq (\alpha_1-\alpha_2-\alpha_3+\alpha_4)^2, \end{cases}$$

which are all permuted by the Galois group S_4, but they are fixed by $(\mathbb{Z}/2\mathbb{Z})^2$. So $y_1,y_2,y_3 \in L$ will be the roots of some cubic (with coefficients in $\mathbb{Q}(i,\omega)$), presumably with Galois group as large as possible inside S_3, and so they will generate our L for degree reasons. We can find which cubic by writing down

$$y^3 + By^3 + Cy^2 + D = 0$$

and solving for the coefficients B,C,D by plugging in y_1,y_2,y_3 and treating this as a massive system of equations. We will not write this out, but we should get

$$y^3 - 2by^2 + \left(b^2 - d\right)y + c^2 = 0.$$

This lets us solve for our $y_\bullet$ by reducing to the cubic case. By choosing our square roots correctly, we can extract $\sqrt{y_1} = \alpha_1+\alpha_2-\alpha_3-\alpha_4$ and its friends.

Remark 5.262 (Nir)**.** It is the fact that we need to solve a cubic in the middle of solving a quartic that makes the quartic formula so painful. In contrast, solving the cubic did need to solve a quadratic, but solving quadratics is significantly more automatic than cubics.

Lastly, we need to convert our $y_\bullet$ to $\alpha_\bullet$. So we set $\sqrt{y_0} \coloneqq \alpha_1+\alpha_2+\alpha_3+\alpha_4 = 0$ and notice that

$$\begin{bmatrix} \sqrt{y_0} \\ \sqrt{y_1} \\ \sqrt{y_2} \\ \sqrt{y_3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}.$$

[14] We have $S_3 \cong S_4/(\mathbb{Z}/2\mathbb{Z})^2$ because it has six elements and is not abelian because $(12)(13) = (132)$ and $(13)(12) = (123)$ are not in the same coset of $(\mathbb{Z}/2\mathbb{Z})^2$.

So we can solve for the $\alpha_\bullet$ by inverting the middle matrix (it is invertible). Alternatively, we can solve by hand

$$\begin{aligned}
\alpha_1 &= \tfrac{1}{4}\left(\sqrt{y_0}+\sqrt{y_1}+\sqrt{y_2}+\sqrt{y_3}\right),\\
\alpha_2 &= \tfrac{1}{4}\left(\sqrt{y_0}+\sqrt{y_1}-\sqrt{y_2}-\sqrt{y_4}\right),\\
\alpha_3 &= \tfrac{1}{4}\left(\sqrt{y_0}-\sqrt{y_1}+\sqrt{y_2}-\sqrt{y_4}\right),\\
\alpha_4 &= \tfrac{1}{4}\left(\sqrt{y_0}-\sqrt{y_1}-\sqrt{y_2}+\sqrt{y_4}\right).
\end{aligned}$$

This finishes the outline. ■

5.7.7 Infinite Galois Extensions: Advertisement

Let's talk a little about infinite Galois extensions M/K.

Example 5.263. Consider $\mathbb{Q} \subseteq \overline{\mathbb{Q}}$. This is an infinite Galois extension: $\overline{\mathbb{Q}}$ is the splitting field of the set of polynomials $\mathbb{Q}[X]$ over $\mathbb{Q}$, and because all polynomials in characteristic 0 are separable, this makes $\overline{\mathbb{Q}}/\mathbb{Q}$ both normal and separable.

Remark 5.264 (Nir)**.** Because of our careful phrasing, most of our normality and separability conditions (Remark 5.53, Proposition 5.56, and Proposition 5.64) will go through, with the exception of Proposition 5.64 (b).

The proof of Proposition 5.64 does need to show (a) implies (c): if L is generated by separable elements $\{\alpha_i\}_{i\in I}$, then fixing any $\alpha \in L$, we can express α in terms of finitely many $\alpha_\bullet$, so α is in a finite extension generated by separable elements, so α is separable by Proposition 5.64.

Remark 5.265 (Nir)**.** Similarly, not all of Proposition 5.76 will go through. We show (a) and (b) are equivalent. For this, we need (b) implies (a): normal is equivalent to being a splitting field of some $S \subseteq K[X]$; separable implies that each polynomial in S is separable. Combining these observations finishes.

However, not all of our Galois theory will go through so smoothly.

Warning 5.266. When M/K is an infinite extension, then subgroups of $\mathrm{Gal}(M/K)$ do not correspond to intermediate extensions by taking fixed fields.

5.7.8 Krull Topology: Galois Edition

We continue to work with M/K an infinite Galois extension. The way to fix the Galois correspondence is to give $\mathrm{Gal}(M/K)$ a topology, and intermediate extensions will correspond to open subgroups.

Here is the idea for the topology we are about to create, the "Krull topology."

Idea 5.267. The Krull topology on $\mathrm{Gal}(M/L)$ is the coarsest topology making restriction maps continuous.

To be explicit, fix some intermediate extension $K \subseteq L \subseteq M$ such that L/K is finite and Galois. Then we have the restriction map

$$\varphi : \mathrm{Gal}(M/K) \to \mathrm{Gal}(L/K).$$

This map is sufficiently natural, so we hope to make it continuous. Here, finite sets like $\mathrm{Gal}(L/K)$ should get the discrete topology, so we see that, given $\sigma \in \mathrm{Gal}(L/K)$, we would like the pre-image

$$\varphi^{-1}(\sigma)$$

to be an open set. Because $\ker \varphi = \mathrm{Gal}(M/L)$, we see that we are asking for $\overline{\sigma}\,\mathrm{Gal}(M/L)$ to be an open set, where here we are fixing $\overline{\sigma} \in \mathrm{Gal}(M/L)$ to be any extension of $\sigma \in \mathrm{Gal}(L/K)$.

In other words, for any $\sigma \in G$ and finite intermediate Galois extension $K \subseteq L \subseteq M$, we are declaring that $\sigma\,\mathrm{Gal}(M/L)$ should be an open set. And these are all of the open sets we ask for.

Definition 5.268 (Krull topology, I)**.** Fix M/K a Galois extension. Then we define the *Krull topology* on $\mathrm{Gal}(M/K)$ as having basis given by the subsets $\sigma\,\mathrm{Gal}(M/L)$, where $\sigma \in \mathrm{Gal}(M/K)$ and L is some finite Galois subextension of K.

We quickly check that these subsets do actually form a basis, and not just a sub-basis.

Lemma 5.269. Fix M/K a Galois extension. Then the subsets $\sigma\,\mathrm{Gal}(M/L) \subseteq \mathrm{Gal}(M/K)$, where $\sigma \in \mathrm{Gal}(M/K)$ and L is some finite Galois extension of K, do in fact form a basis of a topology.

Proof. We have already declared $\mathrm{Gal}(M/K)$ an open set, so we don't have to worry about covering. So fix $\sigma_1, \sigma_2 \in \mathrm{Gal}(M/K)$ and L_1, L_2 finite Galois subextensions of M/K. We would like to study

$$\sigma_1\,\mathrm{Gal}(M/L_1) \cap \sigma_2\,\mathrm{Gal}(M/L_2).$$

Note that any σ in the above intersection will have $\sigma|_{L_1 \cap L_2} = \sigma_1|_{L_1 \cap L_2}$ and $\sigma|_{L_1 \cap L_2} = \sigma_2|_{L_1 \cap L_2}$, so for the above intersection to be nonempty, we must have

$$\sigma_1|_{L_1 \cap L_2} = \sigma_2|_{L_1 \cap L_2}.$$

So we will suppose that the above condition holds, for otherwise the empty union of basis elements will equal to the empty set that we need. Let $\sigma_0 \coloneqq \sigma_1|_{L_1 \cap L_2} \in \mathrm{Gal}((L_1 \cap L_2)/K)$.

So we are now looking for $\sigma \in \mathrm{Gal}(M/K)$ and L/K a finite Galois subextension of M/K such that

$$\sigma\,\mathrm{Gal}(M/K) \subseteq \sigma_\bullet\,\mathrm{Gal}(M/L_\bullet)$$

for each index. At this point we recall that, as noted in our work earlier, $\sigma_\bullet\,\mathrm{Gal}(M/L_\bullet)$ consists of the elements which restrict to $\sigma_\bullet|_{L_\bullet}$ on $L_\bullet$. So now we set $L_0 \coloneqq L_1 \cap L_2$ and $L \coloneqq L_1 L_2$ so that

$$\mathrm{Gal}(L/L_0) \to \mathrm{Gal}(L_1/L_0) \times \mathrm{Gal}(L_2/L_0),$$

is an isomorphism, essentially for size reasons.[15] Now, $\sigma_1|_{L_1} \cdot \sigma_0^{-1} \in \mathrm{Gal}(L_1/L_0)$ and $\sigma_1|_{L_2} \cdot \sigma_0^{-1} \in \mathrm{Gal}(L_2/L_0)$ (namely, fixing L_0 by construction of σ_0), so we can be promised some $\tau \in \mathrm{Gal}(L/L_0)$ such that

$$\tau|_{L_\bullet} = \sigma_\bullet|_{L_\bullet} \cdot \sigma_0^{-1}.$$

We extend τ and σ_0 up to $\mathrm{Gal}(M/K)$ without further remark, and we see that

$$(\tau\sigma_0)|_L\,\mathrm{Gal}(M/L)$$

will restrict to $(\tau\sigma_0)|_{L_\bullet} = \sigma_\bullet|_{L_\bullet}$ on $L_\bullet$. This is what we needed. ∎

While we're here, we might as well check that our restriction maps are actually continuous.

Proposition 5.270. Fix M/K a Galois extension. Then, given a Galois subextension L/K of M (not necessarily finite!) the map

$$\cdot|_L : \mathrm{Gal}(M/K) \to \mathrm{Gal}(L/K)$$

is continuous, where the Galois groups have been given the Krull topology.

[15] Injectivity is because $L = L_1 L_2$. Surjectivity is by size because $[L : L_0] = [L_1 : L_0][L_2 : L_0]$, which we show by showing $\mathrm{Gal}(L/L_2) \to \mathrm{Gal}(L_1/L_0)$ is an isomorphism, which is not easy but not too hard.

Proof. It is sufficient that the pre-image of basis elements of $\mathrm{Gal}(L/K)$ are open in $\mathrm{Gal}(M/K)$. Well, fix $\sigma \in \mathrm{Gal}(L/K)$ and F some finite Galois subextension of L/K so that we want to show that the pre-image of $\sigma\,\mathrm{Gal}(L/F)$ is open in $\mathrm{Gal}(M/K)$.

Well, we note that $\sigma\,\mathrm{Gal}(L/F)$ consists exactly of the $\tau \in \mathrm{Gal}(L/K)$ which restrict as $\tau|_F = \sigma|_F$. Now, $\tau \in \mathrm{Gal}(M/K)$ is in the pre-image of $\sigma\,\mathrm{Gal}(L/F)$ if and only if $\tau|_L$ is in $\sigma\,\mathrm{Gal}(L/F)$ if and only if $\tau|_F = \sigma|_F$. Thus, the pre-image is

$$\overline{\sigma}\,\mathrm{Gal}(M/F),$$

where we have extended σ to some $\overline{\sigma} \in \mathrm{Gal}(M/K)$. This finishes. ∎

As an aside, we note that we may technically remove the Galois condition from our basis elements.

Lemma 5.271. Fix M/K a Galois extension. Then the subsets $\sigma\,\mathrm{Gal}(M/L) \subseteq \mathrm{Gal}(M/K)$, where $\sigma \in \mathrm{Gal}(M/K)$ and L is some finite extension of K (not necessarily Galois!), form a basis for the Krull topology as well.

Proof. Set $\mathcal{B}$ to be the basis for the Krull topology and $\mathcal{B}'$ to be the set of elements defined in the lemma. It is enough to show that the elements of $\mathcal{B}$ are open in the topology induced by $\mathcal{B}'$ and conversely.

- The elements of $\mathcal{B}$ are indeed open in $\mathcal{B}'$ because elements of $\mathcal{B}$ can be written as $\sigma\,\mathrm{Gal}(M/L)$ where $\sigma \in \mathrm{Gal}(M/K)$, and L is some finite (Galois) subextension of M/K.
- Fix some element $\sigma\,\mathrm{Gal}(M/L) \in \mathcal{B}$, where L is some finite subextension of M/K. We may embed L into some finite Galois subextension L'/K, and we note that

  $$\mathrm{Gal}(M/L') \subseteq \mathrm{Gal}(M/L).$$

  In particular, we notice that, for each $\tau \in \mathrm{Gal}(M/L)$, we have $\tau\,\mathrm{Gal}(M/L') \subseteq \mathrm{Gal}(M/L)$, so we see that

  $$\sigma\,\mathrm{Gal}(M/L) = \bigcup_{\tau \in \mathrm{Gal}(M/L)} \{\sigma\tau\} \subseteq \bigcup_{\tau \in \mathrm{Gal}(M/L)} \underbrace{\sigma\tau\,\mathrm{Gal}(M/L')}_{\in \mathcal{B}} \subseteq \bigcup_{\tau \in \mathrm{Gal}(M/L)} \sigma\tau\,\mathrm{Gal}(M/L),$$

  is $\sigma\,\mathrm{Gal}(M/L)$. So we get equalities, so $\sigma\,\mathrm{Gal}(M/L)$ is indeed open under the basis $\mathcal{B}$. ∎

So here is our second version of the Krull topology.

Definition 5.272 (Krull topology, II)**.** Fix M/K a Galois extension. Then we define the *Krull topology* on $\mathrm{Gal}(M/K)$ as having basis given by the subsets $\sigma\,\mathrm{Gal}(M/L)$, where $\sigma \in \mathrm{Gal}(M/K)$ and L is some finite subextension of K.

Remark 5.273. We remark here that the Krull topology on $G \coloneqq \mathrm{Gal}(M/K)$ satisfies the following two properties, which we won't bother checking.

(a) The composition map $G \times G \to G$ is continuous.

(b) The inversion map $G \to G$ is continuous.

These two properties makes G into a "topological group," a notion which we won't use directly but worth knowing about.

5.7.9 Krull Topology: Profinite Edition

As before, we work with M/K a Galois extension. We will build the Krull topology in a more group-centric way. The main idea here is that some $\sigma \in \mathrm{Gal}(M/K)$ can be tracked by its various restrictions to finite Galois subextensions of M/K. That is, we have a map (a homomorphism, in fact)

$$\mathrm{Gal}(M/K) \to \prod_{\substack{K \subseteq L \subseteq M \\ L/K \text{ fin., Gal.}}} \mathrm{Gal}(L/K)$$

induced by gluing all of our restrictions together. Here the product is over finite Galois subextensions of M/K, and we will abbreviate it to $K \subseteq L \subseteq M$ in the discussion that follows.

In fact, we can be more precise about the image. Namely, if we have a chain of finite Galois subextensions $K \subseteq L_1 \subseteq L_2 \subseteq M$, then a given $\sigma \in \mathrm{Gal}(N/K)$ has

$$\sigma|_{L_2}|_{L_1} = \sigma|_{L_1}$$

by the nature of restriction. So we really have a map

$$\mathrm{Gal}(M/K) \to \left\{ (\sigma_L)_L \in \prod_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K) : \sigma_{L_2}|_{L_1} = \sigma_{L_1} \text{ for each } L_1 \subseteq L_2 \right\},$$

where it is not too hard to check that the image on the right is in fact a group by the subgroup test. Alternatively, we can turn the finite Galois subextensions L of M/K into a category by inclusion and note that the right-hand side above is

$$\varprojlim_L \mathrm{Gal}(L/K) \cong \left\{ (\sigma_L)_L \in \prod_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K) : \sigma_{L_2}|_{L_1} = \sigma_{L_1} \text{ for each } L_1 \subseteq L_2 \right\}$$

by Lemma 3.150. Here we are using the fact that the map $L \mapsto \mathrm{Gal}(L/K)$ is functorial, where $L_1 \subseteq L_2$ induces the restriction map $\mathrm{Gal}(L_2/K) \to \mathrm{Gal}(L_1/K)$.

However, the map we've constructed is actually quite nice.

Proposition 5.274. Fix everything as above. Then the map

$$\mathrm{Gal}(M/K) \to \varprojlim_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K)$$

defined above is an isomorphism of groups.

Proof. For concreteness, we immediately unravel $\varprojlim_L \mathrm{Gal}(L/K)$ into the given map

$$\varphi : \mathrm{Gal}(M/K) \to \left\{ (\sigma_L)_L \in \prod_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K) : \sigma_{L_2}|_{L_1} = \sigma_{L_1} \text{ for each } L_1 \subseteq L_2 \right\}$$

by Lemma 3.150. We note that φ is a homomorphism because restriction gives

$$(\sigma\tau)|_L = \sigma|_L \circ \tau|_L$$

for any $\sigma, \tau \in \mathrm{Gal}(M/K)$ and finite Galois subextension L/K.

We now define the inverse map φ^{-1}. Given tuple $(\sigma)_L$ of the product, we define $\sigma \in \mathrm{Gal}(M/K)$ as follows: for some $\alpha \in M$, find any finite Galois extension L containing α (one exists by embedding $K(\alpha)$ into a finite Galois extension) and set

$$\sigma(\alpha) \coloneqq \sigma_L(\alpha).$$

We now have the following checks.

- We check $\sigma(\alpha)$ is well-defined: suppose L_1 and L_2 are both finite Galois subextensions of M/K containing α, then $L_1 \cap L_2$ will also be a finite Galois subextension of M/K.[16] And now we see that
$$\sigma_{L_\bullet}(\alpha) = \sigma_{L_\bullet}|_{L_1 \cap L_2}(\alpha) = \sigma_{L_1 \cap L_2}(\alpha)$$
for either $L_\bullet$, so we are done here.
- We check that $\sigma \in \mathrm{Gal}(M/K)$. Note that σ fixes K because any $\alpha \in K$ has $\sigma(\alpha) = \sigma_K(\alpha) = \alpha$, where $\sigma_K = \mathrm{id}_K$ because this is the only element of $\mathrm{Gal}(K/K)$.
And σ is an automorphism because, for any $\alpha, \beta \in M$, we can embed $K(\alpha, \beta)$ into a finite Galois extension L, and then $\sigma|_L = \sigma_L$ is an automorphism, so
$$\sigma(\alpha + \beta) = \sigma\alpha + \sigma\beta \quad \text{and} \quad \sigma(\alpha\beta) = (\sigma\alpha)(\sigma\beta),$$
by computing with the restriction σ_L.
- The map $\varphi^{-1} : (\sigma_L)_L \mapsto \sigma$ is a homomorphism. Well, fix $(\sigma_L)_L$ and $(\tau_L)_L$. Then, for any $\alpha \in M$, fix L' a finite Galois subextension of M/K so that
$$\varphi^{-1}\left((\sigma_L)_L\right)\varphi^{-1}\left((\tau_L)_L\right)(\alpha) = \varphi^{-1}\left((\sigma_L)_L\right)(\tau_{L'}\alpha) = (\sigma_{L'}\tau_{L'})(\alpha) = \varphi^{-1}\left((\sigma_L\tau_L)_L\right)(\alpha),$$
which finishes.

To finish, we need to check that φ and φ^{-1} are in fact inverses.

- Fix $\sigma \in \mathrm{Gal}(M/K)$. Then, for any $\alpha \in M$, place α into some finite Galois subextension L' of M/K so that
$$\varphi^{-1}(\varphi\sigma)(\alpha) = \varphi^{-1}\left((\sigma|_L)_L\right)(\alpha) = \sigma|_{L'}(\alpha) = \sigma(\alpha),$$
so indeed, $\varphi^{-1} \circ \varphi = \mathrm{id}$.
- Fix $(\sigma_L)_L$ in the inverse limit. Then, fix some finite Galois subextension L' of M/K so that the L' component of $\left(\varphi \circ \varphi^{-1}\right)\left((\sigma_L)_L\right)$ is $\sigma_{L'}$ because this is the restriction of $\varphi^{-1}\left((\sigma_L)_L\right)$ to L', by construction.

Now that we have homomorphisms going in both directions, we have finished verifying the group isomorphism. ■

Remark 5.275. At a high level, we could also imagine showing the above by writing
$$\mathrm{Hom}_K(M, \overline{K}) \simeq \mathrm{Hom}_K\left(\varinjlim_L L, \overline{K}\right) \simeq \varprojlim_L \mathrm{Hom}_K(L, \overline{K}),$$
where the limits are taken over finite Galois subextensions L of M/K. Now, $\mathrm{Hom}_K(L, \overline{K})$ consists of the embeddings $L \hookrightarrow \overline{K}$ fixing K, so because L is normal, we are describing $\mathrm{Gal}(L/K)$, so the above really shows $\mathrm{Gal}(M/K) \simeq \varprojlim_K \mathrm{Gal}(L/K)$.

The reason we did not follow the above remark is because we are going to need to know what the map and its inverse are somewhat shorty.

At this point we note that we have the usual embedding
$$\varprojlim_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K) \subseteq \prod_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K).$$
If we want to add a topology to everything, then we could give the finite groups $\mathrm{Gal}(L/K)$ the discrete topology that they deserve and then give the huge product the product topology. Lastly, the limit could be given the induced topology as a subset.

And now: a miracle occurs.

[16] We see $(L_1 \cap L_2)/K$ is separable because each element is in L_1 and hence separable. The extension is normal because any polynomial $f \in K[X]$ with a root in $L_1 \cap L_2$ fully splits in both L_1 and L_2 and hence in $L_1 \cap L_2$.

Theorem 5.276. Fix everything as above, with the described topologies. Then the map

$$\mathrm{Gal}(M/K) \to \varprojlim_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K)$$

is in fact a homeomorphism.

Proof. Label the given map $\varphi : \mathrm{Gal}(M/K) \to \varprojlim_L \mathrm{Gal}(L/K)$. We already know that φ is a bijection, so we need to check that φ and φ^{-1} are continuous.

- We show that φ is continuous. It suffices to show that any sub-basis element of $\varprojlim_L \mathrm{Gal}(L/K)$ has open pre-image. Well, the product topology will have sub-basis given by

$$\prod_{K \subseteq L \subseteq M} S_L \subseteq \prod_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K),$$

  where all but one of the S_L have $S_L = \mathrm{Gal}(L/K)$. In fact, we can restrict the S_L to only be basis elements of $\mathrm{Gal}(L/K)$ and still generate the full topology (as a sub-basis), which we means we can force S_L to be a a singleton $\sigma_L \in \mathrm{Gal}(L/K)$. Namely, we may define $\{S_L\}_L$ by

$$S_L = \begin{cases} \{\sigma\} & L = L_0, \\ \mathrm{Gal}(L/K) & L \neq L_0, \end{cases}$$

  for some chosen $\sigma \in \mathrm{Gal}(L_0/K)$.

  But now, checking the induced topology on the the inverse limit, we are looking at the open set

$$\prod_{K \subseteq L \subseteq M} S_L \cap \varprojlim_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K)$$

  is

$$\left\{ (\sigma_L)_L \in \prod_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K) : \sigma_{L_2}|_{L_1} = \sigma_{L_1} \text{ for each } L_1 \subseteq L_2 \text{ and } \sigma_L = \sigma \right\}.$$

  Now, pushing this through φ^{-1}, we see that $\tau \in \mathrm{Gal}(M/K)$ has $\varphi(\tau)$ in the above set if and only if $\tau|_L = \sigma$. However, from our earlier discussion, this pre-image is simply $\sigma\,\mathrm{Gal}(M/L_0)$ (where we choose any extension of σ to $\mathrm{Gal}(M/K)$) which is a basis element and therefore open.

- We show that φ^{-1} is continuous. It suffices to show that any basis element of $\mathrm{Gal}(M/K)$ has open pre-image. Well, picking up a basis element $\sigma\,\mathrm{Gal}(M/L_0)$, these are the elements $\tau \in \mathrm{Gal}(M/K)$ such that $\tau|_{L_0} = \sigma|_{L_0}$, so image under φ is

$$\left\{ (\sigma_L)_L \in \prod_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K) : \sigma_{L_2}|_{L_1} = \sigma_{L_1} \text{ for each } L_1 \subseteq L_2 \text{ and } \sigma_L = \sigma|_{L_0} \right\},$$

  which is

$$\prod_{K \subseteq L \subseteq M} S_L \cap \varprojlim_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K),$$

  where

$$S_L = \begin{cases} \{\sigma|_{L_0}\} & L = L_0, \\ \mathrm{Gal}(L/K) & L \neq L_0. \end{cases}$$

  But this is a sub-basis element of the induced topology on the inverse limit, so we are done now. ∎

Remark 5.277. Remark 5.275 is essentially why we would expect this in advance. Alternatively, we have more or less endowed $\varprojlim_L \mathrm{Gal}(L/K)$ with the coarsest topology such that the projection maps to each $\mathrm{Gal}(L/K)$ are continuous, which is precisely the Krull topology.

Namely, we see that $U \subseteq \mathrm{Gal}(M/L)$ if and only if its image in $\varprojlim_L \mathrm{Gal}(L/K)$ is open, so we get the following third definition of the Krull topology.

Definition 5.278 (Krull topology, III)**.** Fix M/K a Galois extension. Then we define the *Krull topology* on $\mathrm{Gal}(M/K)$ as being induced by the product topology under the embedding

$$\mathrm{Gal}(M/L) \cong \varprojlim_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K) \subseteq \prod_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K).$$

5.7.10 Fun with Topology

Let's actually do some fun things with our topology.

Proposition 5.279. Fix M/K a Galois extension. Then $\mathrm{Gal}(M/K)$ is Hausdorff under the Krull topology.

Proof. This is a matter of unwinding the definitions. Fix distinct automorphisms $\sigma_1, \sigma_2 \in \mathrm{Gal}(M/K)$. We need to find disjoint open sets U_1, U_2 such that $\sigma_1 \in U_1$ and $\sigma_2 \in U_2$.

Well, $\sigma_1 \neq \sigma_2$ implies that there exists $\alpha \in M$ such that $\sigma_1\alpha \neq \sigma_2\alpha$. In particular, $\sigma_1|_{K(\alpha)} \neq \sigma_2|_{K(\alpha)}$, so sets U_1 and U_2 defined by

$$U_\bullet \coloneqq \left\{\sigma \in \mathrm{Gal}(M/L) : \sigma|_{K(\alpha)} = \sigma_\bullet|_{K(\alpha)}\right\} = \sigma_\bullet \mathrm{Gal}(M/K(\alpha))$$

are open (in fact, basis elements in our second basis of the Krull topology), disjoint by the nature of restriction, and $\sigma_\bullet \in U_\bullet$. This finishes. ∎

It is also true that $\mathrm{Gal}(M/K)$ will be compact. The easiest way to show this is by embedding into a product.

Proposition 5.280. Fix M/K a Galois extension. The inverse limit $\varprojlim_L \mathrm{Gal}(L/K)$ is closed in the product $\prod_L \mathrm{Gal}(L/K)$.

Proof. We show that the complement of $\varprojlim_L \mathrm{Gal}(L/K)$ is open. For this, it suffices to choose any $(\sigma_L)_L \notin \varprojlim_L \mathrm{Gal}(L/K)$ and find an open set of $\prod_L \mathrm{Gal}(L/K)$ disjoint from $\varprojlim_L \mathrm{Gal}(L/K)$.

Well, $(\sigma_L)_L \notin \varprojlim_L \mathrm{Gal}(L/K)$ must have subfields $L_1 \subseteq L_2$ such that $\sigma_{L_2}|_{L_1} \neq \sigma_{L_1}$. In particular, we define $\{S_L\}_L$ by

$$S_L = \begin{cases} \{\sigma_{L_1}\} & L = L_1, \\ \{\sigma_{L_2}\} & L = L_2, \\ \mathrm{Gal}(L/K) & \text{else}, \end{cases}$$

so that $\prod_L S_L$ contains $(\sigma_L)_L$, but any $(\tau_L)_L \in \prod_L S_L$ has $\tau_{L_2}|_{L_1} \neq \tau_{L_1}$ so that $\prod_L S_L$ is disjoint from $\varprojlim_L \mathrm{Gal}(L/K)$. This finishes. ∎

Theorem 5.281. Fix M/K a Galois extension. The group $\mathrm{Gal}(M/K)$ is compact under the Krull topology.

Proof. We note that the image of $\mathrm{Gal}(M/K)$ under the continuous embedding

$$\mathrm{Gal}(M/K) \cong \varprojlim_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K) \subseteq \prod_{K \subseteq L \subseteq M} \mathrm{Gal}(L/K)$$

is closed by the previous proposition. But the space $\prod_L \mathrm{Gal}(L/K)$ is the product of the compact (finite discrete) spaces $\mathrm{Gal}(L/K)$ and hence compact by Tychonoff's theorem. So it suffices that a closed subset of a compact subset is compact,[17] so we are done. ∎

Remark 5.282. In fact, it is in general true that any profinite group (i.e., inverse limit of finite groups) will be compact under some induced topology, roughly using a proof similar to the one above.

Here is some other magic that our topology can do.

Proposition 5.283. Fix M/K a Galois extension. Then, given $\tau \in \mathrm{Gal}(M/K)$, the maps $x \mapsto \tau x$ and $x \mapsto x\tau$ are both continuous.

Proof. It suffices to show that the pre-images of a basis element $\sigma \,\mathrm{Gal}(M/L)$ will be open, where $\sigma \in \mathrm{Gal}(M/K)$ and L is some finite Galois subextension of M/K. Well, the pre-image under $x \mapsto \tau x$ is

$$\tau^{-1}\sigma \,\mathrm{Gal}(M/L),$$

which is a basis element and hence open. Similarly, the pre-image under $x \mapsto x\tau$ is

$$\sigma \,\mathrm{Gal}(M/L)\tau^{-1} = \sigma\tau^{-1} \cdot \tau \,\mathrm{Gal}(M/L)\tau^{-1} = \sigma\tau^{-1} \,\mathrm{Gal}(M/\tau L),$$

where we are using Lemma 5.108. Again, we see that we have hit a basis element and so are done. ∎

Proposition 5.284. Fix M/K a Galois extension, and set $G \coloneqq \mathrm{Gal}(M/L)$. Then a subgroup $U \subseteq G$ is open if and only if it is closed and has finite index.

Proof. We check the directions one at a time.

(a) Fix $U \subseteq G$ an open subgroup. We show that U is closed and has finite index separately.

- We show that U is closed. We see that, given $\sigma \in \mathrm{Gal}(M/K)$, σU is the pre-image of U under the map $x \mapsto \sigma^{-1}x$, so each σU will also be open. Namely, all cosets in G/U are open, so the complement of U is
$$G \setminus U = \bigcup_{\sigma \notin U} \sigma U,$$
which is also open, so U is closed.
- We show that U has finite index; this is by compactness. Namely, as above, we note that G/U will provide a cover for G, and each $\sigma U \in G/U$ will be open. So G/U is an open cover of G and hence has a finite subcover $\{\sigma_k U\}_{k=1}^n$ by compactness (!).
We finish by claiming $G/U = \{\sigma_k U\}_{k=1}^n$, which will imply that G/U is finite. Indeed, certainly $\{\sigma_k U\}_{k=1}^n \subseteq G/U$. Conversely, for any $\sigma U \in G/U$, we see that σ belongs to some open set in our finite subcover, so say
$$\sigma \in \sigma_k U.$$
Then $\sigma U = \sigma_k U \in \{\sigma_k U\}_{k=1}^n$, finishing.

[17] If V is closed in the compact space X, then any open cover of V can be extended to an open cover of X by adding $X \setminus V$, which can then be refined to a finite subcover and restricted to be a finite subcover of V.

(b) Fix $U \subseteq G$ a closed subgroup of finite index. As before, given $\sigma \in \mathrm{Gal}(M/K)$, $G \setminus \sigma U$ is the pre-image of $G \setminus U$ under the map $x \mapsto \sigma^{-1}x$, so $G \setminus \sigma U$ will also be open.

Now, choosing coset representatives $\{\sigma_1, \dots, \sigma_n\}$ for G/U such that $\sigma_1 U = U$, we see that

$$U = G \setminus \bigcup_{k=2}^{n} \sigma_k U = \bigcap_{k=2}^{n} G \setminus \sigma_k U$$

is the finite intersection of open sets and therefore is open. This finishes. ∎

Remark 5.285. The above also holds, in a weaker form, for general topological groups (i.e., not necessarily compact). Explicitly, in the absence of compactness, the above proof shows that open implies closed and that closed and finite index implies open.

To finish our discussion of the Krull topology, we should actually show that it does salvage the Galois correspondence.

Theorem 5.286 (Galois correspondence)**.** Fix M/K a Galois extension.

(a) Given an intermediate extension F of M/K, the subgroup $\mathrm{Gal}(M/F)$ is closed in $\mathrm{Gal}(M/K)$.

(b) Fix a subgroup $H \subseteq \mathrm{Gal}(M/K)$. Then $\mathrm{Gal}\left(M/M^H\right)$ is the topological closure of H.

(c) In particular, if H is a closed subgroup, then $H = \mathrm{Gal}\left(M/M^H\right)$.

Proof. We take these one at a time.

(a) There is a way to do this by mostly doing topological group theory, imitating Proposition 5.280. Instead, we claim that

$$\mathrm{Gal}(M/F) \overset{?}{=} \bigcap_{L \subseteq F} \mathrm{Gal}(M/L),$$

where the intersection is taken over finite Galois subextensions L of M/K contained in F. Indeed, certainly $\sigma \in \mathrm{Gal}(M/F)$ implies that σ fixes L and hence each L.

And conversely, for any σ in the intersection and $\alpha \in F$, we note that we can place $K(\alpha)$ in a finite Galois extension L of M/K so that σ must fix L and hence fix α. So the equality follows.

But now we see that each $\mathrm{Gal}(M/L)$ is an open subgroup of $\mathrm{Gal}(M/K)$, so it follows that each of these are closed as well. So $\mathrm{Gal}(M/F)$ is some large intersection of closed sets, so $\mathrm{Gal}(M/F)$ is also a closed set.

(b) Let V be the topological closure of H, and we claim $V = \mathrm{Gal}\left(M/M^H\right)$. Note that $\mathrm{Gal}\left(M/M^H\right)$ is a closed set by (a), so it follows that $V \subseteq \mathrm{Gal}\left(M/M^H\right)$.

In the other direction, we note that V consists of H and its limit points, so it suffices to show that each point in $\mathrm{Gal}\left(M/M^H\right)$ is either in H or a limit point. Well, pick up $\tau \in \mathrm{Gal}\left(M/M^H\right)$; if $\tau \in H$, we are done, and otherwise we may take $\tau \notin L$.

We need to show that τ is a limit point of H. It suffices to look at basis elements. Namely, for each basis element $\sigma\,\mathrm{Gal}(M/L)$ containing τ (where $\sigma \in \mathrm{Gal}(M/K)$ and L is a finite Galois subextension of M/K) so that $\sigma\,\mathrm{Gal}(M/L) = \tau\,\mathrm{Gal}(M/L)$ (by group theory), we claim

$$\tau\,\mathrm{Gal}(M/L) \cap (H \setminus \tau) \overset{?}{\neq} \varnothing.$$

Because $\tau \notin H$, it really suffices to show that $\tau\,\mathrm{Gal}(M/L) \cap H \neq \varnothing$.

We are now ready to do Galois theory; the main idea is to reduce to the finite Galois case, where we already have control. Note that τ fixes M^H implies that τ fixes

$$M^H \cap L = \{\alpha \in L : h\alpha = h \text{ for each } h \in H\} = \{\alpha \in L : h|_L \alpha = \alpha \text{ for each } h \in H\} = L^{H|_L},$$

where $H|_L \subseteq \mathrm{Gal}(L/K)$ is H restricted to L. In particular, τ fixing $L^{H|_L}$ implies that $L^{H|_L} \subseteq L^{\langle \tau|_L \rangle}$, so by the Galois correspondence, we see

$$\langle \tau|_L \rangle \subseteq H|_L.$$

In particular, there exists $h \in H$ such that $h|_L = \tau|_L$. But this is exactly what we need to witness $h \in \tau \,\mathrm{Gal}(M/L)$. So we are done.

(c) Because H is a subgroup, $\mathrm{Gal}\left(M/M^H\right)$ will be the topological closure of H. But H is closed, so this topological closure is simply H. ■

5.7.11 Infinite Galois Extensions: Examples

Anyways, let's do some examples. We start with finite fields.

Exercise 5.287. Fix $\mathbb{F}_q$ a finite field. We show that $\mathbb{F}_q \subseteq \overline{\mathbb{F}_q}$ is an infinite extension with Galois group isomorphic to $\widehat{\mathbb{Z}} \cong \prod_p \mathbb{Z}_p$.

Proof. We already know that finite extensions of $\mathbb{F}_q$ are the spitting field of $X^{q^n} - X \in \mathbb{F}_q[X]$ for some $n \in \mathbb{N}$, which is a separable polynomial, so each finite extension of $\mathbb{F}_q$ is a Galois extension. So each $\alpha \in \overline{\mathbb{F}_q}$ lives in a finite extension $\mathbb{F}_q(\alpha)$, which is separable, so α is separable over $\mathbb{F}_p$. Further, each $f(X) \in \mathbb{F}_q[X]$ will split in $\overline{\mathbb{F}_q}$, so any f with a root in $\overline{\mathbb{F}_q}$ will fully split.

So $\overline{\mathbb{F}_q}/\mathbb{F}_q$ is normal and separable and hence Galois, and we see that

$$\mathrm{Gal}\left(\overline{\mathbb{F}_q}/\mathbb{F}_q\right) \cong \varprojlim \mathrm{Gal}\left(\mathbb{F}_{q^n}/\mathbb{F}_q\right) \cong \varprojlim \mathbb{Z}/n\mathbb{Z},$$

where $\mathrm{Gal}\left(\mathbb{F}_{q^n}/\mathbb{F}_q\right)$ is cyclic generated by the Frobenius $x \mapsto x^q$. Technically we have to track what the maps $\mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ in the above inverse limit are, so we do so quickly. Indeed, $\varprojlim \mathrm{Gal}\left(\mathbb{F}_{q^n}/\mathbb{F}_q\right)$ only has the restriction maps

$$\mathrm{Gal}\left(\mathbb{F}_{q^n}/\mathbb{F}_q\right) \to \mathrm{Gal}\left(\mathbb{F}_{q^m}/\mathbb{F}_q\right),$$

which only exist when $\mathbb{F}_{q^m} \subseteq \mathbb{F}_{q^n}$, which is equivalent to $m \mid n$. And when $m \mid n$, the restriction map above will have to take the generator $x \mapsto x^q$ of $\mathrm{Gal}\left(\mathbb{F}_{q^n}/\mathbb{F}_q\right)$ to its restriction $x \mapsto x^q$ on $\mathrm{Gal}\left(\mathbb{F}_{q^m}/\mathbb{F}_q\right)$. Namely, tracking generators shows that the following diagram commutes, for each $m \mid n$.

$$\begin{array}{ccc} \mathrm{Gal}\left(\mathbb{F}_{p^n}/\mathbb{F}_p\right) & \longrightarrow & \mathrm{Gal}\left(\mathbb{F}_{p^m}/\mathbb{F}_p\right) \\ \downarrow & & \downarrow \\ \mathbb{Z}/n\mathbb{Z} & \longrightarrow & \mathbb{Z}/m\mathbb{Z} \end{array}$$

So we do get that

$$\varprojlim \mathrm{Gal}\left(\mathbb{F}_{q^n}/\mathbb{F}_q\right) \cong \varprojlim \mathbb{Z}/n\mathbb{Z} \cong \widehat{\mathbb{Z}}.$$

Lastly, we note that $\widehat{\mathbb{Z}} \cong \prod_p \mathbb{Z}_p$ was shown on Exercise 3.154, finishing. ■

We can also get reasonable control of abelian extensions of $\mathbb{Q}$ using the Kronecker–Weber theorem.

Exercise 5.288. Let $\mathbb{Q}^{\mathrm{ab}}$ be the maximal abelian extension of $\mathbb{Q}$. Assuming the Kronecker–Weber theorem, we show that $\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}$ is a Galois extension with Galois group $\widehat{\mathbb{Z}}^\times \cong \prod_p \mathbb{Z}_p^\times$.

Proof. Note that

$$\mathrm{Gal}\left(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}\right) = \bigcup_{\mathrm{Gal}(K/\mathbb{Q}) \text{ abel.}} K,$$

which is well-defined as a field essentially because the composite of two abelian extensions is another abelian extension[18] so that any $\alpha, \beta \in \mathrm{Gal}\left(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}\right)$ can be placed in some abelian extension $KL/\mathbb{Q}$ where $\alpha \in K$ and $\beta \in L$, giving closure $\mathbb{Q}^{\mathrm{ab}}$ under addition and multiplication.

To show that $\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}$ is a Galois extension, we note that it is separable because every element of $\mathbb{Q}^{\mathrm{ab}}$ comes from a separable extension and hence is separable. Further, each polynomial with a root in $\mathbb{Q}^{\mathrm{ab}}$ has a root in a abelian extension $K/\mathbb{Q}$ and hence will fully split in K and therefore in $\mathbb{Q}^{\mathrm{ab}}$. So $\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}$ is indeed Galois.

It remains to compute the Galois group. Well, by the Kronecker–Weber theorem, each abelian extension $K/\mathbb{Q}$ can be contained in a cyclotomic extension $\mathbb{Q}(\zeta)/\mathbb{Q}$, and each cyclotomic extension is abelian, so it suffices to only focus on cyclotomic extensions. Rigorously, we can start with

$$\mathrm{Gal}\left(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}\right) \cong \left\{ (\sigma_K)_K \in \prod_{K/\mathbb{Q} \text{ abel.}} \mathrm{Gal}(K/\mathbb{Q}) : \sigma_{K_2}|_{K_1} = \sigma_{K_1} \text{ for each } K_1 \subseteq K_2 \right\},$$

but placing each K inside of a cyclotomic extension means that we can fully determine $(\sigma_K)_K$ by the action on on various cyclotomic fields. So we see that

$$\mathrm{Gal}\left(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}\right) \cong \left\{ (\sigma_n)_n \in \prod_{n=1}^{\infty} \mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) : \sigma_n|_m = \sigma_m \text{ for each } \mathbb{Q}(\zeta_m) \subseteq \mathbb{Q}(\zeta_n) \right\} \cong \varprojlim \mathrm{Gal}\left(\mathbb{Q}(\zeta_n)/\mathbb{Q}\right).$$

Quickly, we note that the maps in the inverse limit are restrictions

$$\mathrm{Ga}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \to \mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q}),$$

which exist if and only if $\mathbb{Q}(\zeta_m) \subseteq \mathbb{Q}(\zeta_n)$ if and only if $m \mid n$. Further, we note that $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \cong (\mathbb{Z}/n\mathbb{Z})^\times$, and this isomorphism makes the following diagram of isomorphisms commute.

$$\begin{array}{ccc} \mathrm{Gal}\left(\mathbb{Q}(\zeta_n)/\mathbb{Q}\right) & \longrightarrow & \mathrm{Gal}\left(\mathbb{Q}(\zeta_m)/\mathbb{Q}\right) \\ \downarrow & & \downarrow \\ (\mathbb{Z}/n\mathbb{Z})^\times & \longrightarrow & (\mathbb{Z}/m\mathbb{Z})^\times \end{array}$$

Indeed, we can track $\sigma_k \in \mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ defined by $\sigma_k : \zeta_n \mapsto \zeta_n^k$ through the diagram as follows.

$$\begin{array}{ccc} \zeta_n \mapsto \zeta_n^k & \longmapsto & \zeta_m \mapsto \zeta_m^k \\ \downarrow & & \downarrow \\ k \pmod n & \longmapsto & k \pmod m \end{array}$$

Anyways, the point is that

$$\mathrm{Gal}\left(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}\right) \cong \varprojlim (\mathbb{Z}/n\mathbb{Z})^\times \cong \left(\varprojlim \mathbb{Z}/n\mathbb{Z}\right)^\times = \widehat{\mathbb{Z}}^\times$$

by tracking the isomorphisms through. Technically we should check that $R \mapsto R^\times$ preserves limits for rings R, but this is because $R \mapsto R^\times$ is right adjoint to the group ring functor $G \mapsto \mathbb{Z}[G]$ and hence preserves limits. Anyways, we again see that this is

$$\widehat{\mathbb{Z}}^\times \cong \prod_p \mathbb{Z}_p^\times,$$

where taking the multiplicative group is still safe because it preserves limits (and in particular, products). ■

[18] If $K/\mathbb{Q}$ and $L/\mathbb{Q}$ are Galois, then $\mathrm{Gal}(KL/\mathbb{Q}) \to \mathrm{Gal}(K/\mathbb{Q}) \times \mathrm{Gal}(L/\mathbb{Q})$ is injective, so if $K/\mathbb{Q}$ and $L/\mathbb{Q}$ are abelian, $KL/\mathbb{Q}$ will also be abelian.

It might not be immediately obvious, but having control over the absolute abelian Galois group of $\mathbb{Q}$ is quite useful. To make our presentation more malleable, we will assert but not prove that

$$\mathbb{Z}_p^\times \cong \begin{cases} \{\pm 1\} \times \mathbb{Z}_2 & p = 2, \\ (\mathbb{Z}/(p-1)\mathbb{Z}) \times \mathbb{Z}_p & p \text{ odd}. \end{cases}$$

Essentially these are true because we can build an exponential map $\exp : p\mathbb{Z}_p \to 1 + p\mathbb{Z}_p$ (here, $1 + p\mathbb{Z}_p \subseteq \mathbb{Z}_p^\times$) by

$$\exp(z) := \sum_{k=0}^{\infty} \frac{z^k}{k!}$$

and show by hand that $\exp$ is an injective homomorphism. Anyways, we see

$$\operatorname{Gal}\left(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}\right) \cong \{\pm 1\} \times \mathbb{Z}_2 \times \prod_{p \text{ odd}} \left(\mathbb{Z}/(p-1)\mathbb{Z} \times \mathbb{Z}_p\right).$$

At this point, we understand the group on the right pretty well, and this can translate into some cute results.

Exercise 5.289. We sketch why that there is exactly one chain of Galois extensions $\mathbb{Q} \subseteq L_0 \subseteq L_1 \subseteq \cdots$ such that $\operatorname{Gal}(L_n/\mathbb{Q}) \cong \mathbb{Z}/p^n\mathbb{Z}$ for each n.

Proof. Essentially, the infinite chain is equivalent to asking for an infinite extension $L/\mathbb{Q}$ such that $L := \bigcup_{n\in\mathbb{N}} L_n = \varinjlim L$ such that

$$\operatorname{Gal}(L/\mathbb{Q}) \cong \varprojlim \operatorname{Gal}(L_n/\mathbb{Q}) \cong \varprojlim \mathbb{Z}/p^n\mathbb{Z} \cong \mathbb{Z}_p.$$

Noting that $L/\mathbb{Q}$ is now an abelian extension, we see that we are looking for a surjective, continuous group (restriction) homomorphism

$$\{\pm 1\} \times \mathbb{Z}_2 \times \prod_{p \text{ odd}} \left(\mathbb{Z}/(p-1)\mathbb{Z} \times \mathbb{Z}_p\right) \cong \operatorname{Gal}\left(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}\right) \twoheadrightarrow \operatorname{Gal}(L/\mathbb{Q}) \cong \mathbb{Z}_p.$$

(Certainly $L/\mathbb{Q}$ will induce this map. Conversely, for any continuous group homomorphism, we see that the kernel will be a closed subgroup and hence will induce the desired infinite extension.) But it is not too hard to believe that there is exactly one such continuous surjection.[19] ∎

Exercise 5.290. We sketch why there is no extension $L/\mathbb{Q}$ with Galois group $\mathbb{Z}_2 \times \mathbb{Z}_2$.

Proof. As before, such an extension $L/\mathbb{Q}$ will induce a surjective, continuous group (restriction) homomorphism

$$\{\pm 1\} \times \mathbb{Z}_2 \times \prod_{p \text{ odd}} \left(\mathbb{Z}/(p-1)\mathbb{Z} \times \mathbb{Z}_p\right) \cong \operatorname{Gal}\left(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}\right) \twoheadrightarrow \operatorname{Gal}(L/\mathbb{Q}) \cong \mathbb{Z}_2 \times \mathbb{Z}_2.$$

However, it is again not too hard to believe that no such thing exists because the left-hand side only has one copy of $\mathbb{Z}_2$. ∎

This last result is quite strange because we have just shown that the inverse Galois problem fails if we push to infinite extensions and ask about all profinite extensions. The point here is to show how delicate the inverse Galois problem is.

[19] In particular, one can show that, when $p \neq q$ are primes, the only continuous group homomorphism $\mathbb{Z}_p \to \mathbb{Z}_q$ is the trivial one, essentially by tracking where the dense set $\mathbb{Z} \subseteq \mathbb{Z}_p$ goes.

List of Definitions