MATH 4581
FALL 2025 LECTURE NOTES

RICHARD HABURCAK

Contents

Forward

These are lecture notes from a Fall semester 2025 course of Math 4581, Undergraduate Abstract Algebra 2, at The Ohio State University. This course is a continuation of Math 4580, Undergraduate Abstract Algebra 1.

Acknowledgments. These notes are heavily based on notes by Mike Lipnowski of a course in Spring 2025 at The Ohio State University which were graciously given to me when I was preparing for the course, as well as notes taken by Jeff Hein of a course by David Webb in Fall 2011 at Dartmouth College. I also thank the students in my section who brought up typos or suggestions during the semester.

Remarks. When possible, I tried to make these notes self-contained, assuming a standard course in group theory up and ring theory through the first isomorphism theorem and some examples. I will sometimes write "recall", which should be understood as "if this is familiar, move on; and if not, then I suggest you look it up before proceeding". But don't let it stop you from reading through the notes! Black-boxing is an important skill to develop, though a dangerous one. There may be typos or mistakes scattered throughout these notes, though I hope none are completely derailing. If you notice any mistakes, please let me know. And if you found these notes useful, I would also appreciate any feedback.

1. Introduction

1.1. Groups.

Recall some examples of groups.

Example 1.1.1. We recall some familiar and some new examples of groups.

(1) The group of integers under addition $(\mathbb{Z}, +)$.
(2) The group of integers modulo n under addition $(\mathbb{Z}/n\mathbb{Z}, +)$. For example $\mathbb{Z}/3\mathbb{Z} = \{0, 1, 2\}$.
(3) The group of integers modulo n that have a multiplicative inverse, $(\mathbb{Z}/n\mathbb{Z}^*, \cdot)$, i.e. the integers moduli n that are coprime to n. For example $\mathbb{Z}/3\mathbb{Z}^* = \{1, 2\} \cong \mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/6\mathbb{Z}^* = (\{1, 5\}, \cdot) \cong \mathbb{Z}/2\mathbb{Z}$.
(4) The group of permutations of $\{1, 2, \dots, n\}$, the permutation group S_n. Recall that we can write elements of S_n using cycle notation. For example

$$S_3 = \{(1), (12), (13), (23), (123), (132)\},$$

where the permutation (123) is the permutation

$$\begin{aligned} 1 &\mapsto 2 \\ 2 &\mapsto 3 \\ 3 &\mapsto 1 \end{aligned}$$

cyclically permuting the entries. Recall also that $|S_n| = n!$.

(5) D_n the symmetry group of the regular n-gon. For example D_4 is the symmetry group of the square. There is the identity e, rotation by $90°$ which is denoted by r, and its powers r^2, r^3 and $r^4 = e$, as well as reflections s_h, s_v through the horizontal and vertical axes, and reflections across the diagonals s_d and $s_{d'}$, as shown in Figure 1.

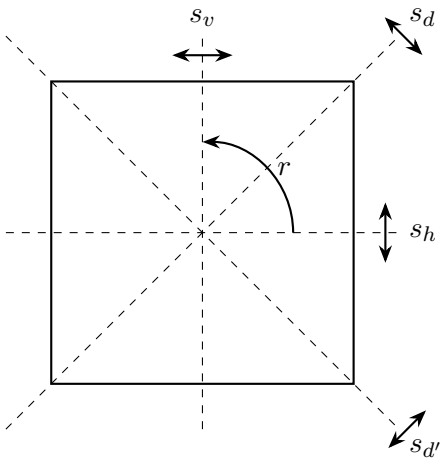


FIGURE 1. How elements of D_4 act on the square

Labeling the vertices you can find relations among these symmetries. For example, $(s_v)^2 = e$.

(6) The group μ_n of n^{th} roots of unity under multiplication, as a subset of the complex numbers. Note that $\mu_n \cong \mathbb{Z}/n\mathbb{Z}$. For example, μ_3 is illustrated in Figure 2.

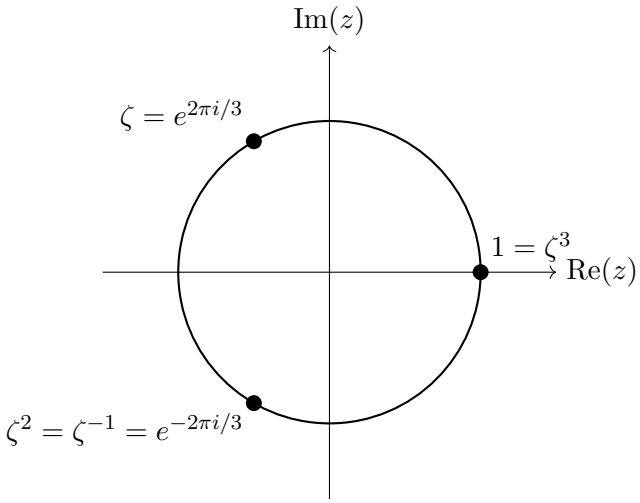


FIGURE 2. The third roots of unity

(7) The group $\mathbb{R}^*$ of non-zero real numbers under multiplication.
(8) The additive structure on any ring R gives a group $(R, +)$.
(9) For any ring R, the invertible elements with respect to multiplication gives a group R^*.
(10) For a vector space V (think $\mathbb{R}^n$), the group

$$\mathrm{GL}(V) \coloneqq \{T : V \to V \mid T \text{ an invertible linear map}\}$$

of invertible linear maps $V \to V$ is a group, called the *general linear group*.

This can be made very concrete if we choose a basis. We'll write

$$\mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^n) \coloneqq \{\text{linear maps } \mathbb{R}^n \to \mathbb{R}^n\},$$

and

$$M_{n\times n}(\mathbb{R}) \coloneqq \{n \times n \text{ matrices with real entries}\}.$$

After choosing a basis, $\mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^n) \cong M_{n\times n}(\mathbb{R})$. We know that for matrices the invertibility is measured by the determinant, which is a map

$$M_{n\times n}(\mathbb{R}) \xrightarrow{\det} \mathbb{R}.$$

The invertible real matrices, $\mathrm{GL}_n(\mathbb{R})$ are those that have non-zero determinant, hence we have a diagram

$$\begin{array}{ccccc} \mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^n) & \cong & M_{n\times n}(\mathbb{R}) & \xrightarrow{\det} & \mathbb{R} \\ \cup & & \cup & & \cup \\ \mathrm{GL}(\mathbb{R}^n) & \cong & \mathrm{GL}_n(\mathbb{R}) & \xrightarrow[\det]{} & \mathbb{R}^* \end{array}$$

showing that $\mathrm{GL}_n(\mathbb{R})$ are those $n \times n$ matrices whose determinant is in $\mathbb{R}^*$. $\mathrm{GL}_n(\mathbb{R})$ is a group under standard matrix multiplication, and is the symmetry group of the vector space $\mathbb{R}^n$ that preserves the linear structure. Moreover, the map

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

is a group homomorphism as for A, B two invertible $n \times n$ matrices,

$$\det(AB) = \det(A) \cdot \det(B).$$

One way to understand groups is by understanding how groups act as symmetries on certain objects, like polygons, vector spaces, etc., and the groups preserve some structure.

Example 1.1.2 (Preserving volume in $\mathbb{R}^n$)**.** Recall that for a matrix A, $\det(A)$ measures how A changes (signed) volume. And matrices with determinant 1 preserve the volume. The *special linear group* is

$$\mathrm{SL}_n(\mathbb{R}) \coloneqq \{A \in \mathrm{GL}_n(\mathbb{R}) \mid \det(A) = 1\},$$

the $n \times n$ matrices with determinant 1. How do we know this is a group? We've just seen that $\det : \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^*$ is a group homomorphism.

Recall 1.1.3. Recall that the kernel of a group homomorphism is a subgroup of the domain.

Proposition 1.1.4. *Let $\varphi : G \to H$ be a group homomorphism. Then*

$$\ker \varphi \coloneqq \{g \in G \mid \varphi(g) = \mathrm{id}_H\} \leq G$$

is a subgroup of G.

Thus $\ker \det = \{A \in \mathrm{GL}_n(\mathbb{R}) \mid \det(A) = 1\} = \mathrm{SL}_n(\mathbb{R})$ is a subgroup of $\mathrm{GL}_n(\mathbb{R})$.
So we could say that $\mathrm{SL}_n(\mathbb{R})$ is the symmetry group of $\mathbb{R}^n$ that preserves volume.

2. Euclidean group

A closely related group to the matrix groups we've seen is the group of *rigid motions*, bijections $\mathbb{R}^n \to \mathbb{R}^n$ that preserve distance.

Definition 2.0.1 (Euclidean group, preliminary definition)**.** We define the *Euclidean group* as the set

$$E(n) \coloneqq \{\text{bijections } \mathbb{R}^n \to \mathbb{R}^n \text{ preserving all distances between points}\}.$$

Remark 2.0.2. It is **not** (yet) clear that $E(n)$ is a group! And the definition is quite sloppy. What do we mean by "preserving all distances"? We'll make this more precise shortly.

Our first task will be to give a more precise definition of $E(n)$ and then to give a nice description of the elements.

But first, let's see some examples.

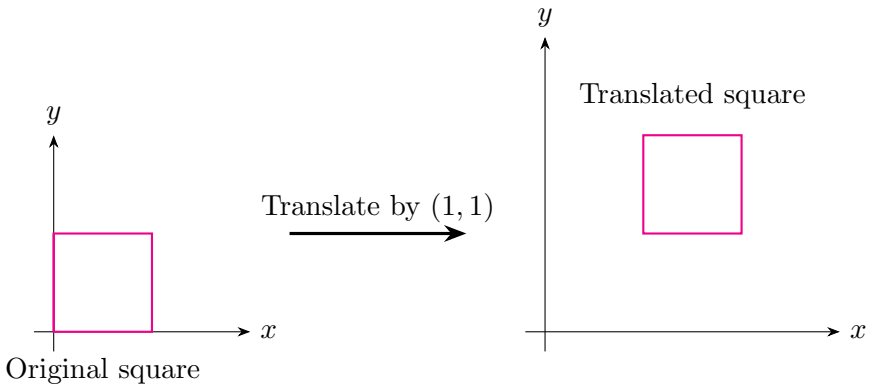


FIGURE 3. Translating by a vector $x \in \mathbb{R}^2$ is a rigid motion

FIGURE 4. Reflecting about a line ℓ is a rigid motion

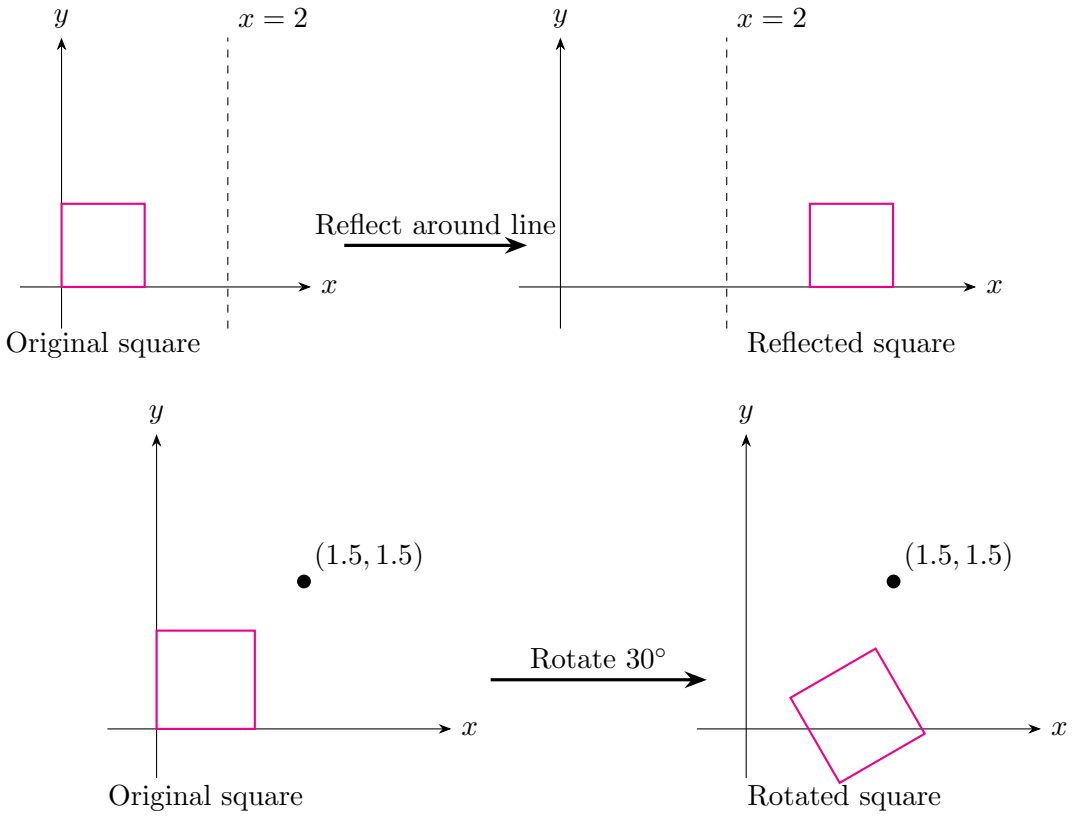


FIGURE 5. Rotating about a point is a rigid motion

2.1. **Distance in $\mathbb{R}^n$.** The way we'll define distance in $\mathbb{R}^n$ is by using the dot product.

Definition 2.1.1 (Dot Product)**.** Let $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^n$. Define the *dot product* of x and y by

$$x \cdot y \coloneqq x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

Equivalently, viewing x as a column vector

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

we have

$$x \cdot y = x^t y = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n} x_i y_i.$$

We will also use the notation

$$\langle x, y \rangle \coloneqq x \cdot y.$$

Using the dot product, and the Pythagorean Theorem, we can define a notion of length of vectors.

Definition 2.1.2 (Length in $\mathbb{R}^n$)**.** The *length* of $x \in \mathbb{R}^n$ is

$$||x|| \coloneqq \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Finally, the distance between two points is just the length of the vector between them.

Definition 2.1.3 (Distance in $\mathbb{R}^n$)**.** The *distance* between $x, y \in \mathbb{R}^n$ is

$$d(x, y) \coloneqq ||x - y||.$$

Remark 2.1.4. For those who have taken a real analysis class, you can check that this defines a "metric" on $\mathbb{R}^n$.

Given this notion of distance, we can now give a precise definition of $E(n)$.

Definition 2.1.5 (Euclidean group)**.** We define the *Euclidean group* as the set

$$E(n) \coloneqq \{\text{bijections } f : \mathbb{R}^n \to \mathbb{R}^n \mid d(f(x), f(y)) = d(x, y) \text{ for all } x, y \in \mathbb{R}^n\}.$$

We're now ready to prove that $E(n)$ is a group.

Proposition 2.1.6. *The set $E(n)$ is a group with group operation composition of maps.*

Proof. The identity map $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ is clearly a bijection, and

$$d(\mathrm{id}(x), \mathrm{id}(y)) = d(x, y).$$

Thus $\mathrm{id} \in E(n)$.

Function composition is associative, so the group operation is associative.

Suppose now that $f, g \in E(n)$. We need to show that $f \circ g \in E(n)$ and there exists a distance preserving bijection $f^{-1} \in E(n)$ such that $f \circ f^{-1} = \mathrm{id} = f^{-1} \circ f$.

If $f, g \in E(n)$, then clearly $f \circ g$ is a bijection and we compute

$$d(f(g(x)), f(g(y))) = d(g(x), g(y)) = d(x, y),$$

thus $f \circ g \in E(n)$.

If $f \in E(n)$, then as $f : \mathbb{R}^n \to \mathbb{R}^n$ is a bijection, there is an inverse bijection $f^{-1} : \mathbb{R}^n \to \mathbb{R}^n$. We now check that f^{-1} preserves distances. Since

$$d(f(x), f(y)) = d(x, y)$$

for any $x, y \in \mathbb{R}^n$, plugging in $f^{-1}(x)$ and $f^{-1}(y)$ gives

$$d(x,y) = d\left(f(f^{-1}(x)), f(f^{-1}(y))\right) = d\left(f^{-1}(x), f^{-1}(y)\right).$$

Thus f^{-1} preserves distances as well, hence $f^{-1} \in E(n)$.

Thus we have shown that $E(n)$ is a group. □

Recall 2.1.7. We recall some useful facts about the dot product and distances.

Proposition 2.1.8. *For $x, y, z \in \mathbb{R}^n$ and $a \in \mathbb{R}$, the following properties hold.*

(1) $\langle x, y\rangle = \langle y, x\rangle$
(2) $\langle x, y+z\rangle = \langle x, y\rangle + \langle x, z\rangle$
(3) $\langle ax, y\rangle = a\langle x, y\rangle = \langle x, ay\rangle$
(4) $\langle x, x\rangle \geq 0$ *with equality if and only if* $x = 0$
(5) if $\langle x, y\rangle = 0$ *for all* $x \in \mathbb{R}^n$, *then* $y = 0$
(6) $\langle x, y\rangle = \frac{1}{2}\left(||x+y||^2 - ||x||^2 - ||y||^2\right)$

Proof. Left as an exercise. The last one is on Homework 1. □

We've already seen some invertible maps $\mathbb{R}^n \to \mathbb{R}^n$ that preserve distance (the matrices with determinant one). It turns out that there are much more! Let's find them all!

2.2. **Distance Preserving Linear Maps.** If a matrix A preserves distance, then since the distance is defined in terms of inner products, it makes sense that it would also preserve the inner product. And vice versa.

Recall that for a matrix

$$A = (a_{ij})_{ij} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{n,n} \end{pmatrix} \in M_{n\times n}(\mathbb{R}),$$

the transpose is given by

$$A^t = (a_{ji})_{ij} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{n,n} \end{pmatrix}.$$

Proposition 2.2.1. *Let $A \in M_{n\times n}(\mathbb{R})$. The following are equivalent:*

(1) *The columns of A for an orthonormal set (a subset $S \subset \mathbb{R}^n$ is called* orthonormal *if for $x \in S$, $\langle x, x\rangle = 1$ and for $x \neq y \in S$ we have $\langle x, y\rangle = 0$.)*
(2) $A^{-1} = A^t$
(3) $\langle Ax, Ay\rangle = \langle x, y\rangle$ *(A preserves the dot product)*
(4) $||Ax - Ay|| = ||x - y||$ *(A preserves distances)*
(5) $||Ax|| = ||x||$ *(A preserves length)*

Proof. We will show that (1) $\iff$ (2) $\iff$ (3), and the remaining equivalences are left as an exercise.

(1) $\implies$ (2)**:** Let $A = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}$ with

$$a_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix} \in \mathbb{R}^n.$$

The set $\{a_i, a_2, \dots, a_n\}$ being orthonormal means that that

$$\langle a_i, a_j \rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}.$$

By definition of matrix multiplication, we have

$$A^t A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} = (\langle a_i, a_j \rangle)_{ij}.$$

Since the columns of A are orthonormal, we have

$$A^t A = I_n \coloneqq \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

as required. (2) $\implies$ (1)**:** From the computation above, we see that the entries of $A^t A$ are the dot products of the columns of A. Thus since $A^t A = I_n$, the columns of A satisfy

$$\langle a_i, a_j \rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j, \end{cases},$$

as required.

(2) $\implies$ (3)**:** By definition, we have

$$\langle Ax, Ay \rangle = (Ax)^t Ay.$$

We now compute

$$\begin{aligned} \langle Ax, Ay \rangle &= (Ax)^t Ay \\ &= x^t A^t A y \\ &= x^t (A^t A) y \\ &= x^t I_n y \\ &= x^t y \\ &= \langle x, y \rangle, \end{aligned}$$

Thus A preserves the inner product.

(3) $\implies$ (2)**:** We first note that

$$\langle x, x \rangle = \langle Ax, Ax \rangle = x^t A^t A x = \langle x, A^t A x \rangle.$$

Thus for any $x \in \mathbb{R}^n$ we have

$$\langle x, (A^t A - I_n)x \rangle = \langle x, A^t A x - x \rangle = \langle x, A^t A x \rangle - \langle x, I_n x \rangle = \langle x, x \rangle - \langle x, x \rangle = 0,$$

and by Proposition 2.1.8(5), we have $(A^t A - I_n)x = 0$ for all $x \in \mathbb{R}^n$. So $A^t A - I_n = 0$, and thus $A^t A = I_n$. Similarly, $AA^t = I_n$, and so $A^t = A^{-1}$.

The rest is left to you, the interested reader. □

Appropriately, the matrices satisfying these conditions are called *orthogonal*.

Definition 2.2.2 (Orthogonal Group)**.** The *orthogonal group* is defined as

$$\mathrm{O}(n) \coloneqq \{A \in M_{n\times n}(\mathbb{R}) \mid A^t = A^{-1}\}.$$

These are the matrices preserving the inner product.

Definition 2.2.3 (Special Orthogonal Group)**.** The *special orthogonal group* is defined as

$$\mathrm{SO}(n) \coloneqq \mathrm{O}(n) \cap \mathrm{SL}_n(\mathbb{R}).$$

These are the orthogonal matrices with determinant 1.

Example 2.2.4 (O(2))**.** Let's see what matrices in O(2) look like. Matrices $A \in \mathrm{O}(2)$ are determined by

$$A\begin{pmatrix}1\\0\end{pmatrix} \text{ and } A\begin{pmatrix}0\\1\end{pmatrix}.$$

Say

$$A\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}a\\b\end{pmatrix},$$

then

$$\left\langle \begin{pmatrix}a\\b\end{pmatrix}, \begin{pmatrix}a\\b\end{pmatrix} \right\rangle = \left\langle \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix} \right\rangle = 1,$$

so $a^2 + b^2 = 1$ and $a, b \in [-1, 1] \subset \mathbb{R}$. Since

$$\left\langle A\begin{pmatrix}1\\0\end{pmatrix}, A\begin{pmatrix}0\\1\end{pmatrix} \right\rangle = \left\langle \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \right\rangle = 0,$$

we have

$$A\begin{pmatrix}0\\1\end{pmatrix} = \pm\begin{pmatrix}-b\\a\end{pmatrix},$$

which you are encouraged to work out for yourself.

If $A\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}-b\\a\end{pmatrix}$, then

$$A = \begin{pmatrix}a & -b\\ba & \end{pmatrix},$$

and since $a \in [-1, 1]$ we can write $a = \cos\theta$ for some angle θ and now

$$A = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix},$$

which is rotation by an angle θ around the origin. In this case we can also check that $\det A = 1$.

If $A\begin{pmatrix}0\\1\end{pmatrix} = -\begin{pmatrix}-b\\a\end{pmatrix}$, then

$$A = \begin{pmatrix}a & b\\ b-a & \end{pmatrix} = \begin{pmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{pmatrix} = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix},$$

and we see that A is the product of a rotation matrix and a reflection matrix. Moreover, we can check that $\det A = -1$, and $A^2 = I_2$.

Exercise 2.2.5. One can use this to show that O(2) has "two pieces". That is, show that O(2) has two connected components, namely SO(2) and $\mathrm{SO}(2) \cdot \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}$.

2.3. **Euclidean Group Presentation.** We now return to the Euclidean group, and show that elements of $E(n)$ can be written in a very nice way. The idea is that $\mathrm{O}(n)$ has to play some role, since these are linear bijections that preserve distance.

Lemma 2.3.1. *An element $f \in E(n)$ such that $f(0) = 0$ is linear. In particular, $f \in \mathrm{O}(n)$.*

Proof. We first show that f preserves the inner product. Since f preserves distances,

$$||f(x)|| = ||f(x) - 0|| = ||x - 0|| = ||x||.$$

Thus, using Proposition 2.1.8(6), we obtain

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

and equating the first and last equality gives $\langle f(x), f(y)\rangle = \langle x, y\rangle$, as desired.

We now show that f is linear. To accomplish this, we note that it suffices to show that f is linear on a basis (if this is not immediately clear, write down a full proof). It will be easier to to show this for an orthonormal basis. So let $\{e_1, \dots, e_n\}$ be an orthonormal basis of $\mathbb{R}^n$. We claim that $\{f(e_1), \dots, f(e_n)\}$ is then also an orthonormal basis. Indeed, to show that $\{f(e_1), \dots, f(e_n)\}$ is linearly independent, suppose that $0 = \sum_{i=1}^n a_i f(e_i)$. Then we compute

$$0 = \langle 0, f(e_j)\rangle = \left\langle \sum_{i=1}^n a_i f(e_i), f(e_j) \right\rangle = \sum_{i=1}^n a_i \langle f(e_i), f(e_j)\rangle = \sum_{i=1}^n a_i \langle e_i, e_j\rangle = a_j$$

and thus $a_j = 0$ for all $0 \leq j \leq n$. Hence $\{f(e_1), \dots, f_(e_n)\}$ are n linearly independent vectors in $\mathbb{R}^n$, and this form a basis. And since f preserves the inner product, this basis is also orthonormal!

Exercise 2.3.2. Verify that if $\{e_1, \dots, e_n\}$ is an orthonormal basis then for $x = \sum_{i=1}^n x_i e_i$, we have

$$x = \sum_{i=1}^n \langle x, e_i\rangle e_i.$$

Now to show linearity, let $x = \sum_{i=1}^n x_i e_i \in \mathbb{R}^n$, then since $\{f(e_1), \dots, f(e_n)\}$ is an orthonormal basis, we have

$$f(x) = \sum_{i=1}^n \langle f(x), f(e_i)\rangle f(e_i) = \sum_{i=1}^n \langle x, e_i f(e_i) = \sum_{i=1}^n x_i f(e_i).$$

So if $y = \sum_{i=1}^n y_i e_i$, and $a \in \mathbb{R}$, then

$$f(x + ay) = f\left(\sum_{i=1}^n (x_a + a y_i) e_i\right) = \sum_{i=1}^n (x_i + a y_i) f(e_i) = \sum_{i=1}^n x_i f(e_i) + a \sum_{i=1}^n y_i f(e_i) = f(x) + a f(y),$$

and so f is linear.

In particular, f is a linear map preserving the inner product, and Proposition 2.2.1 shows that $f \in \mathrm{O}(n)$, as claimed. □

Now we just have to understand what happens to 0 under a map $f \in E(n)$. The idea to get a nice presentation is to "force" the origin to be sent to the origin. So if $f(0) = b$, just translate it back.

Lemma 2.3.3. *Every rigid motion $f \in E(n)$ is affine linear, i.e. $f(x) = Ax + b$ where A is linear and $b \in \mathbb{R}^n$. In particular, A is linear and preserves distances hence $A \in \mathrm{O}(n)$.*

Proof. Let $b \in \mathbb{R}^n$. Consider the rigid motion

$$t_b : \mathbb{R}^n \to \mathbb{R}^n, x \mapsto x + b.$$

We see immedaitely that $t_b^{-1} = t_{-b}$. (If you are not convinced that t_b preserves distances, check it yourself).

Let $f \in E(n)$. Then $t_{-f(0)} \circ f \in E(n)$, and

$$t_{-f(0)} \circ f(0) = t_{-f(0)}(f(0)) = f(0) - f(0) = 0,$$

hence by Lemma 2.3.1 $t_{-f(0)} \circ f = A \in \mathrm{O}(n)$. Thus

$$f = t_{f(0)} \circ A, f(x) = t_{f(0)}(Ax) = Ax + f(0).$$

□

Theorem 2.3.4. *The Euclidean group is the group*

$$E(n) = \{t_b \circ A \mid A \in \mathrm{O}(n), b \in \mathbb{R}^n\}$$

with multiplication given by composition.

Another way to write the elements of $E(n)$ is as a pair $(A, b) \in \emptyset(n) \times \mathbb{R}^n$. However, the multiplication is *very* different than "component-wise multiplication". Indeed, the element $(A, b)(A', b')$ corresponds to the rigid motion $(t_b \circ A) \circ (t_{b'} \circ A')$, and the latter is the map

$$(t_b \circ A) \circ (t_{b'} \circ A')(x) = (t_b \circ A)(A'x + b') = AA'x + Ab' + b,$$

thus

$$(A, b)(A', b') = (AA', Ab' + b).$$

Theorem 2.3.5. *The Euclidean group is the group*

$$E(n) = \{(A, b) \mid A \in \mathrm{O}(n), b \in \mathbb{R}^n\}$$

with group operation given by

$$(A, b)(A'b') = (AA', Ab' + b).$$

Exercise 2.3.6. Let $(A, b) \in E(n)$. Show that $(A, b)^{-1} = (A^{-1}, -Ab)$.

Remark 2.3.7. You will see on Homework that $E(n)$ is a *semidirect product* of $\mathrm{O}(n)$ and $\mathbb{R}^n$, and this is why the product structure looks odd.

2.4. **Finite subgroups of $E(2)$.** To become a bit more familiar with the Euclidean group and get some mileage out of our nice presentation of elements of $E(n)$, we will find the finite subgroups of $E(2)$. In fact, we will show that if $G \leq E(2)$ is a finite subgroup, then it is isomorphic to $\mathbb{Z}/n\mathbb{Z}$ or D_n for some n.

Remark 2.4.1. The proof that a finite subgroup of $E(2)$ is isomorphic to $\mathbb{Z}/n\mathbb{Z}$ or D_n in Judson's Abstract Algebra (the course textbook) contains an error. To alleviate this, we need a result about fixed points of affine actions.

Lemma 2.4.2. *Let $x_1, \ldots, x_d \in \mathbb{R}^n$, and $f \in E(n)$. Then*

$$f\left(\frac{1}{d}\sum_{i=1}^{d} x_i\right) = \frac{1}{d}\sum_{i=1}^{d} f(x_i).$$

Proof. Let $f = (A, b) \in E(n)$. We compute

$$(A,b)\left(\frac{1}{d}\sum_{i=1}^{d} x_i\right) = A\left(\frac{1}{d}\sum_{i=1}^{d} x_i\right) + b$$
$$= \frac{1}{d}\left(\sum_{i=1}^{d} Ax_i\right) + b.$$

We also compute the other side

$$\frac{1}{d}\left(\sum_{i=1}^{d}(A,b)x_i\right) = \frac{1}{d}\left(\sum_{i=1}^{d} Ax_i + b\right)$$
$$= \frac{1}{d}\left(\sum_{i=1}^{d} Ax_i\right) + \frac{1}{d}db$$
$$= \frac{1}{d}\left(\sum_{i=1}^{d} Ax_i\right) + b,$$

and the claim follows. □

Proposition 2.4.3. *Let $G \leq E(n)$ be a finite subgroup. Then G is conjugate to a subgroup of $\mathrm{O}(n)$.*

Proof. Let $x \in \mathbb{R}^n$. By Lemma 2.4.2, any element of G fixes $y := \frac{1}{|G|}\sum_{g\in G} gx$. Indeed, let $h \in G$, then Lemma 2.4.2 shows that

$$hy = \frac{1}{|G|}\sum_{g\in G} hgx = \frac{1}{|G|}\sum_{g'\in G} g'x = y,$$

simply by relabeling $hg = g'$. Thus for all $f \in t_{-y}Gt_y$, f fixes the origin. Hence $f \in \mathrm{O}(n)$. Thus $t_{-y}Gt_y \leq \mathrm{O}(n)$. □

Example 2.4.4 (Finite Subgroups $G \leq \mathrm{O}(2)$)**.** We've seen that $\mathrm{O}(2)$ has two components, given by $\det = \pm 1$. We can divide the subgroups into two kinds.

(1) Every element $g \in G$ has $\det g = 1$. As in Example 2.2.4, every $g \in G$ is given by a rotation matrix R_θ. Since G is finite, there is a smallest angle θ_0 and we claim that R_{θ_0} generates G. Note that $R_\theta^k = R_{k\theta}$. If G is not generated by R_{θ_0}, then for some k, there is an angle θ_1 between $k\theta_0$ and $(k+1)\theta_0$, but then $(k+1)\theta_0 - \theta_1$ is a smaller angle, which is a contradiction. So $G = \langle R_{\theta_0}\rangle \cong \mathbb{Z}/n\mathbb{Z}$, where n is the smallest integer such that $n\theta_0 = 2\pi$.

(2) G has an element $r \in G$ with $\det r = -1$. Then the determinant map restricted to G gives a group homomorphism

$$G \xrightarrow{\det} \mathbb{R}^*,$$

which has image $\{1, -1\}$, and since as a subgroup $\mathbb{Z}/2\mathbb{Z} \cong \{1, -1\} \leq \mathbb{R}^*$, the determinant map defines a surjective group homomorphism

$$G \overset{\det}{\twoheadrightarrow} \{1, -1\} \cong \mathbb{Z}/2\mathbb{Z}.$$

The kernel of a group homomorphism is a normal subgroup so $\ker\det \trianglelefteq G$ is a normal subgroup of G and the first isomorphism theorem tells us that $G/\ker\det \cong \operatorname{im}\det \cong \mathbb{Z}/2\mathbb{Z}$, hence $\ker\det$ has index 2. Thus $\ker\det$ has two cosets, and

$$G = \ker\det \cup r\ker\det$$

for some $r \notin \ker \det$. Every element of $\ker \det$ has, by definition, determinant 1, so $\ker \det \cong \mathbb{Z}/n\mathbb{Z}$ for some n, and so

$$G = \{I_2, R_{\theta_0}, \dots, R_{\theta_0}^{n-1}, rR_{\theta_0}, \dots, rR_{\theta_0}^{n-1}\},$$

with $r^2 = I_2$ and $rR_{\theta_0}r = R_{\theta_0}^{-1}$. This is exactly the group D_n.

Remark 2.4.5. For those familiar with exact sequences, we've just seen that G is a split extension of the form

$$0 \to \mathbb{Z}/n\mathbb{Z} \to G \to \mathbb{Z}/2\mathbb{Z} \to 1$$

so $G \cong \mathbb{Z}/n\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$ with $\mathbb{Z}/2\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ given by

$$+1 \mapsto \mathrm{id}$$
$$-1 \mapsto (-\mathrm{id} : k \mapsto -k).$$

In total, we have shown the following.

Theorem 2.4.6. *Any finite subgroup of $E(2)$ is isomorphic to $\mathbb{Z}/n\mathbb{Z}$ or D_n.*

3. Lattices and Discrete Symmetries of $\mathbb{R}^2$

Here I talked a little about subgroups $\Lambda = \mathbb{Z}x + \mathbb{Z}y \leq R^2$, and mentioned that elliptic curves over the complex numbers, which are solutions to $y^2 = x(x-1)(x-\lambda)$ for some $\lambda \in \mathbb{C}$, correspond to lattices $\Lambda \cong \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2 \subset \mathbb{C} \cong R^2$, and that they look like donuts. I mentioned that elliptic curves are very interesting from many perspectives, and are very useful in cryptography.

Another suggestion is to talk about point groups or tiling of R^2 and wallpaper groups.

4. Finitely Generated Abelian Groups

The main theorem in this section is the structure theorem for finitely generated abelian groups. In the process, we will develop quite a bit of machinery, and the approach we take (Smith normal form) generalizes very nicely to a more abstract setting (modules over a PID).

Remark 4.0.1. Much of our development of free abelian groups is taken from David Webb's algebra 1 course at Dartmouth College, as presented in Jeff Hein's notes from Fall 2011, and from memory of taking the course in Fall 2018.

Let's get a sense of what kind of objects we will be thinking about.

Example 4.0.2. The following groups are all "finitely generated abelian groups":

- $\mathbb{Z}$,
- $\mathbb{Z}/2\mathbb{Z}$,
- $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z} \cong \mathbb{Z}/15\mathbb{Z}$,
- $\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

We already see that some of these are finite, some are infinite, and can be written down in different ways. The goal will be to have a complete list of "finitely generated abelian groups" and, like with the Euclidean group, give a useful presentation of each group.

Recall 4.0.3. If G is a group, and $\{g_i\}_{i \in I} \subset G$ is a collection of elements, then

$$\langle g_i \rangle_{i \in I} \leq G$$

is defined as the *smallest subgroup of G containing the elements $\{g_i\}_{i \in I}$.*

Example 4.0.4. Let $G = S_3$, and consider the subgroup $\langle (12) \rangle \leq S_3$. Since a subgroup is closed under multiplication, we have $\{(12), (12)^2 = (1) = \mathrm{id}\} \subseteq \langle (12) \rangle$. The set $\{(12), (12)^2 = (1) = \mathrm{id}\}$ is already a subgroup, so $\langle (12) \rangle = \{(12), (12)^2 = (1) = \mathrm{id}\}$.

Definition 4.0.5. A group is *finitely generated* is there are $g_1, \dots, g_n \in G$ such that $\langle g_1, \dots, g_n \rangle = G$.

Our goal is to understand all finitely generated abelian groups.

The main theorem is the structure theorem, which we state in two forms, though the proof will occupy us for a significant time.

Theorem 4.0.6 (Structure Theorem for finitely generated abelian groups (invariant factor form))**.** *Let G be a finitely generated abelian group. Then*

$$G \cong \mathbb{Z}^r \times \mathbb{Z}/d_1\mathbb{Z} \times \cdots \mathbb{Z}/d_k\mathbb{Z}$$

where $r \in \mathbb{Z}_{\geq 0}$, $d_1, \dots, d_k \in \mathbb{Z}_{>0}$ such that $d_1 \mid d_2 \mid \cdots \mid d_k$. Moreover, $r, d_1, \dots, d_k$ are uniquely determined.

Remark 4.0.7. Here when we say "uniquely determined" we mean that if G and H are finitely generated abelian groups and $G \cong H$, then the integers $r, d_1, \dots, d_k$ will be the same for both groups. That is, a finitely generated free abelian group is no more than the data of integers $r, d_1, d_2, \dots, d_k$ as above.

There is another form of the structure theorem, which presents the group G using slightly different (but equivalent) data. The idea for the second version is that you can factor each of the integers $d_1, \dots, d_k$ into their prime factors.

Theorem 4.0.8 (Structure Theorem for finitely generated abelain groups (primary factor form))**.** *Let G be a finitely generated abelian group. Then*

$$G \cong \mathbb{Z}^r \times \mathbb{Z}/p_1^{r_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_n^{r_n}\mathbb{Z}$$

where p_i are primes (not necessarily distinct) and $r, r_1, \dots, r_n \in \mathbb{Z} \geq 0$. Moreover, $r, p_1, \dots, p_n, r_1, \dots, r_n$ are uniquely determined.

Remark 4.0.9. This version says that a finitely generated abelian group is no more than the data of

(1) a non-negative integer r, and
(2) a multiset (a set allowing entries to repeat) of non-negative integers recording the exponents of primes, one for each prime, with only finitely many of the multisets being non-empty.

For example, the group $\mathbb{Z}^7 \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$ has the data

$$r = 7, \text{ and the multisets } \underbrace{\{1, 1, 3\}}_{\text{exponents of 2's}}, \underbrace{\emptyset}_{\text{exponents of 3's}}, \dots, \emptyset, \dots.$$

Example 4.0.10. Let G be an abelian group of order 4. Then G is clearly finitely generated (by all 4 of its elements). Let's see how G looks in the forms of the structure theorem. We have $r = 0$, as G is finite. In the invariant factor form, we have $G \cong \mathbb{Z}/d_1\mathbb{Z} \times \cdots \times \mathbb{Z}/d_k\mathbb{Z}$, and so $4 = |G| = d_1 \cdot d_2 \cdots d_k$. Thus we have two possibilities:

- $d_1 = 4$, and so $G \cong \mathbb{Z}/4\mathbb{Z}$. The primary factor form is the same.
- $d_1 = 2$, $d_2 = 2$, and so $G \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. Again, the primary factor form is the same.

The structure theorem tells us that these two groups are not isomorphic!

Example 4.0.11. Let G be an abelian group of order 20, $|G| = 20$. Again, $r = 0$ since G is finite. Since $20 = 2^2 \cdot 5$, and we have $d_1 \cdots d_k = 20$ with $d_1 \mid d_2 \mid \cdots \mid d_k$, we have two possibilities again

- $d_1 = 20$, so $G \cong \mathbb{Z}/20\mathbb{Z}$. In the primary factor form, this is $\mathbb{Z}/20\mathbb{Z} \cong \mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$.
- $d_1 = 2$, $d_2 = 10$, so $G \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/10\mathbb{Z}$. In the primary factor form, this is

$$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/10\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}.$$

It will take is a while to prove the structure theorems. The outline of our proof is as follows.

(1) Prove existence
  (1a) Show that G admits a surjective group homomorphism $\mathbb{Z}^n \overset{\varphi}{\twoheadrightarrow} G$, so $G \cong \mathbb{Z}^n/K$ where $K = \ker \varphi$.
  (1b) understand $K \leq \mathbb{Z}^n$.
    - Show that $K \cong \mathbb{Z}^m$ for some $m \leq n$
    - Show that any group homomorphism $\mathbb{Z}^m \to \mathbb{Z}^n$ can be given by a matrix

$$\begin{pmatrix} d_1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & d_k & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}$$

(2) Prove uniqueness
  (2a) Show that the torsion part and the $\mathbb{Z}^r$ part are determined, here we will see that the $\mathbb{Z}^r$ part is uniquely determined.
  (2b) Understand the torsion part (finite abelian groups)
    - Split up by primes
    - understand finite abelian p-groups

We'll begin by understanding groups that look like $\mathbb{Z}^r$.

4.1. **Free abelian groups.** The notion of a free abelian group is very similar to that of a vector space, in that both have a basis and their structure is very rigid. Basically everything you want to do can be done by using a basis. However, a vector space has some properties that free abelian groups do not enjoy, namely we can't scale elements in any way that we may wish (for example because $\frac{1}{2} \notin \mathbb{Z}$).

The definition and properties of free abelian groups are presented in fairly abstract language. This is in the spirit of a "categorical approach", made popular in the later half of the 20^{textth} century, and takes the view that an object is no more and no less than its properties, rather than something you can "hold in your hands". This usually has a huge benefit when proving theorems, but it does make intuition difficult to build.

Definition 4.1.1 (Free abelian group)**.** Let B be a set. A *free abelian group on* B is an abelian group, $F(B)$ together with a map of sets

$$\iota : B \to F(B)$$

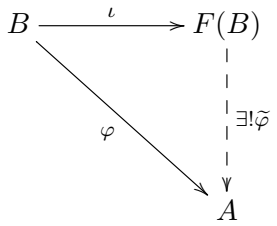
such that for any map

$$\varphi : B \to A$$

to an abelian group A, there exists a unique group homomorphism

$$\widetilde{\varphi} : F(B) \to A$$

such that $\varphi = \widetilde{\varphi} \circ \iota$. This is summarized in the commutative diagram



The set B is usually called the *basis*.

Remark 4.1.2. When we say that diagram is *commutative*, we mean that if you follow two different paths between the same two points, the paths are the same. So the path following φ and the path following ι and then $\widetilde{\varphi}$ are equal, hence $\varphi = \widetilde{\varphi} \circ \iota$.

N.B.: Not every diagram is commutative!!! But a lot are.

The definition may look scary at first, but we'll see shortly that when B is finite say $|B| = n$, then $F(B) \cong \mathbb{Z}^n$.

Definition 4.1.4 (Direct sum of abelian groups)**.** Let $\{A_i\}_{i \in I}$ be a collection of abelian groups. The direct sum $\bigoplus_{i \in I} A_i$ is an abelian group defined as

$$\bigoplus_{i \in I} A_i \coloneqq \{(a_i)_{i \in I} \mid a_i = 0 \text{ for all but finitely many } i\}.$$

That is, $\bigoplus_{i \in I} A_i$ is the collection of sequences (or tuples) of elements of the groups A_i such that only finitely many are non-zero. The group operation is just component-wise addition.

Another way of defining $\bigoplus_{i \in I} A_i$ is as finite (commutative) sums of elements in the groups A_i.

We now say what the groups $F(B)$ look like.

Theorem 4.1.5. *Let B be a set, then $F(B) \cong \bigoplus_{b \in B} \mathbb{Z}$. Moreover, this isomorphism is unique.*

Proof. We first show that $\bigoplus_{b \in B} \mathbb{Z}$ is a free abelian group.

Define a map of sets

$$i : B \to \bigoplus_{b \in B} \mathbb{Z}, \ b \mapsto (0, 0, \dots, 0, \underbrace{1}_{b \text{ entry}}, 0, \dots).$$

Let A be an abelian group and $\varphi : B \to A$ a map of sets. We define

$$\widetilde{\varphi} : \bigoplus_{b \in B} \mathbb{Z} \to A, (a_b)_b \mapsto \sum_{b \in B} a_b \varphi(b).$$

This is a well-defined map as $(a_b)_b$ has finitely many non-zero entries, so the sum in A is a finite sum.

Clearly we have $\varphi = \widetilde{\varphi} \circ i$. And we see that $\widetilde{\varphi}$ is a group homomorphism as

$$\begin{aligned}
\widetilde{\varphi}\left((a_b)_b + (c_b)_b\right) &= \widetilde{\varphi}\left((a_b + c_b)_b\right) \\
&= \sum_{b \in B} (a_b + c_b)\varphi(b) \\
&= \sum_{b \in B} a_b\varphi(b) + c_b\varphi(b) \\
&= \sum_{b \in B} a_b\varphi(b) + \sum_{b \in B} c_b\varphi(b) \\
&= \widetilde{\varphi}\left((a_b)_b\right) + \widetilde{\varphi}\left((c_b)_b\right).
\end{aligned}$$

To see that $\widetilde{\varphi}$ is unique, suppose $f : \bigoplus_{b \in B} \mathbb{Z} \to A$ is another group homomorphism such that $\varphi = f \circ i$. Then

$$\begin{aligned}
f((a_b)_b) &= f\left(\sum_{b \in B} a_b i(b)\right) \\
&= \sum_{b \in B} a_b f(i(b)) \\
&= \sum_{b \in B} a_b \varphi(b) \\
&= \widetilde{\varphi}\left((a_b)_b\right),
\end{aligned}$$

so $f = \widetilde{\varphi}$. Hence $\bigoplus_{b \in B} \mathbb{Z}$ is a free abelian group.

We now show that $F(B) \cong \bigoplus_{b \in B} \mathbb{Z}$.

Consider the diagram

$$\begin{array}{ccc}
B & \xrightarrow{\ i\ } & \bigoplus_{b \in B} \mathbb{Z} \\
& \searrow^{\iota} & \downarrow{\scriptstyle \exists!\ \widetilde{\iota}} \\
& & F(B)
\end{array}$$

where the group homomorphism $\widetilde{\iota} : \bigoplus_{b \in B} / Z \to F(B)$ is the unique group homomorphism obtained from the map of sets $\iota : B \to F(B)$. Similarly, consider the diagram

$$\begin{array}{ccc}
B & \xrightarrow{\ \iota\ } & F(B) \\
& \searrow^{i} & \downarrow{\scriptstyle \exists!\ \widetilde{i}} \\
& & \bigoplus_{b \in B} \mathbb{Z}
\end{array}$$

where the group homomorphism $\widetilde{i} : F(B) \to \bigoplus_{b \in B} \mathbb{Z}$ is the unique group homomorphism obtained from the map of sets $i : B \to \bigoplus_{b \in B} \mathbb{Z}$. Now consider the diagram

$$\begin{array}{ccc}
 & & F(B) \\
 & \nearrow^{\iota} & \downarrow{\scriptstyle \widetilde{i}} \\
B & \xrightarrow{\ i\ } & \bigoplus_{b \in B} \mathbb{Z} \\
 & \searrow^{\iota} & \downarrow{\scriptstyle \exists!\ \widetilde{\iota}} \\
 & & F(B)
\end{array}$$

the homomorphism $\widetilde{\iota} \circ \widetilde{i} : F(B) \to F(B)$ is a group homomorphism such that $\iota = \widetilde{\iota} \circ \widetilde{i} \circ \iota$. The identity map

$$\mathrm{id}_{F(B)} : F(B) \to F(B)$$

is also a group homomorphism satisfying $\iota = \mathrm{id}_{F(B)} \circ \iota$, thus as $F(B)$ is a free abelian group and such homomorphisms are unique, we have

$$\widetilde{\iota} \circ \widetilde{i} = \mathrm{id}_{F(B)}.$$

Mutatis mutandis, we have

$$\widetilde{i} \circ \widetilde{\iota} = \mathrm{id}_{\bigoplus_{b \in B} \mathbb{Z}}.$$

Thus $\widetilde{\iota}$ and $\widetilde{i}$ are inverse isomorphisms and $F(B) \cong \bigoplus_{b \in B} \mathbb{Z}$. Moreover, since the homomorphisms $\widetilde{\iota}$ and $\widetilde{i}$ were obtained as the unique homomorphisms extending the maps ι and i, respectively, the isomorphism is unique. □

As with vector spaces, the number of elements in a basis should be unique.

Theorem 4.1.6. *If $\mathbb{Z}^m \cong \mathbb{Z}^n$, then $m = n$.*

Definition 4.1.7 (Rank of a free abelian group)**.** The number n above is called the *rank* of $\mathbb{Z}^n$.

Proof of Theorem 4.1.6. Let p be a prime. The group $\mathbb{Z}^m$ has a basis $\{b_1, \dots, b_m\}$, thus

$$\mathbb{Z}^m / p\mathbb{Z}^m \cong \underbrace{\mathbb{Z}/p\mathbb{Z} \times \cdots \times \mathbb{Z}/p\mathbb{Z}}_{m \text{ times}},$$

so

$$|\mathbb{Z}^m / p\mathbb{Z}^m| = p^m.$$

Now, if $\mathbb{Z}^n \subset \mathbb{Z}^m$ is the isomorphic copy of $\mathbb{Z}^n$, then

$$\mathbb{Z}^n / p\mathbb{Z}^n \subset \mathbb{Z}^m p\mathbb{Z}^m,$$

hence $p^n \leq p^m$. Similarly, $p^n \geq p^m$. Thus $p^n = p^m$, and so $n = m$. □

Proposition 4.1.8. *If G is a finitely generated abelian group, then G is a quotient of $\mathbb{Z}^n$ for some n.*

Proof. Let G be finitely generated by $B = \{g_1, \dots, g_n\}$. Then the inclusion map $B \to G$ gives a unique group homomorphism $\varphi : \mathbb{Z}^n \to G$, which is surjective as B generates G. This the first isomorphism theorem tells us that

$$G \cong \mathbb{Z}^n / \ker \varphi.$$ □

This finished part (1a) of our proof of the structure theorem for finitely generated abelain groups!

4.2. **Subgroups of $\mathbb{Z}^n$.** We now know that any finitely generated abelian group G is a quotient of $\mathbb{Z}^n$ for some n. That is, there is a surjective map $\mathbb{Z}^n \xrightarrow{\varphi} G$. We want to understand $\ker \varphi$. We'll eventually show that because $\ker \varphi \leq \mathbb{Z}^n$ is a subgroup, then $\ker \varphi \cong \mathbb{Z}^k$ for some $k \leq n$. But we need to build up some theory first.

Question 1. *When is a group a direct product? That is, when is $G \cong G_1 \times G_2$?*

When $G = G_1 \times G_2$, there are the inclusions (which are group homomorphisms)

$$i_1 : G_1 \hookrightarrow G_1 \times G_2, g_1 \mapsto (g_1, \mathrm{id}_{G_2})$$

and

$$i_2 : G_2 \hookrightarrow G_1 \times G_2, g_2 \mapsto (\mathrm{id}_{G_1}, g_2).$$

We also have the quotient map

$$\varphi : G \to G/i_1(G_1) \cong G_2, (g_1, g_2) \mapsto (\mathrm{id}_{G_1}, g_2)$$

where we quotient by G_1. Thus we have a few maps

$$G_1 \xrightarrow{i_1} G_1 \times G_2 \xrightarrow{\varphi} G_2, \quad i_2 : G_2 \to G_1 \times G_2$$

$$g_1 \xmapsto{i_1} (g_1, \mathrm{id}_{G_2})$$

$$(g_1, g_2) \xmapsto{\varphi} g_2$$

$$(\mathrm{id}_{G_1}, g_2) \xmapsfrom{i_2} g_2$$

it is clear that $\varphi \circ i_2 = \mathrm{id}_{G_2}$ and $\mathrm{im} i_1 = \ker \varphi$.

Definition 4.2.1. Let $\varphi : G \to G_2$ be a group homomorphism. A *section of* φ is a group homomorphism $s : G_2 \to G$ such that $\varphi \circ s = \mathrm{id}_{G_2}$.

It turns out that for abelian groups, a group G is a direct product of two groups G_1 and G_2 exactly when there is a surjective group homomorphism $\varphi : G \to G_2$ with a section $s : G_2 \to G$ such that $\ker \varphi \cong G_1$. We will prove this shortly in Lemma 4.2.4.

Remark 4.2.2. It is **not** true that having a surjective map with a section is enough for a non-abelian group to be a direct product. For an example, consider S_3.

Let's see how this direct product nonsense can help us distinguish some groups.

Example 4.2.3. We will show that $\mathbb{Z}/4\mathbb{Z} \ncong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

Proof. Suppose for contradiction that $\mathbb{Z}/4\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. Then we would have homomorphisms

$$\begin{aligned} i_1 &: \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/4\mathbb{Z} \\ \varphi &: \mathbb{Z}/4\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \\ i_2 &: \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/4\mathbb{Z} \end{aligned}$$

such that $\varphi \circ i_2 = \mathrm{id}$ and $\mathrm{im} i_1 = \ker \varphi$. Since i_2 is a group homomorphism, we would need $i_2(0) = 0$, and $i_2(1) \neq 0$ else $\varphi \circ i_2 \neq \mathrm{id}$. But

$$i_2(1) + i_2(1) = i_2(1+1) = i_2(0) = 0,$$

so $i_2(1) = 2 \in \mathbb{Z}/4\mathbb{Z}$. But $2 \in \ker \varphi$, hence $\varphi \circ i_2(1) = \varphi(2) = 0 \neq 1$, which is a contradiction. Hence

$$\mathbb{Z}/4\mathbb{Z} \ncong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}.$$

□

We now show that for abelian groups, having a section is enough to get a direct product.

Lemma 4.2.4. *Let G and Q be a abelian groups with a surjective group homomorphism $f : G \twoheadrightarrow Q$. Suppose there is a group homomorphism $s : Q \to G$ such that $f \circ s = \mathrm{id}_Q$, i.e. that s is a section of f. Then $G \cong \ker f \times Q$.*

Proof. Let $\ker f \overset{i}{\hookrightarrow} G$ be the inclusion. We have a map

$$\varphi : \ker f \times Q \to G, (k, q) \mapsto i(k) + s(q),$$

and we want to show that φ is a isomorhism.

Since both i and s are group homomorphisms, φ is also a group homomorphism.

We first show that φ is injective. Let $(k,q) \in \ker\varphi \leq \ker f \times Q$. Then

$$0 = \varphi((k,q)) = i(k) + s(q).$$

Thus

$$0 = f(i(k) + s(q)) = f \circ i(k) + f \circ s(q) = 0 + q = q,$$

as $\mathrm{im} i = \ker f$. Hence as $q = 0$, $s(q) = 0$, and so $i(k) = 0$, But i is injective, hence $k = 0$, and so $(k,q) = (0,0) \in \ker f \times Q$. Thus φ is injective.

We now show that φ is surjective. Let $g \in G$. Then $g - s \circ f(g) \in \ker f$. Indeed,

$$f(g - s \circ f(g)) = f(g) - f \circ s \circ f(g) = f(g) - \mathrm{id}_Q(f(g)) = f(g) - f(g) = 0.$$

But then as $\ker f = \mathrm{im} i$, we have $g - s \circ f(g) = i(k)$ for some $k \in \ker f$. Now

$$\varphi(k, f(g)) = i(k) + s \circ f(g) = g - s \circ f(g) + s \circ f(g) = g,$$

and so φ is surjective.

Hence φ is an injective and surjective group homomorphism, hence an isomorphism, as claimed. □

Lemma 4.2.5 ("Free abelain groups are projective")**.** *Let G and Q be abelian groups with a surjective group homomorphism $f : G \twoheadrightarrow Q$. Suppose that Q is a free abelian group. Then there is a section $s : Q \to G$ of f, i.e. a group homomorphism such that $f \circ s = \mathrm{id}_Q$. In particular, $G \cong \ker f \times Q$.*

Proof. Since Q is a free abelain group, $Q \cong F(B)$ for some set $B = \{b_i\}_{i \in I}$, and let $\iota : B \to Q$ be the inclusion. As f is surjective, for each $b_i \in B$ there exists some $g_i \in G$ such that $f(g_i) = b_i$. Define a map

$$\varphi : B \to G, b_i \mapsto g_i.$$

Since Q is free abelian, there is a unique group homomorphism $s : Q \to G$ such that $\varphi = s \circ \iota$. In particular, $s(b_i) = g_i$. Let $q = \sum_i a_i b_i \in Q$, then

$$\begin{aligned} f \circ s(q) &= f(\sum_i a_i s(b_i)) \\ &= \sum_i a_i f(s(b_i)) \\ &= \sum_i a_i f(g_i) \\ &= \sum_i a_i b_i \\ &= q, \end{aligned}$$

thus $f \circ s = \mathrm{id}_Q$, and s is a section of f.

The statement that $G \cong \ker f \times Q$ now follows from Lemma 4.2.4. □

We are now ready to show that any subgroup of $\mathbb{Z}^n$ is of the form $\mathbb{Z}^m$ for some $m \leq n$. This

Theorem 4.2.6 (Subgroups of Z^n)**.** *Let leq$\mathbb{Z}^n$ be a subgroup. Then $K \cong \mathbb{Z}^m$ for some $m \leq n$.*

Proof. We will use induction on n.

Base case $n = 1$: If $n = 1$, then $K \leq \mathbb{Z}$ is a subgroup. Hence it is either 0 or $a\mathbb{Z}$ for some integer a. As $0 \cong \mathbb{Z}^0$ and $a\mathbb{Z} \cong \mathbb{Z}$, we have $K \cong \mathbb{Z}^m$ for some $m \leq n = 1$.

Inductive step: Now suppose that all subgroups of $\mathbb{Z}^{n-1}$ are of the form Z^m for some $m \leq n-1$. Let $p : \mathbb{Z}^n \to Z, (a_1, \ldots, a_n) \mapsto a_n$ be the projection onto the last coordinate. We clearly have

$$\ker p = \{(a_1, \ldots, a_{n-1}, 0) \mid a_i \in \mathbb{Z}\} \cong \mathbb{Z}^{n-1}.$$

Suppose $K \leq \mathbb{Z}^n$ is a subgroup.

We have the following commutative diagram

$$\begin{array}{ccccc} & & \ker p|_K & \longrightarrow & K \xrightarrow{p|_K} p(K) \\ & & \cap| & & \cap| \quad \cap| \\ \ker p & \cong & \mathbb{Z}^{n-1} & \hookrightarrow & \mathbb{Z}^n \xrightarrow{p} \mathbb{Z} \end{array}$$

and see that $p(K) \subseteq \mathbb{Z}$. Hence $p(K) \cong \mathbb{Z}^k$ for some $l \leq 1$. By the inductive hypothesis, as $\ker p|_K \leq \mathbb{Z}^{n-1}$ is a subgroup, $\ker p|_K \cong \mathbb{Z}^a$ for some $a \leq n-1$.

Since K and $p(K)$ are abelian groups, and $p|_K : K \to p(K)$ is a surjective group homomorphism, Lemma 4.2.5 shows that $K \cong \ker p|_K \times p(K)$. Thus we have $K \cong \mathbb{Z}^k \times \mathbb{Z}^a \cong \mathbb{Z}^{a+k}$ for some $a + k \leq 1 + n - 1 = n$, as desired. □

This finishes the first part of (1b) of our proof of the structure theorem for finitely generated abelian groups.

4.3. **Smith Normal Form.** So far in our proof of the structure theorem for finitely generated abelain groups, we know that a finitely generated abelian group G is the quotient $G \cong \mathbb{Z}^n/\varphi(\mathbb{Z}^m)$ for some homomorphism $\varphi : \mathbb{Z}^m \to \mathbb{Z}^n$. We now aim to understand these homomorphisms better.

We'll start with an example.

Example 4.3.1. Suppose $f : \mathbb{Z}^2 \to \mathbb{Z}^3, (a, b) \mapsto (2a, 3b, 0)$. What is $\mathbb{Z}^3/f(\mathbb{Z}^2)$?

We can represent f as a matrix

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix}$$

and the image of f is spanned by the columns of the matrix. So

$$\mathbb{Z}^3/f(\mathbb{Z}^2) \cong \mathbb{Z}^3/\left(\mathbb{Z}\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \times \mathbb{Z}\begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix}\right) \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3Z \times \mathbb{Z}.$$

Another way to arrive at this is to perform some column and row operations.

Recall 4.3.2. Let f be a matrix and let c and r be invertible matrices. Invertible matrices correspond to ca change of basis. The matrix fc corresponds to changing the basis of the domain of f, and corresponds to performing column operations on f. While the matrix rf corresponds to changing the basis of the target of f and corresponds to row operations on f.

For example, if $f \in M_{2\times 2}(\mathbb{R})$ and $c = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$, then fc is obtained from f by adding $x \cdot$ column 1 to column 2.

To finish proving the existence part of the structure theorem, we just have to show that for a group homomorphism $f : \mathbb{Z}^m \to \mathbb{Z}^n$, we can use row and column operations until f has a matrix

of the form

$$\begin{pmatrix} d_1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & d_k & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}$$

where $d_1 \mid d_2 \mid \cdots \mid d_k$ and potentially some of the d_i are zero, this is called *Smith Normal Form*.

Remark 4.3.3. But when we use row and column operations, we cannot scale by just any number. We are only allowed to scale by integers! Since the maps c and r have to be isomorphisms of $\mathbb{Z}^m$ or $\mathbb{Z}^n$!

Then for a finitely generated abelian group G, we will have

$$G \cong \mathbb{Z}^n/\text{im} \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_k \end{pmatrix} \cong \underbrace{\mathbb{Z}/d_1\mathbb{Z} \times \cdots \times \mathbb{Z}/d_j\mathbb{Z}}_{d_i \neq 0} \times \underbrace{\mathbb{Z} \times \cdots \mathbb{Z}}_{d_i = 0}.$$

Remark 4.3.4. One can in fact show that Smith Normal Form is unique, which proves the invariant factor form of the structure theorem (Theorem 4.0.6). But we will not do this.

Let's see an example first of how we can get a matrix into Smith Normal Form.

Example 4.3.5. Let's see how we can get the matrix

$$f = \begin{pmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix}$$

into Smith Normal form using row and column operations. There are many ways to do this, so you are encouraged to pause reading and work out a way yourself.

One way to get the matrix f into Smith Normal form is as follows:

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix} \mapsto \begin{pmatrix} 2 & 3 \\ 0 & 3 \\ 0 & 0 \end{pmatrix} \mapsto \begin{pmatrix} 2 & 1 \\ 0 & 3 \\ 0 & 0 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 2 \\ 3 & 0 \\ 0 & 0 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 \\ 3 & -6 \\ 0 & 0 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -6 \\ 0 & 0 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 6 \\ 0 & 0 \end{pmatrix}.$$

And now we see that

$$\mathbb{Z}^3/\text{im}f \cong \mathbb{Z}/1\mathbb{Z} \times \mathbb{Z}/6\mathbb{Z} \times \mathbb{Z} \cong \mathbb{Z}/6\mathbb{Z} \times \mathbb{Z},$$

which you should compare to Example 4.3.1.

Here is how the Smith Normal form algorithm works in general.

4.3.1. *Smith Normal Form algorithm.* We'll show that there is an algorithm that takes a matrix f with integer entries and using row and column operations only scaling by integers we arrive at a matrix in Smith Normal Form.

Smith Normal Form Algorithm Input: matrix f with integer entries.
Output: matrix f' in Smith Normal form obtained by using column and row operations over $\mathbb{Z}$ on the matrix f.

(i) If $f = 0$, the we are done, and the output if f. Else we may assume that $f \neq 0$, and some entry $a_{ij} \neq 0$. Using row and column operations, move a_{ij} to the 11 entry.

Our matrix now looks like

$$\begin{pmatrix} a_{11} & \cdots & * \\ \vdots & \ddots & \vdots \\ * & \cdots & * \end{pmatrix}$$

with $a_{11} \neq 0$.

(ii) If the only non-zero entry in the first row is a_{11}, continue. Otherwise, there is another non-zero entry in the first row. Using column operations, move a non-zero entry into the second position in the first row, into the a_{21} slot. Compute $\gcd(a_{11}, a_{21}) = ma_{11} + na_{21}$ with $m, n \in \mathbb{Z}$, and using column operations make this the a_{21} entry. Now swap the first two columns. Our matrix now looks like

$$\begin{pmatrix} \gcd(a_{11}, a_{21}) & a_{11} & \cdots \\ \vdots & \ddots & \ldots \end{pmatrix}$$

and since $\gcd(a_{11}, a_{21})$ divides a_{11}, we can multiple the first column by an integer and subtract it from the second column to clear out the second entry in the first row. Our matrix now looks like

$$\begin{pmatrix} \bullet & 0 & * & \cdots \\ * & * & \cdots & * \\ \vdots & \cdots & * & * \end{pmatrix}$$

and we continue this process until we zero out the first row except for the first entry. At the end of this step, our matrix looks like

$$\begin{pmatrix} \bullet & 0 & \cdots & 0 \\ * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{pmatrix}$$

(iii) We now do the same process as step (ii) but on the first column.

(iv) In step (iii) we may end up with non-zero entries in the first row again. Then we repeat step (ii) and step (iii) until we end up with a matrix of the form

$$\begin{pmatrix} \bullet & 0 & \cdots & 0 \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{pmatrix}.$$

This process must terminate because at the end of each step (ii) and (iii), the 11-entry is decreasing in absolute value, but it is an integer, so the process cannot continue forever and must terminate.

(v) We now want the 11-entry to divide all of the entries in the matrix. If the 11-entry does not divise an entry a_{ij}, then add row j to the first row move, moving the a_{ij} entry into the first row. Now perform steps (ii), (iii) and (iv). Continue until the 11-entry divides all entries. This process must terminate for the same reason as before.

At the end of this step, we have a matrix

$$\begin{pmatrix} \bullet & 0 & \cdots & 0 \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{pmatrix}$$

where the entry $\bullet$ divides all entries.

(vi) We now repeat this process on the smaller matrix. In doing this the 11 entry still divides all the other entries as we have only taken greatest common divisors, and the 11-entry, dividing all of the entries will certainly divide their greatest common divisors.

We've now seen that we can get any integer matrix into Smith Normal form using row and column operations over $\mathbb{Z}$, and this concludes part (1) of our proof of the structure theorem for finitely generated abelian groups.

Thus so far we have proved the following theorem which is already cause to celebrate!

Theorem 4.3.6. *Let G be a finitely generated abelian group. Then*

$$G \cong \mathbb{Z}^r \times \mathbb{Z}/d_1\mathbb{Z} \times \cdots \times \mathbb{Z}/d_k\mathbb{Z}$$

for some integers $r \in \mathbb{Z}_{\geq 0}$ and $d_1, \ldots, d_k \in \mathbb{Z}_{>0}$ such that $d_1 \mid d_2 \mid \cdots \mid d_k$.

4.4. **Uniqueness in the Structure Theorem.** Our next goal is to show that the integers in the structure theorems Theorem 4.0.6 and Theorem 4.0.8 are uniquely determined.

In particular, from Theorem 4.3.6, we see that every finitely generated abelian group is the direct product of a free part and a torsion part. This will be very useful. To prove the uniqueness statement, we will break up our finitely generated abelian group into exactly these two pieces and show that each is uniquely determined.

Definition 4.4.1. Let G be an abelian group. Let

$$G_{tors} \coloneqq \{g \in G \mid g^n = \mathrm{id}_G \text{ for some } n \in \mathbb{Z}_{>0}\}$$

denote the subset of elements with finite order.

Exercise 4.4.2. Let G be an abelian group. The set G_{tors} is a subgroup.

We want to know that G_{tors} is uniquely determined, and this is true!

Lemma 4.4.3. *Let $\varphi : G \to G'$ be an isomorphism of abelian groups. Then $\varphi(G_{tors}) \subset G'_{tors}$, and $\varphi^{-1}(G'_{tors}) \subset G_{tors}$. In particular, $\varphi|_{G_{tors}} : G_{tors} \to G'_{tors}$ is an isomorphism.*

Remark 4.4.4. This is a good proof to do on your own to make sure you're following. If you have the time, pause here and write a proof on your own.

Proof. Let $x \in G_{tors}$. So $nx = 0$ for some integer $n > 0$. Then since φ is a group homomorphism we have

$$0 = \varphi(0) = \varphi(nx) = n\varphi(x),$$

so $\varphi(x) \in G'_{tors}$.

Similarly with the inverse isomorphism φ^{-1}, we see that $\varphi^{-1}(G'_{tors}) \subset G_{tors}$.

Since φ is a bijection with inverse φ^{-1}, we see that restricting φ to G_{tors} gives an isomorphism $\varphi|_{G_{tors}} : G_{tors} \to G'_{tors}$ with inverse $\varphi^{-1}|_{G'_{tors}}$.

Exercise 4.4.5. Check this last statement carefully.

This completes the proof. □

Exercise 4.4.6. Show that for an abelian group G, G/G_{tors} is also uniquely determined up to isomorphism.

We now want to show that if we get rid of the torsion part, we are left with something free. We'll need a lemma first.

Definition 4.4.7. Let G be a group. G is called *torsion free* if g has no nontrivial elements of finite order.

Lemma 4.4.8. *A finitely generated torsion free abelian group is free.*

Remark 4.4.9. If you have seen a rigorous treatment of linear algebra, you may recognize the proof as finding a basis.

Proof. Let A be a finitely generated torsion free abelian group. Let S be a finite set that generates A, and let $\{x_1, \dots, x_n\} \subset S$ be maximal such that if $a_1, \dots, a_n \in Z$ with $\sum_{i=1}^n a_i x_i$ then $a_i = 0$ for all $0 \leq i \leq n$. Let $B = \langle x_1, \dots, x_n \rangle \leq A$. Then B is a free abelian group. Indeed, there is an isomorphism $B \cong \mathbb{Z}^n$ sending x_i to the standard basis, which is clearly surjective, and injectivity follows from the linear independence of $\{x_1, \dots, x_n\}$ over $\mathbb{Z}$.

We want to show that $B = A$. If $S = \{x_1, \dots, x_n\}$, then we are done, as then $A = B$, the latter of which is free.

Let $y \in S$. By the maximality of $\{x_1, \dots, x_n\}$, there are some integers $m, a_1, \dots, a_n$ not all zero such that

$$my + \sum_{i=1}^{n} a_i x_i = 0.$$

We note that there must be some integers with $m \neq 0$, else $\{y, x_1 \dots, x_n\}$ would be a larger subset of S that is linear independent over $\mathbb{Z}$.

Thus for each generator $y_1, \dots, y_k \in S$, there is some $m_i \in \mathbb{Z}$ such that $m_i y_i \in B$. There is some integer m such that $mA \subseteq B$, for example let $m = m_1 \cdots m_k$.

Define the map $\mu_m : A \to A$, $x \mapsto mx$, the multiplication by m map. This is a group homomorphism. Moreover, $\ker \mu_m = 0$ as A is torsion free. Thus by the first isomorphism theorem

$$A \cong mA \subset B,$$

and so A is isomorphic to a subgroup of B, a free abelian group. Thus by Theorem 4.2.6, we see that $A \cong \mathbb{Z}^m$ for some m, and A is indeed a free abelian group. □

We are now ready to show that finitely generated abelian groups are a direct product of the torsion and free parts.

Theorem 4.4.10. *Let G be a finitely generated abelian group. Then*

(1) *G_{tors} is a finite group,*
(2) *G/G_{tors} is free, and*
(3) *$G \cong G_{tors} \times G/G_{tors}$.*

Proof. We first prove (1). Let $\varphi : \mathbb{Z}^n \twoheadrightarrow G$ be a surjective group homomorphism. Then the preimage of G_{tors}m $\varphi^{-1}(G_{tors}) \leq \mathbb{Z}^n$ is a subgroup, hence finitely generated. Thus as φ is surjective, and

$$\varphi|_{\varphi^{-1}(G_{tors})} : \varphi^{-1}(G_{tors}) \to G_{tors}$$

is surjective, we see that G_{tors} is finitely generated as well.

Exercise 4.4.11. A finitely generated torsion group is finite.

To prove (2), it remains to show that G/G_{tors} is torsion free, and then the result follows from Lemma 4.4.8. Let $\overline{x} \in G/G_{tors}$ such that $m\overline{x} = 0$. Then for any x representing $\overline{x}$, we have $mx \in G_{tors}$, so there is some integer q such that $qmx = 0$. Thus $x \in G_{tors}$, and so $\overline{x} = 0$. Hence G/G_{tors} is indeed torsion free, and thus G/G_{tors} is free by Lemma 4.4.8.

To prove (3), we consider the quotient homomorphism $G \twoheadrightarrow G/G_{tors}$. As G/G_{tors} is free, Lemma 4.2.5 shows that $G \cong G_{tors} \times G/G_{tors}$, as claimed. □

So far we have proven that for a finitely generated abelian group G, G_{tors} and G/G_{tors} are both uniquely determined. Combined with the existence result (Theorem 4.3.6) and the fact that rank is well-defined (Theorem 4.1.6), we now have a part of the uniqueness statement.

Indeed, as

$$(\mathbb{Z}^r \times \mathbb{Z}/d_1\mathbb{Z} \times \cdots \mathbb{Z}/d_k\mathbb{Z})_{tors} \cong \mathbb{Z}/d_1\mathbb{Z} \times \cdots \mathbb{Z}/d_k\mathbb{Z},$$

we have the following theorem.

Theorem 4.4.12. *Let G be a finitely generated abelian group. Then*

$$G \cong \mathbb{Z}^r \times \mathbb{Z}/d_1\mathbb{Z} \times \cdots \times \mathbb{Z}/d_k\mathbb{Z}$$

for some integers $r \in \mathbb{Z}_{\geq 0}$ and $d_1, \ldots, d_k \in \mathbb{Z}_{>0}$ such that $d_1 \mid d_2 \mid \cdots \mid d_k$. Moreover, r is uniquely determined, and the group

$$\mathbb{Z}/d_1\mathbb{Z} \times \cdots \mathbb{Z}/d_k\mathbb{Z}$$

is uniquely determined up to isomorphism. (So far we have not proven that the integers are uniquely determined).

However, this is a big step, and it remains to understand finite abelian groups.

4.5. **Finite Abelian Groups.** We'll begin with an example of breaking up a finite abelian group into pieces based on the prime factorization of the order fo the group.

Example 4.5.1. The map

$$\varphi : \mathbb{Z}/6\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z},\ n \mod 6 \mapsto (n \mod 2, n \mod 3)$$

is an isomorphism. Indeed, you can readily check that φ is a homomorphism and that it is bijective.

Definition 4.5.2. Let A be an abelian group and p a prime. We define the p-torsion subgroup as

$$A(p) \coloneqq \{x \in A \mid \operatorname{order} x = p^k \text{ for some } k\}.$$

Exercise 4.5.3. Show that for an abelian group A, $A(p) \leq A$ is a subgroup and is uniquely determine by A up to isomorphism.

Theorem 4.5.4. *Let A be a torsion abelian group. Then $A \cong \bigoplus_p A(p)$.*

Proof. There is a clear homomorphism in one direction, namely

$$\varphi : \bigoplus_p A(p) \to A, (x_p)_p \mapsto \sum_p x_p.$$

We first show that φ is injective. Let $x \in \ker \varphi$, then $\sum_p x_p = 0$. Let q be prime, so $x_q = -\sum_{p \neq q} x_p$. Let $m = \operatorname{lcm}(\text{ orders of } x_p \text{ for } p \neq q)$, then

$$mx_q = -m\left(\sum_{p \neq q} x_p\right) = -\sum_{p \neq q} mx_p = 0.$$

But $x_q \in A(q)$, so $q^r x_q = 0$ for some r. Now let $d = \gcd(q^r, m)$, and observe that $dx_q = 0$. However, since m has no factors of q, $d = 1$, and so $x_q = 0$. Thus $x_q = 0$ for any prime q, and $\ker \varphi = 0$.

We now show that φ is surjective. Let $A_m = \ker \mu_m$. If $m = rs$ with $\gcd(r, s) = 1$, then $A_m = A_r + A_s$. Indeed, we can write $1 = ur + vs$ for some integers u, v and thus $x \in A_m$ is of the form

$$x = 1 \cdot x = \underbrace{urx}_{\in A_s} + \underbrace{usx}_{\in A_r}.$$

Inductively now, if $m = \prod_{i=1}^{k} p_i^{e_i}$ for distinct primes p_i, then

$$A_m = \sum_{p_i | m} A_{p_i^{e_i}}.$$

Thus the map $\varphi : \bigoplus_p A(p) \to A$ is surjective as every element of A is in A_m for some m. □

Thus we have shown that for a finite abelian group G, we have

$$G \cong G(p_1) \times \cdots \times G(p_k),$$

where $|G| = \prod_{i=1}^{k} p_i^{e_i}$ is the prime decomposition into distinct primes p_i, and these finite groups $G(p_i)$ are uniquely determined.

Our next goal is to understand the finite groups $G(p)$ for a prime p. We'll need to build up some theory. We'll start by giving a slightly weaker result of a theorem we will prove more generally later.

Theorem 4.5.5 (Cauchy's Theorem, abelian version)**.** *Let G be a finite abelian group of order n. If p is a prime dividing n, then G has an element of order p.*

Remark 4.5.6. The stronger result (Cauchy's theorem) drops the assumption that G is abelian.

Proof of Theorem 4.5.5. We use strong induction on the order of G, assuming it is true for all groups of order $< n$, the base case being trivial.

If $p \mid |G|$, then $|G| > 1$. Let $x \in G \setminus \{0\}$. If $|G| = p$, then x has order p, by Lagrange's Theorem. So we may assume $|G| > p$.

If p divides the order of x, so the order of x is np, then the element nx is nonzero and has order p. So we may assume that $p \nmid \operatorname{ord}(x)$.

Let $N = \langle x \rangle$. By Lagrange's Theorem, $p \mid |G/N| = \frac{|G|}{|N|}$, thus $N = (pa)N$ for some $a \in G \setminus N$. If $|N| = n$, then $\gcd(p, n) = 1$, so $1 = xn + yp$ for some integers x, y, and $\operatorname{ord}(pa) \mid n$ by Lagrange's Theorem. Note also that $pna = 0$. We claim that $\operatorname{ord}(na) = p$. It remains to show that $na \neq 0$.

But if $na = 0$, then $a = 1 \cdot a = xna + ypa = ypa = y(pa) \in N$, which is a contradiction. □

Lemma 4.5.7. *Let G be a finite abelian group of order $|G| = \prod_{i=1}^{k} p_i^{e_i}$. Then $|G(p_i)| = p_i^{e_i}$.*

Proof. If $q \mid |G(p_i)|$ for some prime $q \neq p_i$, then by Theorem 4.5.5, $G(p_i)$ would have an element of order q. But every element of $G(p_i)$ has order a power of p_i, so only p_i divides $|G(p_i)|$. The result now follows from Theorem 4.5.4. □

Definition 4.5.8. Let p be a prime. A finite group of order p^e for some integer $e > 0$ is called a p-group.

Remark 4.5.9. We've just seen that for a finite abelian group, every $G(p)$ is a p-group. Thus, in light of Theorem 4.5.4, to finish the proof of the structure theorem it remains to prove a structure theorem for finite p-groups.

Theorem 4.5.10 (Structure Theorem for finite abelian p-groups)**.** *Let p be a prime number and let G be a finite abelian p-group. Then*

$$G \cong \mathbb{Z}/p^{e_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{e_s}\mathbb{Z}$$

with integers $e_1 \geq e_2 \geq e_s > 0$ and $e_1, \ldots, e_s$ uniquely determined.

Before we give a proof we outline how the Structure Theorem for finitely generated abelian groups follows.

Let G be a finitely generated abelian group. We've seen (Theorem 4.4.12) that

$$G \cong G_{tors} \times \mathbb{Z}^r$$

for some integer r that is uniquely determined and the finite torsion group G_{tors} is uniquely determined as well. From Theorem 4.5.4, the finite torsion group $G_{tors} \cong G(p_1) \times \cdots \times G(p_k)$ and the primes and groups $G(p_i)$ are uniquely determined. Now Theorem 4.5.10 gives us a unique way to present the groups $G(p_i)$ as

$$G(p_i) \cong \mathbb{Z}/p_i^{e_{p_i,1}}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_i^{e_{p_i,s_{p_i}}}\mathbb{Z},$$

with the integers $e_{p_i,1}, \dots, e_{p_i,s_{p_i}}$ uniquely determined. This exactly proves the Primary Factor form of the Structure Theorem (Theorem 4.0.8).

Let us explain how to go back and forth between the invariant factor form and the primary factor form, and the uniqueness of the invariant factor form follows from the uniqueness of the primary factor form, thereby proving Theorem 4.0.6 as well.

Recall 4.5.11. We will need the following theorem.

Theorem 4.5.12 (Chinese Remainder Theorem)**.** *$\mathbb{Z}/mn\mathbb{Z} \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ if and only if $\gcd(m,n) = 1$.*

Thus to obtain the primary factor form from the invariant factor form, simply factor the d_i's

$$\mathbb{Z}/d_1\mathbb{Z} \times \cdots \times \mathbb{Z}/d_k\mathbb{Z} \cong \underbrace{\mathbb{Z}/p_1^{e_{1,1}}\mathbb{Z} \times \cdots}_{\mathbb{Z}/d_1\mathbb{Z}} \times \cdots \times \underbrace{\mathbb{Z}/p_1^{e_{k,1}} \times \cdots}_{\mathbb{Z}/d_k\mathbb{Z}}$$

and grouping the same primes together gives the primary factor form.

To obtain the invariant factor form from the primary factor form, we will illustrate the idea with an example, as the notation becomes unwieldy.

Example 4.5.13 (Invariant factors from primary factors)**.** Suppose we want to find the primary factor decomposition of the group

$$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z} \times \mathbb{Z}/25\mathbb{Z}.$$

We group the primary factors together to obtain the $\mathbb{Z}/d_i\mathbb{Z}$. Since $d_1 \mid d_2 \mid \cdots \mid d_k$, we obtain the d_i's inductively from the largest one, taking the largest primary factor remaining for each prime.

That is, we would group them like this

$$\underbrace{\mathbb{Z}/2\mathbb{Z}}_{d_1} \times \underbrace{\mathbb{Z}/2\mathbb{Z}}_{d_2} \times \underbrace{\mathbb{Z}/4\mathbb{Z}}_{d_3} \times \underbrace{\mathbb{Z}/3\mathbb{Z}}_{d_3} \times \underbrace{\mathbb{Z}/5\mathbb{Z}}_{d_1} \times \underbrace{\mathbb{Z}/5\mathbb{Z}}_{d_2} \times \underbrace{\mathbb{Z}/25\mathbb{Z}}_{d_3},$$

thus the invariant factor form is

$$\underbrace{(\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z})}_{\mathbb{Z}/d_1\mathbb{Z}} \times \underbrace{(\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z})}_{\mathbb{Z}/d_2/Z} \times \underbrace{(\mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/25\mathbb{Z})}_{\mathbb{Z}/d_3\mathbb{Z}} \cong \mathbb{Z}/10\mathbb{Z} \times \mathbb{Z}/10\mathbb{Z} \times \mathbb{Z}/300\mathbb{Z},$$

the last isomorphism coming from the Chinese Remainder Theorem (Theorem 4.5.12).

Our final task is to prove the Structure theorem for finite p-groups. We will need a lemma first. The idea of the proof of the structure theorem for finite p-groups, and why the lemma is useful, is that we will inductively peel off the largest primary piece.

Lemma 4.5.14. *Let p be a prime and let A be a finite abelian p-group, and let $a \in A$ be an element of maximal order, say of order p_1^e. Let $A_1 = \langle a \rangle \leq A$, and let $\overline{b} \in A/A_1$ be an element of order p^r. Then there exists a representative $b \in A$ of $\overline{b}$ of order p^r.*

Proof. Let $b \in A$ be a representative of $\overline{b}$. Since $p^r\overline{b} = 0 \in A/A_1$, we know that $p^r b \in A_1$, so $p^r b = na_1$ for some $a_1 \in A_1$ and $n \geq 0$. Note that $\operatorname{ord}(\overline{b}) \leq \operatorname{ord}(b)$. If $n = 0$, then we are done, as then

$$p^r = \operatorname{ord}(\overline{b}) \leq \operatorname{ord}(b) \leq p^r,$$

and so b is the representative for which we are looking. Thus we may assume that $n = p^k\mu$ with $\gcd(p, \mu) = 1$, and so $\langle a_1\rangle = \langle \mu a_1\rangle = A_1$. Thus $\text{ord}(\mu a_1) = p^{e_1}$.

We now claim that b has order p^{r+e_1-k}. Indeed, $p^r b \neq 0$, and $p^r b = na_1 = p^k \mu a_1$ has order p^{e_1-k}. So $p^{r+e_1-k}b = p^{e_1}\mu a_1 = 0$, hence $\text{ord}(b) \leq p^{r+e_1-k}$. Moreover, if $p^x = \text{ord}(b)$, then $x \geq r$ as $p^r b \neq 0$. And if $x < r + e_1 - k$, then $0 \leq x - r \leq e_1 - k$, hence $p^{x-r} = \text{ord}(p^r b) = \text{ord}(p^k \mu a_1) = p^{e_1-k}$, which is a contradiction to $x < r + e_1 - k$. Thus we must have $r + e_1 - k \leq x$, and so

$$p^{r+e_1-k} \leq p^x = \text{ord}(b) \leq p^{r+e_1-k},$$

hence $\text{ord}(b) = p^{r+e_1-k}$, as claimed.

Finally, we note that since r_1 was assumed the largest order of an element of A, we have $r+e_1-k \leq e_1$, so $r \leq k$. Thus there is an element $c \in A_1$ such that $p^r c = p^r b$(for example take $c = p^{e_1-k}a_1$). Now let $a = b - c$. Note that since $c \in A_1$ we have $\overline{a} = \overline{b}$ so a is a representative for $\overline{b}$, and thus $\text{ord}(a) \geq p^r$. As $p^r a = p^r b - p^r c = 0$, $\text{ord}(a) = p^r$, as desired. □

We are now prepared to prove the Structure theorem for finite p-groups. Let us restate it here and then give a proof.

Theorem 4.5.15 (Structure Theorem for finite abelian p-groups)**.** *Let p be a prime number and let G be a finite abelian p-group. Then*

$$G \cong \mathbb{Z}/p^{e_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{e_s}\mathbb{Z}$$

with integers $e_1 \geq e_2 \geq e_s > 0$ and $e_1, \ldots, e_s$ uniquely determined.

Proof. Let G be a finite abelian p-group,a nd let $a_1 \in G$ be an element of largest order, say p^{e_1}. Let $A_1 = \langle a_1\rangle$. By induction,

$$G/A_1 \cong \underbrace{\mathbb{Z}/p^{e_2}\mathbb{Z}}_{\overline{A}_2} \times \cdots \times \underbrace{\mathbb{Z}/p^{e_s}\mathbb{Z}}_{\overline{A}_s}$$

with $e_2 \leq \cdots \geq e_s$. Let $\overline{a}_i$ generate $\overline{A}_i$. By Lemma 4.5.14, there is a representative $a_i \in A$ of $\overline{a}_i$ with $\text{ord}(a_i) = p^{e_i}$. Let $A_i = \langle a_i\rangle$.

We claim that $A \cong A_1 \times \cdots \times A_s$. Let $x \in A$, with class $\overline{x} \in A/A_1$. Then

$$\overline{x} = \sum_{i=2}^{s} m_i \overline{a}_i,$$

so

$$x - \sum_{i=2}^{s} m_i a_i \in A_1.$$

Say $x - \sum_{i=2}^{s} m_i a_i = m_1 a_1$. Then $x = \sum_{i=1}^{s} m_i a_i$, so $A = A_1 + A_2 + \cdots + A_s$. Let $0 = \sum_{i=1}^{s} m_i a_i$,a nd we may assume that $m_i < p^{e_i}$ as a_i has period p^{e_i}. So

$$\overline{0} = \sum_{i=2}^{s} m_i \overline{a}_i,$$

and since $A/A_1 \cong \overline{A}_2 \times \cdots \times \overline{A}_s$ we have $m_i = 0$ for $2 \leq i \leq s$. Hence $0 = m_1 a_1$, and since $m_i < p^{e_1}$, we have $m_1 = 0$. So $m_i = 0$ for all $1 \leq i \leq s$, and thus

$$A \cong A_1 \times \cdots \times A_s,$$

as claimed.

We now prove uniqueness, again by induction.

Suppose that

$$A \cong \mathbb{Z}/p^{e_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{e_s}\mathbb{Z} \cong \mathbb{Z}/p^{r_1}\mathbb{Z} \times \cdots \mathbb{Z}/p^{r_t}\mathbb{Z}$$

with $e_1 \geq e_2 \geq \cdots \geq e_s \geq 1$ and $r_1 \geq r_2 \geq \cdots \geq r_t \geq 1$. If all e_i and r_i are 1, then $|A| = p^s = p^t$, and so $s = t$. For the inductive step we may assume that some e_i and some r_i are at least 2, consider the subgroup $pA \leq A$ which is also a p-group of smaller order, so

$$pA \cong \mathbb{Z}/p^{e_1-1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{e_u-1}\mathbb{Z} \cong \mathbb{Z}/p^{r_1-1}\mathbb{Z} \times \cdots \mathbb{Z}/p^{r_v-1}\mathbb{Z},$$

where if $e_i = 1$ or $r_i = 1$ then the factor is not included. By induction $e_1 - 1 \geq \cdots \geq e_u - 1 \geq 1$ and $r_1 - 1 \geq \cdots r_v - 1 \geq 1$ are uniquely determined, so $u = v$ and for $e_i \geq 2$ and $r_i \geq 2$, we have $e_i = r_i$. Thus the tuples $(e_1, \ldots, e_s)$ and $(r_1, \ldots, r_t)$ can only differ where $e_i = 1 = r_i$. But now

$$|A| = p^{e_1+\cdots+e_u+\#\{e_i=1\}} = p^{e_1+\cdots+e_u+(s-u)} = p^{e_1+\cdots+e_u+\#\{r_i=1\}} = p^{e_1+\cdots+e_u+(t-u)},$$

and we immediately see that $s = t$ and $(e_1, \ldots, e_s) = (r_1, \ldots, r_t)$, as claimed. □

This concludes our proof of the Structure Theorem for finitely generated abelian groups, Theorem 4.0.6 and Theorem 4.0.8.

4.6. **History of the Structure Theorem for Finitely Generated Abelian Groups.** The structure theorem for finitely generated abelian groups has a long history. It was proven in past before group theory was even established (before groups were called groups!).

There were early cases proven by Gaußaround 1801, and we should note the Galois (credited with starting to formalize groups) lived 1811–1832. In the 1870's, Kronecker have a proof for finite abelian groups, though not in the language of group theory. Frobenius and Stickelberger in 1878 gave a proof for finite abelian groups using the language of group theory. And further in 1900, Poincaré gave a proof for finite abelian groups using matrices, inspired by geometric developments in computing what we would now call homology. In 1926, Emmy Noether gave a proof for finitely generated abelian groups, generalizing Poincaré's approach. The developments in Noether's proof led in large part to advances in algebra and geometry.

4.7. **Applications of the Structure theorem.** We will prove the following theorem.

Theorem 4.7.1. *Let F be a field, and let $G \leq F^*$ be a finite (abelian) subgroup. Then G is cyclic.*

Example 4.7.2. If $G \leq \mathbb{C}^*$ is a finite group, then $G \subset \{z \in \mathbb{C} \mid |z| = 1\}$, since otherwise powers of z would have larger and larger or smaller and smaller length, and G would be infinite! And so if $G \subset \{z \in \mathbb{C} \mid |z| = 1\}$ we've already seen that G is generated by the element that makes the smallest angle with the x-axis. Thus G is cyclic.

Remark 4.7.3. We note that the assumption that G be finite is required. For example, consider $\mathbb{Q}^*$, and show that it is not cyclic.

Remark 4.7.4. We note that the field assumption is also required. For example, consider $(\mathbb{Z}/8\mathbb{Z})^*$, which you should be able to show is not cyclic. Hint: it is a group of order 4 since

$$(\mathbb{Z}/8\mathbb{Z})^* = \{a \in \mathbb{Z}/8\mathbb{Z} \mid \gcd(a, 8) = 1\} = \{1, 3, 5, 7\}.$$

We will first need a fact which we will prove later.

Lemma 4.7.5. *Let F be a field, and $p(x) \in F[x]$ a polynomial $p(x) = a_n x^n + \cdots a_1 x + a_0$ with $a_n \geq 0$, so $\deg(p) = n$. Then $p(x)$ has at most n roots in F.*

Proof of Theorem 4.7.1. Since $G \leq F^*$ is finite and abelian, the Structure Theorem gives an isomorphism

$$G \xrightarrow{\varphi} \mathbb{Z}/d_1\mathbb{Z} \times \mathbb{Z}/d_{k-1}/Z \times \mathbb{Z}/d_k\mathbb{Z}$$

with $d_1 \mid d_2 \mid \cdots \mid d_{k-1} \mid d_k$. Let $n = d_k$. We claim that $G \cong Z/n\mathbb{Z}$. Consider the polynomial $x^n - 1 \in F[x]$. For $x \in G$, we have

$$\varphi(x) = (a_i, \ldots, a_{k-1}, a_k)$$

with $a_i \in \mathbb{Z}/d_i\mathbb{Z}$. Now

$$\varphi(x^n) = n(a_i, \dots, a_{k-1}, a_k) = (\frac{n}{d_1}d_1a_1, \dots, \frac{n}{d_k}d_ka_k) = (0, \dots, 0),$$

and so $x^n = \varphi^{-1}(0, \dots, 0) = 1 \in F^*$. Thus $x^n - 1$ for any $x \in G$. But $x^n - 1$ has $\leq n$ roots, so $|G| \leq n$. However, we have $|G| = d_1 \cdots d_k = d_1 \cdots d_2 \cdots d_{k-1} \cdot n \leq n$, and so

$$d_1 = 1 = d_2 = \cdots = d_{k-1},$$

and $G \cong \mathbb{Z}/n\mathbb{Z}$ as claimed. □

5. Group Actions

Group actions are how we understand many groups, and the things on which they act. We'll begin with a few examples before giving a precise definition of what we mean by a "group action". Intuitively, think of group actions as the thing that groups were designed to do, to act as symmetries.

Example 5.0.1. The group $\mathrm{O}(n)$ acts on $\mathbb{R}^n$. Let $A \in \mathrm{O}(n)$, then we have a symmetry of $\mathbb{R}^n$ given by the map associated to the matrix A. That is, A defines a map $A : \mathbb{R}^n \to \mathbb{R}^n, x \mapsto Ax$. We note that $I_n x = x$ and for $A, B \in \mathrm{O}(n)$ we have $(AB)x = A(Bx)$.

Group actions generalize this.

Definition 5.0.2. Let X be a set, and G a group. A *(left) group action of G on X is a map*

$$G \times X \to X, (g, x) \mapsto g.x$$

satisfying

(1) $\mathrm{id}_G.x = x$ for every $x \in X$
(2) $(gh).x = g.(h.x)$ for every $x \in X$ and every $g, h \in G$.

Such a set X is called a G-set, and we write $G \curvearrowright X$ to mean "G acts on X".

Remark 5.0.3. Fixing $g \in G$, we have a map

$$g : X \to X, x \mapsto g.x,$$

which is a bijection with inverse g^{-1}.

Exercise 5.0.4. Check that the map

$$g : X \to X, x \mapsto g.x$$

above is indeed a bijection.

Exercise 5.0.5. A group action $G \times X \to X$ is equivalent to a group homomorphism

$$\varphi : G \to \mathrm{Perm}(X) \coloneqq \{\text{bijections } X \to X\}.$$

Work out the details of the map

$$\{\text{Group actions } G \times X \to X\} \overset{1:1}{\longleftrightarrow} \{\text{group homomorphisms } G \to \mathrm{Perm}(X)\}.$$

Let's see some examples of group actions.
A familiar group action is the action of the permutation group S_n on the set $\{1, \dots, n\}$.

Example 5.0.6 ($S_n \curvearrowright \{1, \dots, n\}$)**.** The group $S_n \coloneqq \mathrm{Perm}(\{1, \dots, n\})$ acts on the set $\{1, \dots, n\}$ in the natural way, as permutations.

Example 5.0.7 (Actions of $\mathrm{O}(n)$)**.** We've seen that $\mathrm{O}(n)$ acts on $\mathbb{R}^n$ simply by applying the matrix. Since matrices in $\mathrm{O}(n)$ also preserve length, they send vectors of length 1 to vectors fo length 1. Hence $\mathrm{O}(n) \curvearrowright$ unit sphere centered at $0 \in \mathbb{R}^n$.

From this action, there are a few more induced actions. Namely

- $\mathrm{O}(n) \curvearrowright \{$ pairs of opposite points on unit sphere$\}$
- $\mathrm{O}(n) \curvearrowright \{$ equators of unit sphere$\}$

Example 5.0.8 ($G \curvearrowright G$ by left translation)**.** Here our group will be a group G, and our set will be $X = G$. The group action is simply multiplication in the group

$$G \times X \to X, (g, x) \mapsto gx,$$

which gives a group action as the group operation is associative.

Example 5.0.9 ($\mathbb{R}$ acts on itself by left translation)**.** For example, if we take the group $(\mathbb{R}, +)$, then shifting by a number $a \in \mathbb{R}$ defines a bijection $\mathbb{R} \to \mathbb{R}, x \mapsto a + x$, which is left translation by a.

Example 5.0.10 ($G \curvearrowright G$ by conjugation)**.** Here our group will again be a group G, and our set will again be $X = G$, but the action is different than left translation. Instead, we will conjugate. That is, the conjugation action of a group on itself is given by

$$G \times G \mapsto G, (g, x) \mapsto g.x \coloneqq gxg^{-1}.$$

Exercise 5.0.11. Check that the conjugation action is indeed a group action. Hint: $(gh)^{-1} = h^{-1}g^{-1}$.

Example 5.0.12 (Group action on cosets)**.** Let $H \leq G$ be a subgroup, and denote the set of (left) cosets of H by G/H. Then $G \curvearrowright G/H$ by left translation again, namely

$$g.(aH) = (ga)H.$$

5.1. **Orbits and Stabilizers.** Given a group action, there are some naturally associated sets and subgroups.

Definition 5.1.1. Let $G \curvearrowright X$. For $x \in X$, the *stabilizer of* x is the set

$$G_x \coloneqq \{g \in G \mid g.x = x\}.$$

Proposition 5.1.2. *Let $G \curvearrowright X$ and $x \in X$. Then G_x is a subgroup of G.*

Proof. We must check the G_x contains id_G, and is closed under multiplication and inverses.

By definition, $\mathrm{id}_G.x = x$, so $\mathrm{id}_G \in G_x$.

Now let $g, h \in G_x$. We compute

$$g^{-1}.x = g^{-1}.(g.x) = (g^{-1}g).x = \mathrm{id}_G.x = x,$$

thus $g^{-1} \in G_x$. We also compute

$$(gh).x = g.(h.x) = g.x = x,$$

whereby $gh \in G_x$.

Hence G_x is indeed a subgroup of G. □

Example 5.1.3. Let $X = \{1, 2, 3\}\}$, and let $S_3 \curvearrowright X$ in the natural way.

Then $(S_3)_3 = \{\sigma \in S_3 \mid \sigma(3) = 3\} = \{(1), (12)\}$.

Example 5.1.4. Let $G \curvearrowright G$ by conjugation, and let $x \in G$. Then

$$G_x = \{g \in G \mid g.x = x\} = \{g \in G \mid gxg^{-1} = x\} = Z_G(x),$$

the *centralizer of x in G*.

Example 5.1.5. Let $G \curvearrowright G/H$. The stabilizer of the coset H is

$$G_H = \{g \in G \mid gH = H\} = H.$$

The stabilizer of another coset aH is

$$\begin{aligned} G_{aH} &= \{g \in G \mid gaH = aH\} \\ &= \{g \in G \mid a^{-1}gaH = H\} \\ &= \{g \in G \mid a^{-1}ga \in H\} \\ &= \{g \in G \mid g \in aHa^{-1}\} \\ &= aHa^{-1} \\ &= aG_Ha^{-1}. \end{aligned}$$

This turns out to be true in general.

Lemma 5.1.6. *Let $G \curvearrowright X$. For $x \in X$ and $g \in G$, we have*

$$G_{g.x} = gG_xg^{-1}.$$

Proof. We simply unwind the definitions.

$$\begin{aligned} h \in G_{g.x} &\iff h.(g.x) = g.x \\ &\iff g^{-1}.(h.(g.x)) = x \\ &\iff (g^{-1}hg).x = x \\ &\iff g^{-1}hg \in G_x \\ &\iff h \in gG_xg^{-1}. \end{aligned}$$

□

Definition 5.1.7. Let $G \curvearrowright X$, and let $x \in X$. The *orbit of x* is

$$Gx = \mathcal{O}_x \coloneqq \{y \in X \mid y = g.x \text{ for some } g \in G\}.$$

Example 5.1.8. Let $\mathrm{O}(2) \curvearrowright \mathbb{R}^2$. For $x \in \mathbb{R}^2$, the orbit

$$\mathcal{O}_x = \{y \in \mathbb{R}^2 \mid ||y|| = ||x||\},$$

since $\mathrm{O}(2)$ preserves lengths and has all the rotation matrices. In fact, we can see that this group action breaks up $\mathbb{R}^2$ into orbits that look like concentric circles of different radii and an orbit which is just the origin $0 \in \mathbb{R}^2$. So the orbits of $\mathrm{O}(2) \curvearrowright \mathbb{R}^2$ do in fact look like orbits!

Let's find some stabilizers!

- As mentioned before, we have $\mathcal{O}_0 = \{0\}$, and since every $A \in \mathrm{O}(2)$ preserves the origin, we have $G_0 = \mathrm{O}(2)$.
- To find $G_{e_1} = \{A \in \mathrm{O}(2) \mid Ae_1 = e_1\}$, recall the for a matrix A, Ae_1 is the first column of A. Thus for $A \in G_{e_1}$ we have

  $$A = \begin{pmatrix} 1 & a \\ 0 & b \end{pmatrix},$$

  and since $A \in \mathrm{O}(2)$ we have $A^t = A^{-1}$, and so

  $$\begin{pmatrix} 1 & a \\ 0 & b \end{pmatrix}\begin{pmatrix} 1 & 0 \\ a & b \end{pmatrix} = \begin{pmatrix} 1+a^2 & ab \\ ab & b^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

thus $a = 0$ and $b = \pm 1$. Hence

$$G_{e_1} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\}.$$

We note that G_{e_1} is exactly the reflections through the e_1-axis.

- Similarly, for $x \in \mathbb{R}^2$, we have $G_x = \{I_2, \text{ reflection across line of } x\}$.

In these examples, we see notice a few things seem to be similar. Namely,

- The orbits of a group action partition X.
- The sizes of orbits and of stabilizers are inversely related. If an orbit is small then the stabilizer is large, and if the orbit is large then the stabilizer is small.

We'll make these observations very precise shortly.

Lemma 5.1.9. *Let $G \curvearrowright X$. Two orbits of are either disjoint or equal.*

Proof. Let $\mathcal{O}_x$ and $\mathcal{O}_y$ be two orbits. If the two orbits are disjoint, then we are done. If $z \in \mathcal{O}_x \cap \mathcal{O}_y$, then

$$z = g.x = h.y$$

and so

$$y = h^{-1}g.x \in \mathcal{O}_x, \text{ and } x = g^{-1}h.y \in \mathcal{O}_y,$$

so $\mathcal{O}_x = \mathcal{O}_y$. □

Since every element of X is in some orbit (namely $\mathcal{O}_x$), the orbits partition X.

Lemma 5.1.10. *Let $G \curvearrowright X$. Then X is a disjoint union of orbits, $X = \bigsqcup_{i \in I} \mathcal{O}_{x_i}$ where $\{x_i\}_{i \in I}$ is a set of representatives for the orbits.*

Lemma 5.1.11 (Class equation for orbits)**.** *Let $G \curvearrowright X$ with X a finite set. Let $X = \bigsqcup_{i \in I} \mathcal{O}_{x_i}$ be the partition of X into disjoint orbits. Then $|X| = \sum_{i \in I} |\mathcal{O}_{x_i}|$.*

Proof. Since the set X is finite, and the union is disjoint, this follows by taking cardinalities. □

Example 5.1.12. Let $H \leq G$ be a subgroup, and let $H \curvearrowright G$ by left translation, that is $h.g = hg$. The orbits are $\mathcal{O}_g = \{hg \in G \mid h \in H\} = Hg$, which are just the cosets of H. And the set of orbits is just the set of cosets $H\backslash G$.

Recall 5.1.13. Recall that the cosets of a subgroup partition the group.

Hence, via group actions or group theory, we have $G = \bigsqcup_{i \in I} Hg_i$. Note that as sets, all the cosets are bijective. Indeed, $H \to Hg, h \mapsto hg$ is a bijection.

Now if G is finite, then all of the cosets are have the same size, namely $|H|$, and we have

$$|G| = \sum_{Hg_i \in H\backslash G} |Hg_i| = \sum_{Hg_i \in H\backslash G} |H| = |G/H| \cdot |H|,$$

which is Lagrange's Theorem.

More generally, much of this holds true if a set X is endowed with some equivalence relation $\sim$, as then equivalence classes partition X.

Definition 5.1.14. An equivalence relation $\sim$ on a set X is a relation $\sim$ between elements of X such that for all $x, y, z \in X$

- $x \sim x$ (reflexive)
- $x \sim y \implies y \sim x$ (symmetric)

- if $x \sim y$ and $y \sim z$, then $x \sim z$ (transitive).

Definition 5.1.15. For a set X with an equivalence relation $\sim$. The *equivalence class of* $x \in X$ is

$$[x] \coloneqq \{y \in X \mid x \sim y\}.$$

Theorem 5.1.16. *Let X be a set with an equivalence relation. Then the equivalence classes partition X.*

Exercise 5.1.17. Let $G \curvearrowright X$. Define an equivalence relation on X by

$$x \sim y \iff \exists g \in G \text{ such that } y = g.x$$

and verify that this does indeed define an equivalence relation.

Definition 5.1.18. An action $G \curvearrowright X$ is called *transitive* if X is a single orbit, i.e. $X = \mathcal{O}_x$ for some $x \in X$.

Remark 5.1.19. Observe that a single orbit $\mathcal{O}_x$ is a transitive G-set. Since the orbits partition X, any G-set is the disjoint union of transitive G-sets.

We now turn our attention to the relationship between orbits and stabilizers.

Example 5.1.20. Let $G \curvearrowright X$. We'll consider some extreme cases of orbits and stabilizers.

- Suppose that every element fixes $x \in X$, so $G_x = G$, then $\mathcal{O}_x = \{x\}$
- Suppose that only the identity fixes $x \in X$, so $G_x = \{\mathrm{id}_G\}$, then $G \to \mathcal{O}_x, g \mapsto g.x$ is a bijection.

Theorem 5.1.21 (Orbit-Stabilizer Theorem)**.** *Let $G \curvearrowright X$ and $x \in X$. Then*

$$\varphi_x : G/G_x \to \mathcal{O}_x,\ gG_x \mapsto g.x$$

is a bijection.

Proof. We first need to check that the map φ_x is even well-defined, since a coset gG_x can have many representatives, and the map φ_x is defined in terms of the representative. So we must show that if $gG_x = hG_x$, then $g.x = h.x$. We simply unwind the definitions

$$\begin{aligned} gG_x = hG_x &\iff h^{-1}gG_X = G_x \\ &\iff h^{-1}g \in G_x \\ &\iff h^{-1}g.x = x. \end{aligned}$$

In particular,

$$g.x = hh^{-1}g.x = h.x,$$

as we wanted to show. Thus φ_x is well-defined.

We can define an inverse map

$$G/G_x \underset{\varphi_x}{\overset{\psi_x}{\rightleftarrows}} \mathcal{O}_x\ ,$$

$$gG_x \xmapsto{\varphi_x} g.x$$

$$gG_x \xmapsto{\psi_x} g.x$$

though we must again check that $\psi_x : \mathcal{O}_x \to G/G_x$ is well-defined as the orbit $\mathcal{O}_x$ can have many different elements of G that send x to the same point of the orbit$\mathcal{O}_x$. So suppose that $y = h.x = g.x \in \mathcal{O}_x$ for some elements $h, g \in G$. Then $h^{-1}g.x = x$, so $h^{-1}g \in G_x$, and so $h^{-1}gG_x = G_x$, thus $gG_x = hG_x$ and

$$\psi_x(g.x) = gG_x = hG_x = \psi_x(h.x),$$

so ψ_x is well-defined.

Clearly be definition φ_x and ψ_x are inverses, so $\varphi_x : G/G_x \to \mathcal{O}_x$ is a bijection. □

Remark 5.1.22. If X is a transitive G-set and $x \in X$, then $G/G_x \cong X$. Moreover, both G/G_x and X have actions by G.

$$G \curvearrowright G/G_x \text{ by } g.(hG_x) = \underbrace{(gh)}G_x$$

and

$$G \curvearrowright X \text{ by } g. \underbrace{hx}_{\text{since } X = GX} = \underbrace{gh}\, x$$

and so

$$g.\varphi_x(hG_x) = g.(h.x) = (gh).x = \varphi(ghG_x) = \varphi_x(g.(hG_x)),$$

so the map φ_x respects the two G-actions.

Definition 5.1.23. Let X and Y be G-sets. A map $\varphi : X \to Y$ is called *G-equivariant* if $\varphi(g.x) = g.\varphi(x)$. An *isomorphism of G-sets* is a G-equivariant map $\varphi : X \to Y$ with a G-equivariant inverse $\psi : Y \to X$, i.e. $\varphi \circ \psi = \mathrm{id}_Y$ and $\psi \circ \varphi = \mathrm{id}_X$.

Exercise 5.1.24. Show that the map $\varphi_x : G/G_x \to \mathcal{O}_x$ is an isomorphism of G-sets.

Example 5.1.25. Let $X = \{x \in \mathbb{R}^2 \mid ||x|| = 1\}$ be the unit circle in $\mathbb{R}^2$ and let $Y = \mathbb{R}^2$. Both X and Y are O(2)-sets, and the inclusion $X \hookrightarrow Y$ is O(@)-equivariant, but clearly $X \ncong Y$.

5.2. **Applications of class equation and Orbit-Stabilizer.**

Proposition 5.2.1. *Let X be a finite set, and $G \curvearrowright X$. Then $|X| = \sum_{i=1}^{r}[G : G_{s_i}]$ where $\{s_i\}_{i=1}^{r}$ represent the G-orbits.*

Proof. Let $X = \bigsqcup_{i=1}^{r} \mathcal{O}_{s_i}$. By the Orbit-Stabilizer theorem $\mathcal{O}_{s_i} \cong G/G_{s_i}$. So

$$|\mathcal{O}_{s_i}| = |G/G_{s_i}| = [G : G_{s_i}],$$

and thus

$$|X| = \sum_{i=1}^{r} |\mathcal{O}_{s_i}| = \sum_{i=1}^{r} [G : G_{s_i}],$$

as claimed. □

Example 5.2.2. Let $G \curvearrowright G$ by conjugation. An orbit is a conjugacy class, that it

$$\mathcal{O}_h = \{ghg^{-1} \mid g \in G\}.$$

For

$$h \in Z(G) \coloneqq \{g \in G \mid gh = hg \text{ for all } h \in G\},$$

we have $\mathcal{O}_h = \{h\}$ and $G_h = G$.

Proposition 5.2.3 (Also definition)**.** *For $h \notin Z(G)$, the* centralizer subgroup

$$Z_G(h) \coloneqq \{g \in G \mid gh = hg\} \leq G$$

is a subgroup of G, and is the stabilizer of h under the conjugation action.

Proof. Since stabilizer subgroups are subgroups, it suffices to show that $Z_G(h) = G_h$. This is a simple unwinding of the definitions

$$\begin{aligned} G_h &= \{g \in G \mid ghg^{-1} = h\} \\ &= \{g \in G \mid gh = hg\} \\ &= Z_G(h). \end{aligned}$$

□

Proposition 5.2.4. *Let G be a finite group and let $G \curvearrowright G$ by conjugation. Then*

$$|G| = |Z(G)| + \sum_{i=1}^{m} [G : Z_G(h_i)]$$

where $h_1, \ldots, h_m$ represent the conjugacy classes with more than one element.

Proof. We apply the class equation (Lemma 5.1.11) to the conjugation action.

$$|G| = \sum_{i=1}^{r} |\mathcal{O}_{h_i}| = \sum_{\substack{i=1 \\ h_i \notin Z(G)}}^{m} |\mathcal{O}_{h_i}| + \underbrace{\sum_{\substack{i=m+1 \\ h_i \in Z(G)}}^{r} \overbrace{|\mathcal{O}_{h_i}|}^{=1 \text{ as } h_i \in Z(G)}}_{=|Z(G)|} = \sum_{i=1}^{m} [G : Z_G(h_i)] + |Z(G)|,$$

where the last equality comes from the Orbit-Stabilizer theorem and the fact that $[G : H] = |G/H|$. □

We're now ready to prove a few results about p-groups using group actions that will be useful for us later.

Theorem 5.2.5. *Let p be a prime and G a group of order p^n. Then $Z(G) \neq \{1\}$.*

Proof. We will show that $|Z(G)| \geq p$ (≥ 2), which implies that $Z(G)$ cannot be the trivial subgroup.
From Proposition 5.2.4, we have

$$|G| = |Z(G)| + \sum_{i=1}^{m} [G : G_{h_i}]$$

where $G_{h_i} \lneq G$. As $p^n = |G| = [G : G_{h_i}] \cdot \underbrace{|G_{h_i}|}_{<p^n}$, p must divide $[G : G_{h_i}]$. Hence

$$p^n = |G| = |Z(G)| + \underbrace{[G : G_{h_1}] + \cdots + [G : G_{h_m}]}_{\text{divisible by } p},$$

so p must also divide $|Z(G)|$. As $1 \in Z(G)$, we have $|Z(G)| \geq 1$, and so $|Z(G)| \geq p$, as claimed. □

Lemma 5.2.6. *Suppose that $G/Z(G)$ is cyclic. Then G is abelian.*

Proof. Let $aZ(G)$ generate $G/Z(G)$. Let $g, h \in G$, say $g \in a^m Z(G)$ and $h \in a^n Z(G)$. Thus we may write

$$g = a^m x, \ h = a^n y \ \text{ for some } x, y \in Z(G).$$

We now compute

$$\begin{aligned} gh &= (a^m x)(a^n y) \\ &= a^m(xa^n)y \\ &= a^m(a^n x)y \\ &= (a^{m+n})(xy) \\ &= (a^n a^m)(yx) \\ &= a^n(a^m y)x \\ &= a^n(ya^m)x \\ &= (a^n y)(a^m x) \\ &= hg, \end{aligned}$$

thus $gh = hg$ for all $g, h \in G$, and so G is abelian. □

Corollary 5.2.7. *Let p be prime and G be a group of order p^2. Then G is abelian.*

Proof. By Theorem 5.2.5, $|Z(G)| \geq p$. Hence as

$$p^2 = |G| = [G : Z(G)]|Z(G)|,$$

$|Z(G)| = p$ or p^2. If $|Z(G)| = p^2$, then $G = Z(G)$, and G is abelian.

Thus suppose that $|Z(G)| = p$. Then $Z(G)$ and $G/Z(G)$ are both groups of order p, hence cyclic. The result now follows from Lemma 5.2.6. □

Remark 5.2.8. In class, the proof of the Corollary included just the proof of Lemma 5.2.6.

5.3. **Burnside's Counting Lemma.**

Remark 5.3.1. This is somewhat of a misnomer, as Burnside was not the firs to prove this lemma.

The idea behind Burnside's lemma is that one may count the number of orbits by understanding how the group acts on the set.

Definition 5.3.2. Let $G \curvearrowright X$ and let $g \in G$. The fixed set of g is

$$\operatorname{Fix}(g) \coloneqq \{x \in X \mid g.x = x\} \subseteq X.$$

Theorem 5.3.3 (Burnside's Counting Lemma)**.** *Let G be a finite group and X a finite G-set. Let X/G denote the set of orbits. Then*

$$|X/G| = \frac{1}{|G|} \sum_{g \in G} |\operatorname{Fix}(g)|.$$

Proof. The proof follows an ingenious double-counting argument, where we count one set in two different ways.

Define

$$S \coloneqq \{(g, x) \in G \times X \mid g.x = x\}.$$

We will count S by fixing g and then by fixing x.

Fixing g, we have $|S| = \sum_{g \in G} |\{x \in X \mid g.x = x\}| = \sum_{g \in G} |\operatorname{Fix}(g)|$.

Fixing x, we have $|S| = \sum_{x \in X} |\{g \in G \mid g.x = x\}| = \sum_{g \in G} |G_x|$. As $X = \bigsqcup_{\mathcal{O} \in X/G} \mathcal{O}$, we have

$$|S| = \sum_{\mathcal{O} \in X/G} \sum_{x \in \mathcal{O}} |G_x|.$$

Now by Lemma 5.1.6, for $x, y \in \mathcal{O}$, the stabilizer subgroups G_x and G_y are conjugate subgroups of G, hence $|G_x| = |G_y|$. Thus

$$\begin{aligned}
|S| &= \sum_{\mathcal{O} \in X/G} \sum_{x \in \mathcal{O}} |G_{x_0}| \text{ for some fixed } x_0 \text{ representing } \mathcal{O} \\
&= \sum_{\mathcal{O} \in X/G} \underbrace{|\mathcal{O}_{x_0}||G_{x_0}|}_{=|G|} \\
&= \sum_{\mathcal{O} \in X/G} |G| \\
&= |G| \cdot |X/G|.
\end{aligned}$$

Thus we have

$$\sum_{g \in G} |\operatorname{Fix}(g)| = |S| = |G| \cdot |X/G|,$$

and dividing by $|G|$ yields the result. □

5.3.1. *Applications of Burnside's Lemma and Fermat's little theorem.*

Example 5.3.4 (Induced group action on maps)**.** Let $G \curvearrowright X$ and let Y be a set. Let

$$\operatorname{Map}(X, Y) \coloneqq \{ \text{ maps } f : X \to Y \}.$$

Then there is an induced action $G \curvearrowright \operatorname{Map}(X, Y)$ given by

$$g.f = (x \mapsto f(g^{-1}.x)).$$

We now check that this is indeed a group action. We compute

$$\begin{aligned}
(g.(h.f))(x) &= g.f(h^{-1}.x) \\
&= f(h^{-1}.(g^{-1}.x)) \\
&= f(h^{-1}g^{-1}.x) \\
&= f((gh)^{-1}.x) \\
&= (gh).f(x).
\end{aligned}$$

Example 5.3.5 (Fermat's Little Theorem)**.** Let p be prime, and let $G = \mathbb{Z}/p\mathbb{Z}$ and $X = \mathbb{Z}/p\mathbb{Z}$, with $G \curvearrowright X$ by left translation. Let Y be a finite set of size $|Y| = a$. The action of $G \curvearrowright \operatorname{Map}(X, Y)$ is given by

$$g.f(x) = f(x - g).$$

We now compute Burnside's Lemma for this action.

Let us compute $|\operatorname{Fix}(g)|$ for the elements of G.

- For $g = 0$, $x - g = x$, so $g.f = f$, and so $|\operatorname{Fix}(0)| = |\operatorname{Maps}(X, Y)| = a^p$, where the last equality comes from the fact that a map $X \to Y$ is obtained by a choice of an element of Y for every element of X.
- Now suppose $0 \neq g \in G$. Observe that for $f \in \operatorname{Fix}(g)$, we have $f = g^{-1}g.f = g^{-1}.f$ and so $f \in \operatorname{Fix}(g^{-1})$ as well. Thus we have

  $$f = g.f = g^2.f = \cdots = g^{-1}.f = g^{-2}.f = \cdots,$$

  hence

  $$f(x) = f(x + g) = f(x + 2g) = \cdots = f(x + kg) \text{ for all } k \in \mathbb{Z}.$$

  Hence as $g \neq 0$, and p is prime $G = \langle g \rangle$, so $f(x) = f(x+h)$ for all $h \in G$ and so $f(x) \in \operatorname{Fix}(g)$ is constant. Hence for $g \neq 0$, $|\operatorname{Fix}(g)| = a$.

By Burnside's lemma (Theorem 5.3.3),

$$\frac{1}{|G|}\sum_{g\in G}|\operatorname{Fix}(g)| = |X/G| \in \mathbb{Z},$$

thus as

$$\begin{aligned}\frac{1}{|G|}\sum_{g\in G}|\operatorname{Fix}(g)| &= \frac{1}{p}(\underbrace{a^p}_{|\operatorname{Fix}(0)|} + (p-1)\underbrace{a}_{|\operatorname{Fix}(g)|, g\neq 0})\\ &= \frac{1}{p}(a^p - a) \in \mathbb{Z},\end{aligned}$$

$a^p - a$ is divisible by p, so

$$a^p - a \equiv = \mod p.$$

And we have proved Fermat's Little Theorem.

Theorem 5.3.6 (Fermat's Little Theorem)**.** *Let p be prime. Then $a^p \equiv a \mod p$ for all $a \in \mathbb{Z}$.*

5.4. **Further applications of group actions and Cauchy's Theorem.** We now give a proof of the full Cauchy's theorem, for arbitrary finite groups, not just for abelian groups as we did in Theorem 4.5.5. There are many proofs of Cauchy's theorem, though this one using group actions is particularly slick. We'll need a few useful facts first.

Definition 5.4.1. Let $G \curvearrowright X$. An element $x \in X$ is called a *fixed point* if $\mathcal{O}_x = \{x\}$.

Theorem 5.4.2. *Let p be prime and G be a finite group of order $|G| = p^k$ for some k. Suppose that $G \curvearrowright X$ on a finite set X. Then*

$$|X| \equiv |\{ \textit{ fixed points}\}| \mod p.$$

Proof. This is a straightforward application of the Orbit-Stabilizer theorem. We have

$$|X| = \sum_{i=1}^{r}|\mathcal{O}_{x_i}| = \sum_{i=1}^{r}[G : G_{x_i}].$$

Since $p^k = |G| = [G : G_{x_i}]|G_{x_i}|$, p divides $[G : G_{x_1}]$ unless $[G : G_{x_i}] = 1$ in which case x_i is a fixed point. So we have

$$|X| = \sum_{i=1}^{r}[G : G_{x_i}] = |\{\text{fixed points}\}| + \sum_{\substack{x_i\\ |\mathcal{O}_{x_i}|>1}} \underbrace{[G : G_{x_i}]}_{p \text{ divides}},$$

which reads $\mod p$ as

$$|X| \equiv |\{\text{fixed points}\}| \mod p. \qquad \square$$

Corollary 5.4.3. *Let G be a finite p-group acting on a finite set X. Then,*

(a) *if $|X| \not\equiv 0 \mod p$, there is at least one fixed point;*
(b) *if $X \equiv 0 \mod p$, then $|\{$fixed points$\}|$ is divisible by p (maybe 0).*

Proof. From Theorem 5.4.2, $|X| \equiv |\{\text{fixed points}\}| \mod p$. Thus

(a) $|\{\text{fixed points}\}| \not\equiv 0 \mod p$, hence $|\{\text{fixed points}\}| \geq 1$;
(b) $|\{\text{fixed points}\}| \equiv 0 \mod p$, hence p divides $|\{\text{fixed points}\}|$. $\square$

Theorem 5.4.4 (Cauchy's Theorem)**.** *Let G be a finite group, and suppose p is a prime dividing $|G|$. Then G has an element of order p.*

Proof. Consider the set

$$X \coloneqq \{(g_1, \dots, g_p) \in G^p \mid g_1 \cdots g_p = \mathrm{id}_G\}.$$

Since $g_1^{-1} = g_2 \cdots g_p$,

$$g_1 g_2 \cdots g_p = g_1 g_1^{-1} = g_1^{-1} g_1 = g_2 \cdots g_p g_1,$$

and we have an action of $\mathbb{Z}/p\mathbb{Z} \curvearrowright X$ by cyclically permuting the entries.

Again, since $g_1 = (g_2 \cdots g_p)^{-1}$, $|X| = |G|^{p-1}$ as we cam choose the last $p-1$ entries freely. Thus $\mathbb{Z}/p\mathbb{Z} \curvearrowright X$ is an action of a p-group on a finite set and since $|X| = |G|^{p-1} \equiv 0 \mod p$, Corollary 5.4.3 shows that p divides $|\{\text{fixed points}\}|$. Since $(\mathrm{id}_G, \dots, \mathrm{id}_G) \in X$, which is clearly a fixed point, we see that $|\{\text{fixed points}\}| \geq p \geq 2$.

Note that a fixed point $(g_1, g_2, \dots, g_p) \in X$ satisfies

$$(g_1, g_2 \dots, g_p) = (g_2, \dots, g_p, g_1) = \cdots = (g_p, g_1, g_2 \dots, g_{p-1}),$$

and thus $g_i = g$ for some fixed $g \in G$. Hence a fixed point $(g, \dots, g) \in X$ is exactly given by an element $g \in G$ such that $g^p = 1$. Since $|\{\text{fixed points}\}| \geq 2$, there is some $\mathrm{id}_G \neq g \in G$ such that $g^p = 1$, and so $\mathrm{ord}(g) = p$, as desired. □

Remark 5.4.5. While we will now depart from group actions, I want to stress that we have just scratched the surface. Group actions have become an integral part of modern mathematics, appearing in our understanding of geometry, number theory, analysis, combinatorics, differential equations, and many other fields. In many contexts, having a group action around is a useful way of "rigidifying" the structures you consider, since the group action moves things around. And many questions arise by asking what the set of orbits looks like, or finding the fixed points or invariant structures of a group action.

6. Sylow Theorems

Remark 6.0.1. This section is somewhat historical. In the 19$^{\text{th}}$ and 20$^{\text{th}}$ centuries, there was a big push to understand finite groups. This was done in parts, first by breaking up finite groups into "simple" pieces following the philosophy of Jordan and Hölder, and then understanding the "simple" pieces.

Definition 6.0.2. A group G is called *simple* if the only normal subgroups of G are $\{\mathrm{id}_G\}$ and G itself.

There is a way of breaking up a group into simple pieces, this is the Jordan–Hölder Theorem, which we unfortunately did not have time to cover. The idea is that a finite group G can be built as a tower

$$1 = G_0 \trianglelefteq G_1 \trianglelefteq \cdots \trianglelefteq G_n = G,$$

with $G_i \trianglelefteq G_{i+1}$ a normal subgroup and G_{i+1}/G_i a simple group. The Jordan–Hölder theorem states that for finite groups such a tower exists and that for a fixed group G the simple groups G_{i+1}/G_i that appear are uniquely determined by G (though the groups G_i themselves need not be and the quotients may appear in different orders), much like the structure theorem for finitely generated abelian groups.

The trouble with understanding *all* finite groups, now, is to first understand all finite simple groups and then somehow stitch the simple pieces back together. The first of these problems, has occupied mathematicians for the greater part of the 20$^{\text{th}}$-century, culminating in a complete classification fo all finite simple groups in 2004, spanning over 150 years since the work of Jordan and Hölder in the mid-19$^{\text{th}}$-century, and by the end, the work of hundreds of mathematicians.

The second problem, of stitching the simple pieces together, called the "extension problem" for finite groups is widely considered unsolvable, too complicated for a reasonable statement to be made or algorithm.

Nevertheless, in the understanding of finite groups, especially groups of small order, the theorems of Sylow, which we will now discuss and prove, were instrumental.

Definition 6.0.3. Let p be prime. A group G is called a p-group if $|G| = p^k$ for some k.

Exercise 6.0.4. The following proposition follows readily from Lagrange's theorem and Cauchy's theorem

Proposition 6.0.5. *A finite group G is a p-group if and only if every element has order a power of p.*

In proving the structure theorem for finite p-groups, we made a lot of progress by peeling off a subgroup generated by an element of largest order. For arbitrary finite groups, we can try to do something similar and find a subgroup of largest prime order.

Definition 6.0.6. Let p be prime, and G a finite group of order $|G| = p^a b$ with $\gcd(p, b) = 1$. A subgroup $S \leq G$ of order $|S| = p^a$ is called a *Sylow p-subgroup*.

We'll not state and prove the Sylow Theorems, a collection of statement about the existence and structure of Sylow p-subgroups.

Theorem 6.0.7 (Sylow Theorems)**.** *Let p be prime, G a finite group of order $|G| = p^a b$ with $\gcd(p, b) = 1$. Then,*

(a) *G admits at least one Sylow p-subgroup, say $\mathcal{P}$;*
(a') *if $Q \leq G$ is a subgroup of order p^m, then Q is conjugate to a subgroup of a Sylow p-subgroup, i.e. there is some G and a Sylow p-subgroup $\mathcal{P}$ such that $gQg^{-1} \leq \mathcal{P}$;*
(b) *all Sylow p-subgroups are conjugate;*
(c) *let $n_p \coloneqq \#\{\text{Sylow } p\text{-subgroups}\}$ and let $\mathcal{P}$ be a Sylow p-subgroup, then*

$$n_p = [G : N_G(\mathcal{P})] \mid |G|$$

where $N_G(\mathcal{P}) \coloneqq \{g \in G \mid g\mathcal{P}g^{-1} = \mathcal{P}\}$ is the normalizer of $\mathcal{P}$, moreover $n_p \equiv 1 \mod p$ and $n_p \mid b$.

Before the proof, we give a corollary.

Corollary 6.0.8. *Let G be a group of order pq for primes $p < q$. Suppose that $q \not\equiv 1 \mod p$, then $G \cong \mathbb{Z}/pq\mathbb{Z}$.*

Proof. By Theorem 6.0.7, G contains a Sylow q-subgroup $K \leq G$, and $K \cong \mathbb{Z}/q\mathbb{Z}$. Thus $n_q = [G : N_G(K)]$ divides $|G| = pq$. But $n_q = 1$ or $1 + q$ or $1 + 2q \ldots$, so n_q must divide p and thus $n_q = 1$, whereby $N_G(K) = G$ and $K \trianglelefteq G$ is normal.

Now for the prime p, G also has a Sylow p-subgroup $H \cong \mathbb{Z}/p\mathbb{Z}$, and again $n_p = 1$ or 1_p or $\ldots$ and n_p divides pq. Thus n_p divides q, hence either $n_p = 1$ or $q = 1 + kp$ for some k. The latter case implies that $q \equiv 1 \mod p$, which we assume is not the case. Hence $n_p = 1$ as well, and $H \trianglelefteq G$ is also a normal subgroup.

Exercise 6.0.9. Show that $KH = G$, $K \cap H = \{\mathrm{id}_G\}$, and $kh = hk$ for all $k \in K$ and $h \in H$

Thus $G \cong K \times H \cong \mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \cong \mathbb{Z}/pq\mathbb{Z}$ as $\gcd(p, q) = 1$. □

We now prove the Sylow theorems.

Proof of Theorem 6.0.7.

(a): We induct on the order of G. If $|G| = p$, then we are done. Now suppose that $|G| = p^a b$ with $\gcd(p, b) = 1$, and (a) holds for all groups of order $< p^a b$ such that p divides the order.

Let $G \curvearrowright G$ by conjugation. Then

$$|G| = |Z(G)| + |[G : Z_G(x_1)]| + \cdots + |G : Z_G(x_k)|.$$

If $p \nmid [G : Z_G(x_i)]$ for some i, then $p^a \mid |Z_G(x_i)|$ as $p^a \mid |G| = [G : Z_G(x_i)]|Z_G(x_i)|$, and so by induction there eists a Sylow p-subgroup $\mathcal{P} \leq Z_G(x_i) \lneq G$. Thus we may assume that $p \mid [G : Z_G(x_i)]$ for all i. Hence $p \mid |Z(G)|$.

By Cauchy's theorem (Theorem 5.4.4), $Z(G)$ has an element g of order p, and we let $N = \langle g \rangle \leq Z(G)$. Since $g \in Z(G)$, N is normal of order p. If $a = 1$, we are done (take $\mathcal{P} = N$). Hence we may assume $a \geq 2$. Consider the quotient map

$$\pi : G \to G/N,$$

and noting that $|G/N| = p^{a-1}b < |G|$, the inductive hypothesis provides a Sylow p-subgroup $\mathcal{P}' \leq G/N$ of order $|\mathcal{P}'| = p^{a-1}$. Let $\mathcal{P} = \pi^{-1}(\mathcal{P}')$. Then $N \leq \mathcal{P}$ and

$$\pi|_{\mathcal{P}} : \mathcal{P} \to \mathcal{P}'$$

is surjective with kernel N, hence

$$\mathcal{P}/N \cong \mathcal{P}'$$

and so

$$p^{a-1} = |\mathcal{P}'| = \frac{|\mathcal{P}|}{|N|} = \frac{|\mathcal{P}|}{p},$$

so $\mathcal{P}$ is a Sylow p-subgroup of G, as desired.

(a'): Let $\mathcal{P} \leq G$ be a Sylow p-subgroup, and let $Q \curvearrowright G/\mathcal{P}$ by left translation. Now

$$|G/\mathcal{P}| = b,$$

which is not divisible by p. By the class equation, we have

$$b = |G/\mathcal{P}| = \sum_{\substack{\mathcal{O} \text{ orbit} \\ \text{for } q \curvearrowright G/\mathcal{P}}} |\mathcal{O}|,$$

thus as p does not divide b, there must be at least one orbit $\mathcal{O}$ such that p does not divide $|\mathcal{O}|$, say $\mathcal{O}_x$.

By the Orbit-Stabilizer theorem, $Q/Q_x \cong \mathcal{O}_x$, so $\frac{Q}{Q_x} = |\mathcal{O}_x|$, hence $p^m = |Q| = |\mathcal{O}_x||Q_x|$ Since p does not divide $|\mathcal{O}_x|$, we must have $|\mathcal{O}_x| = 1$, and so $\mathcal{O}_x = \{x\}$ and x is a fied point of the action $Q \curvearrowright G/\mathcal{P}$. Since $x \in G/\mathcal{P}$ is a coset, say $x = g\mathcal{P}$, for all $q \in Q$ we have

$$\begin{aligned} qg\mathcal{P} = g\mathcal{P} &\iff g^{-1}qg\mathcal{P} = \mathcal{P} \\ &\iff g^{-1}qg \in \mathcal{P} \\ &\iff q \in g\mathcal{P}g^{-1}. \end{aligned}$$

Thus $Q \subseteq g\mathcal{P}g^{-1} \iff g^{-1}Qg \subseteq \mathcal{P}$, as desired.

(b): If $\mathcal{P}$ and $\mathcal{P}'$ are Sylow p-subgroups, then $g\mathcal{P}g^{-1} \subset \mathcal{P}'$ but these have the same order and so $g\mathcal{P}g^{-1} = \mathcal{P}'$.

(c): Let $S \coloneqq \{\text{Sylow } p\text{-subgroups of } G\}$. Let $G \curvearrowright S$ by conjugation. We note that $S = \mathcal{O}_{\mathcal{P}}$ since all Sylow p-subgroups are conjugate. The stabilizer of $\mathcal{P}$ is

$$G_{\mathcal{P}} = \{g \in G \mid g\mathcal{P}g^{-1} = \mathcal{P}\}$$

si the normalizer subgroup of $\mathcal{P}$. The Orbit-Stabilizer theorem now gives

$$n_p = |S| = |\mathcal{O}_{\mathcal{P}}| = [G : N_G(\mathcal{P})],$$

in particular by Lagrange's theorem,

$$n_p \text{ divides } [G : N_G(\mathcal{P})][N_g(\mathcal{P}) : \mathcal{P}] = [G : \mathcal{P}] = b.$$

It remains to show that $n_p \equiv 1 \mod p$. For this, let $\mathcal{P} \curvearrowright S$ by conjugation. From the class equation,

$$|S| = \sum_{|\mathcal{O}|=1} |\mathcal{O}| + \sum_{|\mathcal{O}|>1} |\mathcal{O}|,$$

and note that if $|\mathcal{O}| > 1$ then the Orbit-Stabilizer theorem gives

$$|\mathcal{O}| = |\mathcal{P}/\mathcal{P}_{\mathcal{O}}| = p^m$$

for some $m \geq 1$.

Thus

$$\begin{aligned} |S| &\equiv \sum_{|\mathcal{O}|=1} |\mathcal{O}| \mod p \\ &\equiv \sum_{\substack{\text{fixed points} \\ \text{of } \mathcal{P} \curvearrowright S}} 1 \\ &\equiv |\{Q \in S \mid gQg^{-1} = Q \text{ for all } g \in \mathcal{P}\}| \mod p. \end{aligned}$$

So we have

$$n_p = |S| \equiv |\{Q \in S \mid gQg^{-1} = Q \text{ for all } g \in \mathcal{P}\}| \mod p. \tag{1}$$

For such Q, $\mathcal{P} \leq N_G(Q)$, thus Q and $\mathcal{P}$ are Sylow p-subgroups of order p^a of $N_G(Q)$. By (b), Q and $\mathcal{P}$ are conjugate subgroups of $N_G(Q)$, say $nQn^{-1} = \mathcal{P}$ with $n \in N_G(Q)$. But since $n \in N_G(Q)$, $nQn^{-1} = Q$, and so $Q = \mathcal{P}$. Thus if $Q \in S$ is a fixed point of the action $\mathcal{P} \curvearrowright S$, then $Q = \mathcal{P}$. Thus

$$|\{Q \in S \mid gQg^{-1} = Q \text{ for all } g \in \mathcal{P}\}| = 1,$$

and Equation (1) now reads $n_p \equiv 1 \mod p$, as desired. □

We'll finish by giving a nice example of how the Sylow theorems could be used to determine the structure of some small groups, though we'll focus on a very small group, S_3.

Example 6.0.10. We will show that if G is a group of order 6, then $G \cong \mathbb{Z}/6\mathbb{Z}$ or $G \cong S_3$, and these two are not isomorphic.

Note that G is either abelian or not. If G is abelian, then by Theorem 4.0.6, $G \cong \mathbb{Z}/6\mathbb{Z}$.

If G is not abelian, then clearly $G \ncong \mathbb{Z}/6\mathbb{Z}$, the latter being abelian.

Proposition 6.0.11. *Let G be a non-abelian group of order 6. Then $G \cong S_3$.*

Proof. Now $n_2 \equiv 1 \mod 2$ and $n_2 \mid 6$, so $n_2 \in \{1, 3\}$. Similarly, $n_3 \equiv 1 \mod 3$ and $n_3 \mid 6$, so $n_3 = 1$. Let $\mathbb{Z}/3\mathbb{Z} \cong K \trianglelefteq G$ be the unique (normal) Sylow 3-subgroup. Let $\mathbb{Z}/2\mathbb{Z} \cong H \leq G$ be a Sylow 2-subgroup.

If $n_2 = 1$, H is normal and $K \curvearrowright H$ by conjugation, which is equivalent to a group homomorphism $\mathbb{Z}/3\mathbb{Z} \to \mathrm{Perm}(H) \cong S_2 \cong \mathbb{Z}/2\mathbb{Z}$. But this must be trivial, hence $kh = hk$ for all $h \in H$ and all $k \in K$. Hence $G \cong H \times K \cong \mathbb{Z}/6\mathbb{Z}$, which cannot be the case as G is not abelian.

Thus $n_2 = 3$, and we let $G \curvearrowright G/H$, which gives a group homomorphism $\varphi : G \to S_3$ as there are three cosets of H. Clearly $\ker\varphi \subseteq H$, so either $\ker\varphi = H$ or $\ker\varphi = \{\mathrm{id}_G\}$. But if $\ker\varphi = H$, then H would be normal and so $n_2 = 1$, which is not the case. Hence φ is injective, and since $|G| = 6 = |S_3|$, φ must also be surjective. Hence $G \cong S_3$, as claimed. □

7. Linear Algebra

Remark 7.0.1. This section serves as a rigorous trek through linear algebra over fields, as many first introductions to linear algebra focus on vector spaces over $\mathbb{R}$ or $\mathbb{C}$. Perhaps a majority of this section will be familiar to you. We include this section to go through the standard facts about matrices and vector spaces without assuming anything. In particular, I wanted to include the existence of bases, which is generally not covered in a first linear algebra course. I also include a discussion of determinants, as I find it quite enlightening that the determinant is uniquely defined by its properties rather than a mess of a formula.

Historically, the motivation for developing linear algebra is as a technical shorthand for solving systems of linear equations. That is, we can package the system of linear equations into a matrix

$$\begin{aligned} 3x + y + 2z &= 0 \\ 2x + y + z &= 0 \end{aligned} \quad \longleftrightarrow \quad \begin{pmatrix} 3 & 1 & 2 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and then we can perform row operations on the matrix (swapping rows, adding rows to other rows, or scaling rows by an element of the field) to find a simpler system to solve. In scaling rows, when we work over a field, we can also divide by non-zero field elements.

Recall 7.0.2. We will need the notion of a *field*, which you should have seen in the first semester. If it is not familiar, here is a definition.

Definition 7.0.3. A *field* F is a commutative ring with unity, denoted 1_F, in which every nonzero element is invertible. That is, if $0_F \neq a \in F$, then there is $a^{-1} \in F$ such that $aa^{-1} = a^{-1}a = 1_F$.

Familiar fields are $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. Perhaps some less familiar fields are

$$\text{for a prime } p,\ \mathbb{Z}/p\mathbb{Z} \coloneqq \{0, 1, 2, \ldots, p-1\} \text{ with addition and multiplication } \mod p,$$

and some very large fields such as

$$\mathbb{Q}(x) \coloneqq \left\{ \frac{f(x)}{g(x)} \;\middle|\; f(x), g(x) \in \mathbb{Q}[x] \text{ polynomials, and } g(x) \neq 0 \right\},$$

the field of rational functions with rational coefficients.

In doing these row operations, a few nice facts which simplify our solution method become apparent. First, when we are solving systems of the form $Ax = 0$ (homogeneous system), the sum of any two solutions is also a solution; likewise, if we scale any solution we get another solution. Secondly, if we want to solve a system $Ax = b$, then any solution to $Ax = 0$ can be added and we obtain another solution. This is central to obtaining solutions to linear ordinary differential equations! It is so useful inf act, that we want to capture the abstract notion of a set where we can add and scale elements.

Definition 7.0.4 (Vector Space)**.** A *vector space over a field* F is the data $(V, 0_V, +_V, \cdot_V)$ of

- an abelian group $(V, 0_V, +_V)$,
- a map $\cdot : F \times V \to V,\ (c, v) \mapsto c \cdot v$ called *scalar multiplication*

such that for all $a, b \in F$ and all $v, w \in V$

- $1_F \cdot v = v$,
- $(a +_F b) \cdot v = (a \cdot v) +_V (b \cdot v)$,
- $a \cdot (v +_V w) = (a \cdot v) +_V (a \cdot w)$,
- $(ab) \cdot v = a \cdot (b \cdot v)$.

The elements of a vector space are called *vectors*, and generally field elements are called *scalars*. When we want to specify the field F, we may say that V is an F-vector space, or say "V is a vector space over F".

Vector spaces are ubiquitous in math. In some sense, vector spaces are the only things we really understand. One way we attempt to understand more complicated objects is by associated vector spaces that capture a part of the complicated structure. In some sense, this is really all that calculus is, an attempt to take complicated functions and associate some kind of linear structure by using the derivative. A fancier example is the advent of "cohomology" in understanding geometric objects, for example associating vector spaces of functions by using some differential structure (as in de Rham cohomology).

Example 7.0.5 (Examples of vector spaces)**.** Throughout, F will be a field.

(0) The simplest example of a vector space is $0 \coloneqq \{0_F\}$.
(1) The next simplest example is simply the field F itself.
(2) Similarly, $V = F^n \coloneqq \{(x_1, \dots, x_n) \mid x_i \in F\}$ is an F-vector space with $+_V$ defined coordinate-wise and $\cdot_V$ defined by
$$a \cdot (x_1, x_2, \dots, x_n) \coloneqq (a \cdot_F x_1, a \cdot_F x_2, \cdots a \cdot_F x_n).$$
(3) A familiar example is $\mathbb{R}^n$, the n-dimensional real Euclidean space of multivariable calculus.
(4) Polynomials with coefficients in a field ($F[x]$). For example,
$$\mathbb{R}[x] = \{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \mid a_i \in \mathbb{R}, n \in \mathbb{Z}_{\geq 0}\},$$
polynomials with real coefficients.

Examples abound, and you are encouraged to come up with some more!

Let's see some immediate first properties of vector spaces.

Proposition 7.0.6. *Let V be an F-vector space. Then for any $v \in V$,*

(i) $0_F \cdot v = 0_V$,
(ii) $-v = (-1_F) \cdot v$.

Proof. To prove (i), first note that
$$\begin{aligned} 0_F \cdot v &= (0_F +_F 0_F) \cdot v \\ &= (0_F \cdot v) +_V (0_F \cdot v), \end{aligned}$$
and now adding $-(0_F \cdot v)$ to both sides yields
$$\begin{aligned} 0_V &= (0_F \cdot v) +_V (0_F \cdot v) - (0_F \cdot v) \\ &= 0_F \cdot v +_V 0_V \\ &= 0_F \cdot v, \end{aligned}$$
and so $0_V = 0_F \cdot v$.

To show (ii), recall that $-v$ is the additive inverse of v in the abelian group V. So we want to show that $v +_V (-1_F) \cdot v = 0_V$. We compute
$$\begin{aligned} v +_V (-1_F) \cdot v &= 1_F \cdot v +_V (-1_F \cdot v) \\ &= (1_F +_F (-1_F)) \cdot v \\ &= (0_F) \cdot v \\ &= 0_V, \end{aligned}$$
as desired. □

Exercise 7.0.7. Let V be an abelian group. Show that the set $\mathrm{Hom}_{\mathbb{Z}}(V,V)$ of group homomorphisms $V \to V$ form a ring. Show that the data of a vector space over F is the same as a ring homomorphism $F \to \mathrm{Hom}_{\mathbb{Z}}(V,V)$.

Vector spaces are interesting objects, but much like with groups, the really important aspect is maps between the objects!

Definition 7.0.8. Let V, W be F-vector spaces. An *F-vector space homomorphism* $T : V \to W$ is a group homomorphism $T : V \to W$ that is also *F-linear*, i.e. for all $v, v' \in V$ and for all $a \in F$,

$$T(v + a \cdot v') = T(v) + a \cdot T(v').$$

For simplicity, when F is understood, we call an F-vector space homomorphism a *linear map*.

Definition 7.0.9. We say that two F-vector spaces V and W are *isomorphic* if there are linear maps

$$T : V \to W,\ T' : W \to V$$

such that

$$T' \circ T = \mathrm{id}_V,\ T \circ T' = \mathrm{id}_W,$$

and we call the maps T, T' isomorphisms, with $T' = T^{-1}$ the inverse of T, and write $T \cong W$.

Example 7.0.10. Let A be an $m \times n$ matrix with entries in F. Then we can define a linear map $F^n \to F^m$ by $x \mapsto Ax$ given by standard matrix multiplication. When $n = m$ and A is an invertible matrix, then the map $X \mapsto Ax$ is an isomorphism of F^n with itself.

Example 7.0.11. Let

$$V = \{\text{polynomials of degree} \leq 2 \text{ with real coefficients}\} = \{a_2x^2 + a_1x + a_0 \mid a_i \in \mathbb{R}\}.$$

Check that V is a vector space over $\mathbb{R}$. In fact, $V \cong \mathbb{R}^3$ via the map

$$T : V \to \mathbb{R}^3,\ a_2x^2 + a_1x + a_0 \mapsto (a_2, a_1, a_0).$$

Check that T is linear, define T^{-1} and check that it is also linear.

Example 7.0.12. Let F be a field, and let

$$V = \{\text{functions } \{1, \ldots, n\} \to F\}.$$

Then with pointwise addition $(f + g)(x) = f(x) + g(x)$ and scaling $(a \cdot f)(x) = af(x)$, we see that V is a vector space over F. Check that

$$T : V \to F^n,\ f \mapsto (f(1), f(2), \ldots, f(n))$$

defines an isomorphism with inverse

$$T^{-1} : F^n \to V,\ (c_1, \ldots, c_n) \mapsto (f : \{1, \ldots, n\} \to F, f(i) \mapsto c_i).$$

Remark 7.0.13. This last example highlights the fact that vector spaces are like free abelian groups, with a bit more structure.

As with groups, there is a notion of a subset that has the same structure.

Definition 7.0.14. Let V be a vector space over F. A subset $W \subseteq V$ is called a *subspace* if

(i) $W \leq V$ is a subgroup (W is closed under addition), and
(ii) for any $a \in F$ and $w \in W$, $a \cdot w \in W$ (W is closed under scaling).

Example 7.0.15. Let $W = \{a_2x^2 + a_1x + a_0 \mid a_i \in \mathbb{R}\}$, then $W \subseteq \mathbb{R}[x]$ is a subspace.

Definition 7.0.16. Let V and W be F-vector spaces and $T : V \to W$ a linear map. The *kernel of* T, $\ker T$ is

$$\ker T \coloneqq \{v \in V \mid T(v) = 0_W\},$$

and the *image of* T, $\operatorname{im} T$ is

$$\operatorname{im} T \coloneqq \{w \in W \mid w = T(v) \text{ for some } v \in V\}.$$

Exercise 7.0.17. Let $T : V \to W$ be a linear map. Show that $\ker T \leq V$ is a subspace, and $\operatorname{im} T \leq W$ is a subspace.

To motivate our next main topic, dimension, let's see if we can find all the subspaces of $\mathbb{R}^2$.

Example 7.0.18 (Motivating Example, Subspaces of $\mathbb{R}^2$)**.** We know that $\mathbb{R}^2$ has a few trivial subspaces, namely $\{0\} \subseteq \mathbb{R}^2$ and $\mathbb{R}^2 \subseteq \mathbb{R}^2$. Are there any others?

We take out hint from the name of this section, linear algebra, and guess that perhaps we should look at lines. This isn't a bad guess, since if $W \subseteq \mathbb{R}^2$ is a subspace, and $w \in W$, then $\mathbb{R}w = \{a \cdot w \mid a \in \mathbb{R}\} \subset W$. But not every line is a subspace, since a subspace must contain 0 (as it is in particular a subgroup). So maybe all the other subspaces are lines through the origin.

Proposition 7.0.19. *Let $W \subseteq \mathbb{R}^2$ be a subspace. then W is one of the following*

- $\{0\}$
- $\mathbb{R}w$ *for some* $0 \neq w \in W$
- $\mathbb{R}^2$,

Proof. What we want to show is that if W is not $\{0\}$ and is also not $\mathbb{R}^2$, then $W = \mathbb{R}w$ for some $0 \neq w \in W$.

Well, if $W \neq \{0\}$, then there is clearly some $0 \neq w \in W$. And we've also observed that since W is a subspace, we have $\mathbb{R}w \subseteq W$. So all we have to show is that if there is some $w' \in W$ such that $w' \notin \mathbb{R}w$, then $W = \mathbb{R}^2$.

So suppose that $w' \in W$ but $w' \neq aw$ for any $a \in \mathbb{R}$. Then clearly $\mathbb{R}w + \mathbb{R}w' \subseteq W \subset \mathbb{R}^2$. And we would be done if we could say that $\mathbb{R}w + \mathbb{R}w' = \mathbb{R}^2$. But how do we know?? We need to understand "how big" subspaces are... We postpone the proof until we develop this notion, which is called *dimension.* □

7.1. **Dimension theory.** While we may have developed an intuition for dimension, saying things like "$\mathbb{R}^2$ is 2-dimensional", here we will develop the notion of dimension rigorously.

Definition 7.1.1 (Linear Combination)**.** Let V be a F-vector space, and let $v_1, \dots, v_n \in V$. An (F)-linear combination of $v_1, \dots, v_n$ is a vector in V of the form

$$a_1 v_1 + \cdots + a_n v_n, \ a_i \in F.$$

Definition 7.1.2. Let V be an F-vector space, and $X \subseteq V$ a subset. The (F)-span of X is

$$\operatorname{span}_F(X) \coloneqq \{a_1 v_1 + \cdots + a_n v_n \mid a_i \in F, v_i \in X, n \in \mathbb{Z}_{\geq 0}\},$$

the set of finite linear combinations of elements in X. When $X = \{v_1, \dots, v_n\}$ we write $\operatorname{span}_F(v_1, \dots, v_n)$.

Remark 7.1.3. The subset $X \subseteq V$ could be infinite! However, every element of $\operatorname{span}_F(X)$ can be written as a finite sum.

Proposition 7.1.4. *Let V be an F-vector space, and $S \subseteq V$ a subset. Then* $\operatorname{span}_F(S)$ *is the smallest subspace of V containing S.*

Proof.

Exercise 7.1.5. Check that $\operatorname{span}_F(S)$ is a subspace of V.

Let $W \subseteq V$ be a subspace containing S. We want to show that $\operatorname{span}_F(S) \subseteq W$. Since W is closed under addition and scaling, it contains all linear combinations of elements of S, i.e. $\operatorname{span}_F(S) \subseteq W$, as desired. □

In light of Proposition 7.1.4, we can give a new definition of $\operatorname{span}_F(S)$.

Definition 7.1.6. Let V be an F-vector space, and $S \subseteq V$ a subset. The (F)-*span of* S, denoted $\operatorname{span}_F(S)$ is a subspace of V such that

(i) $S \subseteq \operatorname{span}_F(S)$
(ii) if $W \subseteq V$ is any subspace such that $S \subseteq W$, then $\operatorname{span}_F(S) \subseteq W$.

Remark 7.1.7. We note that Definition 7.1.6 doesn't actually tell us what $\operatorname{span}_F(S)$ *is*, but rather says that $\operatorname{span}_F(S)$ is some subspace that satisfies a certain property, much like our definition of free abelian groups (Definition 4.1.1). This is somewhat of a philosophical approach, saying that an object can be defined by its properties rather than as an object itself. The reader is invited to contemplate which (if any) of their properties are defining. Nevertheless, Definition 7.1.6 is very useful in proofs.

Exercise 7.1.8. Let $W \subseteq V$ be a subspace. Show that $\operatorname{span}_F(W) = W$ using Definition 7.1.6.

Exercise 7.1.9. Let $S \subset V$ be a subset, and define

$$K_S \coloneqq \bigcap_{\substack{W \subseteq V \text{ subspace} \\ S \subseteq W}} .$$

Show that $K_S = \operatorname{span}_F(S)$.

Example 7.1.10. We note that different subsets may have the same span! Indeed, let

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

Then since

$$(-1)\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

we have

$$\begin{aligned}\operatorname{span}_{\mathbb{R}}(S) &= \left\{ a\begin{pmatrix}1\\0\\1\end{pmatrix} + b\begin{pmatrix}1\\1\\1\end{pmatrix} + c\begin{pmatrix}0\\1\\0\end{pmatrix} \;\middle|\; a,b,c \in \mathbb{R} \right\} \\ &= \left\{ a-c\begin{pmatrix}1\\0\\1\end{pmatrix} + b - c\begin{pmatrix}1\\1\\1\end{pmatrix} \;\middle|\; a,b,c \in \mathbb{R} \right\} \\ &= \left\{ a'\begin{pmatrix}1\\0\\1\end{pmatrix} + b'\begin{pmatrix}1\\1\\1\end{pmatrix} \;\middle|\; a',b' \in \mathbb{R} \right\} \\ &= \operatorname{span}_{\mathbb{R}}\left(\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix} \right).\end{aligned}$$

So S and $\left\{ \begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix} \right\}$ have the same $\mathbb{R}$-span!

One way to answer the "how large is a subspace" question is to know when we have the fewest possible vectors whose span is the subspace of interest.

Remark 7.1.11. Observe that if $s \in S$ such that $s \in \operatorname{span}_F(S \setminus \{s\})$, then

$$\operatorname{span}_F(S) = \operatorname{span}_F(S \setminus \{s\}),$$

as in Example 7.1.10. Moreover, if $0 \neq s \in \operatorname{span}_F(S)$, then there are $v_1, \dots, v_{n-1} \in S$ and $a_1, \dots, a_{n-1} \in F$ not all zero such that

$$a_1 v_1 + \cdots + a_{n-1} v_{n-1} - s = 0.$$

Definition 7.1.12. Let V be an F-vector space. A subset $S \subseteq V$ is *linearly dependent (over F)* if exist elements $a_1, \dots, a_n \in F$ not all zero and elements $s_1, \dots, s_n \in S$ such that

$$a_1 s_1 + \cdots + a_n s_n = 0.$$

If no such elements of F and S exist, then S is *linearly independent (over F)*. We generally omit "over F" when the field is understood.

Exercise 7.1.13. Show that if $S \subset V$ is linearly independent, then $0 \notin S$.

Example 7.1.14. Let $V = F^n$, and let

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

which are called the *standard unit vectors*. Then $\{e_1, \dots, e_n\}$ is linearly independent. Indeed, suppose $a_1, \dots, a_n \in F$ such that $\sum_{i=1}^n a_i e_i = 0$. Then since

$$a_1 e_1 + a_2 e_2 + \cdots + a_n e_n = (a_1, a_2, \dots, a_n),$$

we have $a_i = 0$ for all $1 \leq i \leq n$.

Exercise 7.1.15. Show that $\{\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}3\\4\end{pmatrix}\} \subset \mathbb{R}^2$ is linearly independent.

Example 7.1.16. $\left\{\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}3\\4\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}\right\} \subset \mathbb{R}^2$ is linearly dependent. Indeed,

$$2c_3\begin{pmatrix}1\\2\end{pmatrix} - c_3\begin{pmatrix}3\\4\end{pmatrix} + c_3\begin{pmatrix}1\\0\end{pmatrix} = 0 \text{ for any } c_3 \in \mathbb{R}.$$

This can be found considering when

$$c_1\begin{pmatrix}1\\2\end{pmatrix} + c_2\begin{pmatrix}3\\4\end{pmatrix} + c_3\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix},$$

leading to the conditions

$$\begin{aligned} c_1 + 3c_2 + c_3 &= 0 \\ 2c_1 + 4c_2 &= 0. \end{aligned}$$

We now turn our attention to defining a notion of dimension using linearly independent sets.

Lemma 7.1.17. *Let V be an F-vector space and $S \subset V$ be linearly independent. If*

$$v = c_1 s_{i_1} + \cdots c_n s_{i_n} = d_1 s_{j_i} + \cdots + d_m s_{j_m}$$

with $s_i, s_j \in S$ and $0 \neq c_i, d_j \in F$ for all i, j, then $n = m$ and $\{i_1, \ldots, i_n\} = \{j_1, \ldots, j_n\}$.

Proof. We compute

$$0 = v - v = c_1 s_{i_1} + \cdots c_n s_{i_n} - d_1 s_{j_i} - \cdots - d_m s_{j_m},$$

hence as S is linearly independent, $c_i = d_j = 0$ for all i, j unless for each i_k there is a $j_{k'}$ such that $s_{i_k} = s_{j_{k'}}$ and $c_{i_k} = d_{j_{k'}}$. Thus $n = m$ and the s_i's are a rearrangement of the s_j's, as claimed. □

Definition 7.1.18. Let V be an F-vector space. A subset $\mathcal{B} \subset V$ is called a *basis* for V (over F) if

- $\operatorname{span}_F(\mathcal{B}) = V$, and
- $\mathcal{B}$ is linearly independent (over F).

Example 7.1.19. The standard unit vectors $e_1, \ldots, e_n$ are a basis of F^n.

Lemma 7.1.20. *Let $\mathcal{B} = \{v_1, \ldots, v_m\}$ be a basis of V. Suppose that $S = \{s_1, \ldots, s_n\} \subset V$ is linearly independent. Then $n \leq m$.*

Proof. We proceed by induction. Since $\operatorname{span}_F(\mathcal{B}) = V$, we can write

$$x_1 = c_1 v_1 + \cdots c_m v_m$$

for some $c_i \in F$. Since S is linearly independent, $x_i \neq 0$, so there is some $c_i \neq 0$. Without loss of generality, we may assume $c_1 \neq 0$. Thus

$$v_1 = \frac{1}{c_1}\left(-x_1 + c_2 v_2 + \cdots + c_n v_n\right) = \frac{-1}{c_1}x_1 + \frac{c_2}{c_1}v_2 + \cdots + \frac{c_n}{c_1},$$

hence $v_1 \in \operatorname{span}_F(x_1, v_2, \ldots, v_n) = V$. Thus $\{x_1, v_2 \ldots, v_n\}$ spans V.

Moreover, $\{x_1, v_2, \ldots, v_n\}$ is linearly independent. Indeed suppose that

$$b_1 x_1 + b_2 v_2 + \cdots + b_n v_n = 0,$$

then if $b_1 \neq 0$, we cal solve for x_1 and obtain

$$x_1 = \frac{-1}{b_1}(b_2 v_2 + \cdots + b_n v_n)$$

hence

$$c_1v_1 + (c_2 - \frac{b_2}{b_1})v_2 + \cdots + (c_n - \frac{b_n}{b_1})v_n = 0,$$

which cannot occur as $\mathcal{B}$ is linearly independent; thus $b_1 = 0$ and so $b_i = 0$ for all $2 \leq i \leq n$ as $\mathcal{B}$ is linearly independent.

Thus $\{x_1, v_2, \ldots, v_m\}$ is a basis.

Proceeding inductively, we see that if $\{x_1, \ldots, x_r, v_{r+1}, \ldots, v_m\}$ is a basis, then so is $\{x_1, \ldots, x_{r+1}, v_{r+2}, \ldots, v_m\}$. Hence, as we can always continue this, we must have $n \leq m$. □

Theorem 7.1.21. *Let V be an F-vector space and $\mathcal{B}$ a basis. Then $|\mathcal{B}|$ is independent of the choice of basis.*

Proof. Assume that V has a finite basis $\{v_1, \ldots, v_m\}$. Then by Lemma 7.1.20, we have $m = |\mathcal{B}|$, as desired.

If instead V has an infinite basis, $\mathcal{C}$, then any finite subset of $\mathcal{B}$ is contained in the span of a finite subset of $\mathcal{C}$. Hence $|calB| \leq |\mathcal{C}|$, and vice versa. Hence $|\mathcal{B}| = |\mathcal{C}|$, as was to be shown. □

Definition 7.1.22 (Dimension)**.** Let V be an F-vector space. The *dimension of V (over F)* is defined as

$$\dim_F(V) = |\mathcal{B}|$$

for any F-basis $\mathcal{B}$ of V.

Example 7.1.23. The vector space $\mathbb{F}^n$ has dimension n over F.

Example 7.1.24. Let $\mathbb{Q}(\sqrt{2}) := \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. Then $\dim_{\mathbb{Q}} \mathbb{Q}(\sqrt{2}) = 2$, and indeed $\{1, \sqrt{2}\} \subset \mathbb{Q}(\sqrt{2})$ is a basis.

Example 7.1.25. $\mathbb{R}[x]$ has basis $\{1, x, x^2, x^3 \ldots\}$. Thus $\dim_{\mathbb{R}} \mathbb{R}[x] = \infty$ (it has countablly infinite dimension).

Exercise 7.1.26. Show that $\mathbb{R}$ is a vector space over $\mathbb{Q}$.

Example 7.1.27. In fact, $\dim_{\mathbb{Q}} \mathbb{R} = \infty$, though this infinity is uncountable.

Note that we have never actually shown that every vector space even has a basis! We'll need some setup before showing that bases exist.

The first is a way to recognize when a subset is a basis. We'll give two criteria, the first being a maximal linearly independent subset; the second being a minimal spanning set.

Proposition 7.1.28. *Let V be an F-vector space and suppose that $S \subseteq V$ be a maximal linearly independent subset, i.e.*

- *S is linearly independent, and*
- *S is maximal with this property, that is, if $S \subsetneq T \subseteq V$, then T is linearly dependent.*

Then S is a basis for V.

Proof. By assumption, S is linearly independent. Thus it remains to show that $\text{span}_F(S) = V$. Let $0 \neq v \in V$. If $v \in S$, then $v \in S \subseteq \text{span}_F(S)$. If $v \notin S$, then $S \subsetneq S \cup \{v\} \subseteq V$, thus $S \cup \{v\}$ is linearly dependent. So there exist $c_1, \ldots, c_n \in F$ not all zero and $v_1, \ldots, v_n \in S \cup \{v\}$ such that

$$c_1v_1 + \cdots c_nv_n = 0.$$

If $v \neq v_i$ for any i, then $c_i = 0$ for all i as S is linearly independent. Hence we must have $v = v_i$ for some i. Without loss of generality, suppose $v = v_1$. Moreover, we must have $c_1 \neq 0$, otherwise S would not be linearly independent. Thus

$$v = \frac{-1}{c_1}(c_2v_2 + \cdots + c_nv_n) \in \text{span}_F(S),$$

thus $v \in \text{span}_F(S)$, and $\text{span}_F(S) = V$. □

Proposition 7.1.29. *Let V be an F-vector space and suppose that $S \subseteq V$ is a minimal spanning set, i.e.*

- $\text{span}_F(S) = V$, *and*
- S *is minimal with this property, that is, if* $T \subsetneq S$ *then* $\text{span}_F(T) \neq V$.

Then S is a basis for V.

Proof. By assumption, S spans V. Thus it remains to show that S is linearly independent. Suppose that there are some $c_i \in F$ and $s_i \in S$ such that

$$c_1 s_1 + \cdots c_n s_n = 0.$$

If $c_i \neq 0$, then solving for s_i, we see that $s_i \in \text{span}_F(s_1, \ldots s_{i-1}, s_{i+1}, \ldots, s_n)$. But then

$$\text{span}_F(s_1, \ldots s_{i-1}, s_{i+1}, \ldots, s_n) = \text{span}_F(S),$$

which contradicts S being a minimal spanning set. Thus $c_i = 0$ for all i, and S is linearly independent. □

We're almost ready to prove the fact that every vector space has a basis. We'll need one technical notion first.

Definition 7.1.30. A set $\mathcal{P}$ is called *partially ordered* if there is a relation $\leq$ on $\mathcal{P}$ such that for all $x, y, z \in \mathcal{P}$ we have

- $x \leq x$,
- if $x \leq y$ and $y \leq x$, then $x = y$, and
- if $x \leq y$ and $y \leq z$, then $x \leq z$.

The relation $\leq$ is called a *partial order* as it need not be the case that any two elements of $\mathcal{P}$ are comparable. A partially ordered set is sometimes called a "poset".

Example 7.1.31. Let X be a set. Let $\mathcal{P}$ be the set of subsets of X, and define a partial order by inclusion. That is, for subsets $A, B \subset X$ we say $A \leq B$ if $A \subseteq B$.

Definition 7.1.32. Let $\mathcal{P}$ be a partially ordered set. A *chain* is a totally ordered subset $\mathcal{C} \subset \mathcal{P}$, that is, for any $x, y \in \mathcal{C}$ we have $x \leq y$ or $y \leq x$.

Remark 7.1.33. Posets are fundamental, sadly we don't have time to delve into properties or examples of posets. One example is the natural numbers ordered by size. Another example is subsets of $\mathbb{R}$ ordered by inclusion.

Lemma 7.1.34 (Zorn's Lemma)**.** *Let $\mathcal{P}$ be a partially ordered set. Suppose $\mathcal{P}$ is nonempty and every chain in $\mathcal{P}$ has an upper bound, i.e., for a chain $\mathcal{C} \subset \mathcal{P}$ there is an element $x \in \mathcal{P}$ such that $a \leq x$ for all $a \in \mathcal{C}$. Then $\mathcal{P}$ has at least one maximal element, i.e., an element $m \in \mathcal{P}$ such that $x \leq m$ for all $x \in \mathcal{P}$.*

Remark 7.1.35. Lemma 7.1.34 is not really a lemma, it is an axiom of set theory, which is equivalent to the Axiom of Choice (which says that there is a way to pick an element out of any set). In fact, the Axiom of Choice and Zorn's Lemma are both equivalent to the statement "every vector space has a basis". So in some sense, we are assuming what we want to prove. On the other hand, if we want to prove that every vector space has a basis, we have to assume this.

If you are uncomfortable with the Axiom of Choice, then assume all vector spaces are finite dimensional in this section.

We are now ready to show that every vector space has a basis.

Theorem 7.1.36 (Every vector space has a basis)**.** *Let V be a nonzero F-vector space. Then V has a basis.*

Proof. Let $\mathcal{P}$ be the poset of linearly independent subsets of V, partially ordered by inclusion. Clearly $\mathcal{P}$ is non-empty as there is some nonzero $v \in V$, and $\{v\}$ is linearly independent. Thus $\{v\} \in \mathcal{P}$. If $\mathcal{C} \subset \mathcal{P}$ is a chain, then

$$\bigcup_{X \in \mathcal{C}} X$$

is an upper bound for $\mathcal{C}$.

Exercise 7.1.37. Show that $\bigcup_{X \in \mathcal{C}} X$ is linearly independent.

Thus we see that every chain in $\mathcal{P}$ has an upper bound. Hence by Zorn's Lemma, $\mathcal{P}$ has a maximal element, let us denote it by $\mathcal{B}$. Thus $\mathcal{B}$ is a maximal linearly independent subset of V, and by Proposition 7.1.28, $\mathcal{B}$ is a basis of V. □

Example 7.1.38. We have shown that F^n has a basis, namely $\{e_1, \dots, e_n\}$.

Example 7.1.39. It is not immediately clear that $\mathbb{R}$ even has a basis over $\mathbb{Q}$, nevertheless, Theorem 7.1.36 shows that $\mathbb{R}$ does indeed have a basis over $\mathbb{Q}$.

Remark 7.1.40. Having defined the notion of basis, we can not settle the question of identifying all subspaces of $\mathbb{R}^2$. They are indeed either 0, $\mathbb{R}^2$, or $\mathbb{R}w$ for some $0 \neq w \in \mathbb{R}^2$.

Having a basis around is quite handy.

Theorem 7.1.41. *Let $T : V \to W$ be a linear map, and $\mathcal{B} = \{v_1, \dots, v_n\}$ a basis of V. If $S : V \to W$ is another linear map such that $S(v_i) = T(v_i)$ for all $1 \leq i \leq n$, then $S = T$.*

Proof. Let $v \in V$. Then since $\mathcal{B}$ is a basis, there is a unique way of writing $v = \sum_{i=1}^n a_i v_i$ with $a_i \in F$. We now compute

$$S(v) = S\left(\sum_{i=1}^n a_i v_i\right) = \sum_{i=1}^n a_i S(v_i) = \sum_{i=1}^n a_i T(v_i) = T\left(\sum_{i=1}^n a_i v_i\right) = T(v),$$

thus $S = T$. □

Corollary 7.1.42. *Let $\mathcal{B} = \{v_1, \dots, v_n\}$ be a basis of V. Let $\tau : \mathcal{B} \to W$ be a map of sets. Then there is a unique linear map $T : V \to W$ such that $T(v_i) = \tau(v_i)$.*

Proposition 7.1.43. *Let V be an F-vector space of dimension n. Then $V \cong F^n$.*

Proof. If $n = 0$, then $V = \{0\} = F^0$. So we may assume $n \geq 1$, V is nonzero, and thus by Theorem 7.1.36 V has a basis $\mathcal{B} = \{v_1, \dots, v_n\}$.

We define a map $T : F^n \to V$, obtained as the unique linear map extending the map

$$\tau : \{e_i, \dots, e_n\} \to \{v_1, \dots, v_n\},\ e)i \mapsto v_i.$$

Similarly, we define a map $T' : V \to F^n$ as the unique linear map extending

$$\tau' : \{v_1, \dots, v_n\} \to \{e_1, \dots, e_n\},\ v_i \mapsto e_i.$$

Then $T' \circ T : F^n \to F^n$ satisfies $T' \circ T(e_i) = e_i$. But id_{F^n} also extends the map

$$\{e_1, \dots, e_n\} \to \{e_1, \dots, e_n\},\ e_i \mapsto e_i.$$

Hence by Corollary 7.1.42, we have $T \circ T' = \mathrm{id}_{F^n}$. Similarly, $T' \circ T : V \to V$ is the unique linear map extending $\mathrm{id}_{\mathcal{B}}$, hence $T' \circ T = \mathrm{id}_V$. Thus T and T' are inverse isomorphisms, and $V \cong F^n$. □

Remark 7.1.44. The isomorphism $V \cong F^n$ required a *choice*, namely a basis $\mathcal{B}$ of V. For different bases, the isomorphisms can be different!

Another way of saying this is that, using a basis, we can represent vectors with coordinates. But different bases may give us different coordinates.

Definition 7.1.45. Let V be a vector space with a basis $\mathcal{B} = \{v_i\}_{i\in I}$. For $v \in V$, we write

$$[v]_{\mathcal{B}} = (a_i)_{i\in I} \text{ if } v = \sum_{i\in I} a_i v_i.$$

Example 7.1.46. Let $\mathcal{B} = \{e_1, e_2\}$ be the standard basis of $\mathbb{R}^2$. Then for $v = (v_1, v_2) \in \mathbb{R}^2$, we have $[v]_{\mathcal{B}} = (v_1, v_2)$.

Let $\mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$. This is a basis of $\mathbb{R}^2$. For $v = w_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + w_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, we have $[v]_{\mathcal{C}} = (w_1, w_2)$.

Note that $w_2 \neq v_2$. Indeed, you should check that $v_1 = w_1$ and $v_2 = w_1 + w_2$.

7.2. **Linear maps.** We now turn to understanding linear maps between vector spaces. The main theorem is that, once we choose bases, every linear map can be represented by a matrix. We'll need to set some notation first.

Definition 7.2.1 (Representing Matrix)**.** Let $T : V \to W$ be a linear map. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis of V, and let $\mathcal{B}' = \{w_1, \ldots, w_m\}$ be a basis of W. Then we may write

$$T(v_j) = a_{1j} w_1 + \cdots + a_{mj} w_m$$

for some unique $a_{ij} \in F$. We write

$$[T]^{\mathcal{B}}_{\mathcal{B}'} \coloneqq \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

called the *representing matrix* for T in the bases $\mathcal{B}$ of V and $\mathcal{B}'$ of W.

Example 7.2.2. Let $\mathcal{B}$ be the standard unit basis of $\mathbb{R}^2$, and let $\mathcal{B}' = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$. What is $[\mathrm{id}_{\mathbb{R}^2}]^{\mathcal{B}}_{\mathcal{B}'}$?

Well, $[\mathrm{id}_{\mathbb{R}^2}]^{\mathcal{B}}_{\mathcal{B}'}$ is defined by the numbers a_{ij} defined by

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mathrm{id} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = a_{11} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + a_{21} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \mathrm{id} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = a_{12} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + a_{22} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Thus we see that

$$[\mathrm{id}_{\mathbb{R}^2}]^{\mathcal{B}}_{\mathcal{B}'} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

Definition 7.2.3 (Linear map associated to matrix)**.** Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis of V, and let $\mathcal{B}' = \{w_1, \ldots, w_m\}$ be a basis of W. Given an $m \times n$ matrix $A = (a_{ij})$ with $a_{ij} \in F$, we obtain a linear map $T_A : V \to W$ obtained as the unique linear map extending the map

$$\tau : \mathcal{B} \to W, \; v_j \mapsto \sum_{i=1}^{m} a_{ij} w_i.$$

Theorem 7.2.4 (Matrix-Map Correspondence)**.** *Given vector spaces V and W and bases $\mathcal{B}$ of V, and $\mathcal{B}'$ of W, the association*

$$\begin{array}{ccc} \mathrm{Hom}_F(V,W) & & \mathcal{M}_{m\times n}(F) \\ \| & & \| \\ \{\text{linear maps } T: V\to W\} & \longrightarrow & \{m\times n \text{ matrices with entries in } F\} \end{array}$$

$$T \longmapsto [T]^{\mathcal{B}}_{\mathcal{B}'} \qquad (\,[\,]^{\mathcal{B}}_{\mathcal{B}'})$$

$$T_A \longmapsto A \qquad (T_\cdot)$$

is an isomorphism (of vector spaces).

Proof.

Exercise 7.2.5. Show that $\mathrm{Hom}_F(V,W)$ is an V-vector space, with addition and scaling defined pointwise as with functions.

That $M_{m\times n}(F)$ is a vector space is seen by component-wise addition and scaling.
Since T and $T_{[T]^{\mathcal{B}}_{\mathcal{B}'}}$ both extend the map $\mathcal{B}\to V,\ v_i\mapsto v_i$, we have

$$T = T_{[T]^{\mathcal{B}}_{\mathcal{B}'}},$$

hence $T_\cdot \circ [\,]^{\mathcal{B}}_{\mathcal{B}'} = \mathrm{id}_{\mathrm{Hom_F}(V,W)}$.

We observe that $[T_A]^{\mathcal{B}}_{\mathcal{B}'}$ is the matrix associated to the linear map

$$T_A(v_j) = \sum a_{ij} w_i,$$

thus the j^{th} column of $[T_A]^{\mathcal{B}}_{\mathcal{B}'}$ is the j^{th} column of A. Hence $A = [T_A]^{\mathcal{B}}_{\mathcal{B}'}$, and so $[\,]^{\mathcal{B}}_{\mathcal{B}'}\circ T_\cdot = \mathrm{id}_{\mathcal{M}_{m\times n}(F)}$.

Exercise 7.2.6. Check that the maps $[\,]^{\mathcal{B}}_{\mathcal{B}'}: \mathrm{Hom}_F(V,W)\to\mathcal{M}_{m\times n}(F)$ and $T_\cdot:\mathcal{M}_{m\times n}(F)\to \mathrm{Hom}_F(V,W)$ are linear.

Thus $[\,]^{\mathcal{B}}_{\mathcal{B}'}$ and $T_\cdot$ are isomorphisms, as wash to be shown. □

Remark 7.2.7. Another perspective on Theorem 7.2.4 is that this correspondence goes via a choice of isomorphisms $\mathcal{B}: F^n\to V$ and $\mathcal{B}': F^m\to W$ obtained by choosing the bases $\mathcal{B}$ and $\mathcal{B}'$ of V and W, respectively. Namely, we have a commutative diagram

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ {\scriptstyle \mathcal{B}}\uparrow & & \uparrow{\scriptstyle \mathcal{B}'} \\ F^n & \xrightarrow[[T]^{\mathcal{B}}_{\mathcal{B}'}]{} & F^m \end{array} \qquad \begin{array}{c} v_i \\ \uparrow \\ e_i \end{array} \qquad \begin{array}{c} w_i \\ \uparrow \\ e_i \end{array}$$

and we see that both T and $[T]^{\mathcal{B}}_{\mathcal{B}'}$ really represent the same linear maps with the bases isomorphisms $\mathcal{B}$ and $\mathcal{B}'$ in between. That is

$$[T]^{\mathcal{B}}_{\mathcal{B}'} = \mathcal{B}'^{-1} \circ T \circ \mathcal{B}.$$

7.3. **Determinants.** We give a rigorous treatment of determinants of matrices, motivated by the question "When is a linear map invertible?" We'll focus on finite dimensional vector spaces in this subsection.

Remark 7.3.1. If V and W are finite dimensional vector spaces, if $T : V \to W$ is invertible, then $V \cong W$, and so $\dim_F(V) = \dim_F(W)$. So for any choice of bases $\mathcal{B}$ of V and $\mathcal{B}'$ of W $[T]^{\mathcal{B}}_{\mathcal{B}'}$ is a square $n \times n$ matrix.

Moreover, from Remark 7.2.7, we see that T is invertible if and only if $[T]^{\mathcal{B}}_{\mathcal{B}'}$ is invertible, namely, $[T]^{\mathcal{B}}_{\mathcal{B}'}{}^{-1} = \mathcal{B}^{-1} \circ T^{-1} \circ \mathcal{B}'$.

Thus to answer the question of invertibility of linear maps, we focus only on matrices $A : F^n \to F^n$.

We first recall how invertibility interacts with row operations.

Definition 7.3.2. If a matrix A can be transformed into A' using row operations, we call A *row equivalent* to A'.

Definition 7.3.3. A matrix A is in *row echelon form* if

(1) all zero rows are below all non-zero entries, and
(2) for any non-zero row, the first non-zero entry (called a *pivot*) is strictly to the right of the leading entry in the row above.

Example 7.3.4. The matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is in row echelon form.

The matrix $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{pmatrix}$ is in row echelon form.

The matrix $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 2 & 0 & 0 \end{pmatrix}$ is not in row echelon form.

Theorem 7.3.5. *Any matrix can be put into row echelon form using row operations. Call this form of A the* row echelon form of A.

Definition 7.3.6. A matrix A is in *reduced row echelon form* if all non-zero rows are non-zero only in pivots, and all pivots are 1.

Theorem 7.3.7. *Let A be an $n \times n$ matrix with entries in F. The following are equivalent:*

(1) A is invertible
(2) The columns of A are a basis of F^n
(3) The echelon form of A has a pivot in every column and every row
(4) A is row-equivalent to the identity matrix $I_n \coloneqq \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$
(5) The reduced row echelon form of A is the identity matrix

Proof.

Exercise 7.3.8. Give a proof.

□

We now recall how row operations correspond to multiplication by certain matrices. Namely, row operations correspond to multiplying on the left by certain matrices.

Definition 7.3.9. An elementary matrix is an $n \times n$ matrix of the form

(I) $P_{ij} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots & & \vdots \\ \vdots & & 0 & \cdots & 1 & \cdots \\ 0 & \cdots & \vdots & \ddots & \vdots & \\ \vdots & & 1 & \cdots & 0 & \\ 0 & \cdots & \cdots & & & 1 \end{pmatrix}$ where the j^{th} row is e_i and the i^{th} row is e_j, which switches rows i and j.

(II) $D_i(\lambda) = \begin{pmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ 0 & & & & 1 \end{pmatrix}$ which scales row i where λ appears by $\lambda \in F$.

(III) $E_{ij}(\lambda) = \begin{pmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & 1 & & \\ & & \lambda & \ddots & \\ 0 & & & & 1 \end{pmatrix}$ which replaces row i by row i $+ \lambda \cdot$ row j, where λ is in position ij.

Remark 7.3.10. The elementary matrices are clearly invertible because we can undo row operations.

Theorem 7.3.11. *Any invertible matrix is a product of elementary matrices.*

Proof. Let A be invertible. Then A is row equivalent to the identity matrix I_n, so there are elementary matrices $E_1, \dots, E_N$ such that $E_N \cdots E_1 A = I_n$, thus $A = E_1^{-1} \cdots E_N^{-1}$. □

Our goal in understanding determinants will be to define a map $\mathcal{M}_{n\times n} \overset{\det}{\to} F$ such that $A \in \mathcal{M}_{n\times n}$ is invertible if and only if $\det(A) \neq 0$. Thus $\det(A)$ *determines* whether A is invertible.

What properties should det have? It turns out that det plays slightly more nicely with columns than with rows. Recall that for a matrix $A = (A^1 \cdots A^n)$ with columns $A^i \in F^n$, A is invertible if and only if the columns $\{A^1, \dots, A^n\}$ are a basis of F^n. Moreover,

$$\{A^1, \dots, A^n\} \text{ is a basis } \iff \{A^1, \dots A^i + \alpha A^j, \dots, A^j, \dots, A^n\} \text{ is a basis for any } \alpha \in F.$$

Thus, while it may be too much to ask, let us add the property

$$\det(A^1 \cdots A^n) = \det(A^1 \cdots A^i + \alpha A^j \cdots A^j \cdots A^n) \text{ for any } \alpha \in F \tag{D1}$$

to our list of what we want det to satisfy.

Let us also add

$$\det(A^1 \cdots \alpha A^i \cdots A^n) = \alpha \det(A^1 \cdots A^i \cdots A^n), \tag{D2}$$

as scaling by some constant $\alpha \in F$ should make det zero exactly if $\{A^1, \dots, \alpha A^i, \dots, A^n\}$ is no longer a basis, which is exactly when $\alpha = 0$ and that is when we want det to vanish.

It turns out that Equation (D1) and Equation (1) for a basis $\{A^1, \dots, A^n\}$ also imply

$$\det(A^1 \cdots A^i + B \cdots A^n) = \det(A^1 \cdots A^i \cdots A^n) + \det(A^1 \cdots B \cdots A^n) \text{ for any } B \in F^n \tag{D3}$$

which, together with Equation (D2), says that

$$\det : \underbrace{F^n \times \cdots \times F^n}_{n \text{ times}} \to F$$

is *multilinear*, that is, det is linear in each column.

Lemma 7.3.12. *If $D : F^n \times \cdots \times F^n \to F$ is multilinear, then $D(v_1, \ldots, 0, \ldots, v_n) = 0$.*

Proof. We compute

$$D(v_1, \ldots, 0, \ldots, v_n) = D(v_1, \ldots, 0 + 0, \ldots, v_n) = D(v_1, \ldots, 0, \ldots, v_n) + D(v_1, \ldots, 0, \ldots, v_n),$$

and subtracting $D(v_1, \ldots, 0, \ldots, v_n)$ from both sides gives the result. □

Since we want $\det(A) = 0$ if A is not invertible, we want det to vanish if A has any repeated columns. Thus we want det to satisfy

$$\det(A^1 \cdots A^n) = 0 \text{ if } A^i = A^j \text{ for some } i \neq j \tag{D4}$$

and in fact Equation (D4) follows from Equations (D1) to (D3).

Proof. We compute

$$\begin{aligned} \det(A^1 \cdots A^i \cdots A^i \cdots A^n) &= \det(A^1 \cdots A^i - A^i \cdots A^i \cdots A^n) \\ &= \det(A^1 \cdots 0 \cdots A^i \cdots A^n) \\ &= 0 \text{ by Lemma 7.3.12.} \end{aligned}$$

□

We now want to capture all of the properties we have discussed thus far.

Definition 7.3.13. A multilinear map $D : F^n \times \cdots \times F^n \to F$ is called *alternating* if $D(v_1, \ldots, v_n) = 0$ if $v_i = v_j$ for some $i \neq j$.

Lemma 7.3.14. *If $D : F^n \times \cdots \times F^n \to F$ is multilinear and alternating, then*

$$D(v_1, \ldots, v_n) = D(v_1, \ldots v_i, \ldots \alpha v_j, \ldots, v_n)$$

for any $\alpha \in F$.

Proof.

Exercise 7.3.15. Give a proof

□

Remark 7.3.16. Thus a multilinear alternating map $D : F^n \times \cdots \times F^n \to F$ satisfies properties $(D1)$–$(D4)$.

Proposition 7.3.17. *Let $D : (F^n)^n \to F$ be an alternating multilinear map. Then*

$$D(v_1, \ldots, v_i, \ldots, v_j \ldots, v_n) = -D(v_1 \ldots, v_j \ldots, v_i \ldots, v_n).$$

Proof. We compute

$$\begin{aligned} 0 &= D(v_1, \ldots, v_i + v_j, \ldots, v_i + v_j \ldots, v_n) \\ &= D(v_1, \ldots, v_i, \ldots, v_i + v_j \ldots, v_n) + D(v_1, \ldots, v_j, \ldots, v_i + v_j \ldots, v_n) \\ &= D(v_1, \ldots, v_i, \ldots, v_i \ldots, v_n) + D(v_1, \ldots, v_i, \ldots, v_j \ldots, v_n) \\ &\quad + D(v_1, \ldots, v_j, \ldots, v_i \ldots, v_n) + D(v_1, \ldots, v_j, \ldots, v_j \ldots, v_n) \\ &= D(v_1, \ldots, v_i, \ldots, v_j \ldots, v_n) + D(v_1, \ldots, v_j, \ldots, v_i \ldots, v_n), \end{aligned}$$

thus

$$0 = D(v_1, \ldots, v_i, \ldots, v_j \ldots, v_n) + D(v_1, \ldots, v_j, \ldots, v_i \ldots, v_n),$$

as desired. □

Finally, we add one final condition for the determinant, namely that

(D5) $$\det(I_n) = 1.$$

With this, we can state what determinants are.

Theorem 7.3.18. *There is a unique alternating multilinear map* $\det : (F^n)^n \to F$ *such that* $\det(e_1, \dots, e_n) = 1$.

We will prove Theorem 7.3.18 shortly. First, let us see that this definition of the determinant actually does what we want.

Proposition 7.3.19. *Let* $\det : (F^n)^n \to F$ *be an alternating multilinear map. Then*

(1) $\det(0, v_2, \dots, v_n) = 0$
(2) $\det(v, v, v_3, \dots, v_n) = 0$
(3) $\det(v, \alpha v, v_3 \dots, v_n) = 0$
(4) if $v \in \operatorname{span}_F(v_2, \dots, v_n)$ *then* $\det(v, v_2, \dots, v_n) = 0$.

Proof. (1)–(3) are immediate as det is alternating multilinear. To show (4), write $v = \sum_{i=2}^n a_i v_i$ and now

$$\begin{aligned} \det(v, v_2, \dots, v_n) &= \sum_{i=2}^n a_i \det(v_i, v_2, \dots, v_n) \\ &= 0, \end{aligned}$$

as det is alternating. □

Proposition 7.3.20. *If* $D : (F^n)^n \to F$ *is alternating multilinear, then*

$$D(x_1, \dots, x_i + \alpha x_j, \dots, x_n) = D(x_1 \dots, x_n).$$

Proof. We compute

$$\begin{aligned} D(x_1, \dots, x_i + \alpha x_j, \dots, x_n) &= D(x_1, \dots, x_i, \dots, x_n) + \alpha D(x_1, \dots, x_j, \dots, x_n) \\ &= D(x_1, \dots, x_i, \dots, x_n), \end{aligned}$$

as x_j appears twice once we expand the sum. □

Theorem 7.3.21. *For* $A \in \mathcal{M}_{n\times n}(F)$, $\det(A) = 0$ *if and only if* A *is not invertible.*

Proof.

Claim. If A and B are column equivalent, then $\det(A) = 0$ if and only if $\det(B) = 0$.

Assuming the claim, we have

$$\begin{aligned} A \text{ not invertible} &\iff \text{reduced row echelon form of } A \text{ has a zero row} \\ &\iff \text{reduced row echelon form of } A \text{ has a zero column} \\ &\iff \det(\text{ reduced row echelon form of } A) = 0. \end{aligned}$$

As A is column equivalent to the reduced row echelon form of A, then A if not invertible if and only if $\det(A) = 0$, as desired.

Proof of claim. Column operations are scaling by a non-zero $\alpha \in F$, switching columns, or adding a scale of a column to another column. Note that by Remark 7.3.16, det satisfies Equation (D2). Hence for $\alpha \neq 0$ we have

$$0 = \det(A^1 \cdots \alpha A^i \cdots A^n) = \alpha \det(A^1 \cdots A^n) \iff \det(A^1 \cdots A^n) = 0.$$

Similarly, det vanishing is preserved under switching columns by Proposition 7.3.17 and under adding a scale of a column to another column by Remark 7.3.16. □

As the claim is proved, this finishes the proof. □

Recall 7.3.22. Elements of the permutation group S_n have a *sign*, namely there is a unique non-trivial homomorphism $s\,\mathrm{sgn} : S_n \to \mathbb{Z}/2\mathbb{Z} = \{\pm 1\}$ for $n \geq 2$. In particular, for a transposition $\tau \in S_n$ we have $\mathrm{sgn}(\tau) = -1$.

Unsurprisingly, because of multilinearity, alternating multilinear maps are determined by their value on a basis. However, there is a nice relationship using permutations.

Lemma 7.3.23. *Let $D : (F^n)^n \to F$ be aternating multilinear. For $1 \leq i \leq n$ let $w_i = \sum_{j=1}^{n} a_{i,j} v_j$. Then*

$$D(w_1 \ldots, w_n) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} D(v_1, \ldots, v_n).$$

Proof. We simply expand

$$D(a_{1,1}v_1 + \cdots + a_{1,n}v_n, \ldots, a_{n,1}v_1 + \cdots + a_{n,n}v_n)$$

using multilinearity into terms of the form

$$\square D(v_?, \ldots, v_?),$$

and note that terms where there is a repeated entry vanish as D is alternating. Thus the terms remaining are of the form

$$a_{i,\sigma(1)} \cdots a_{n,\sigma(n)} D(v_{\sigma(1),\ldots,v_{\sigma_n}})$$

for some permutation $\sigma \in S_n$. Writing σ as a product of transpositions, we see that

$$D(v_{\sigma(1),\ldots,v_{\sigma_n}}) = \mathrm{sgn}(\sigma) D(v_1, \ldots, v_n),$$

thus

$$D\left(\sum_{j=1}^{n} a_{1,j}v_j, \ldots, \sum_{j=1}^{n} a_{n,j}v_j\right) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} D(v_1, \ldots, v_n),$$

as claimed. □

We are now ready to prove that the map $\det : \mathcal{M}_{n\times n}(F) \to F$ satisfying (D1)–(D5) is uniquely determined.

Theorem 7.3.24. *There is a unique map $\det : \mathcal{M}_{n\times n}(F) \to F$ that is alternating multilinear in the columns such that $\det(I_n) = 1$.*

Proof. Writing $v_i = \sum_{i=1}^{n} a_{i,j} e_j$, Lemma 7.3.23 gives

$$\begin{aligned} \det(v_1, \ldots, v_n) &= \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} \det(e_1, \ldots, e_n) \\ &= \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)}, \end{aligned}$$

thus $\det(v_1, \ldots, v_n)$ is completely determined by being alternating multilinear and the property that $\det(I_n) = 1$. □

Using this we can give the usual cofactor expansion formula for computing determinants.

Definition 7.3.25. Given an $n \times n$ matrix A, denote by $A^{(i)}$ the $(n-1) \times (n-1)$ obtained by deleting the first column and the i^{th} row of A.

Exercise 7.3.26. Show that switching two rows changes the determinant by a negative sign.

Proposition 7.3.27. *For $A = (a_{ij}) \in \mathcal{M}_{n\times n}(F)$, we have*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det(A^{(i)}).$$

Proof. By moving columns around (and adjusting by the appropriate sign), we may assume that $a_{11} \neq 0$. Writing the first column as $A^1 = \sum_{i=1}^n a_{i1} e_i$, and using multilinearity of det, we have

$$\begin{aligned}
\det(A) &= a_{11} \det \begin{pmatrix} 1 & * & \cdots & * \\ 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{pmatrix} + \cdots + a_{n1} \det \begin{pmatrix} 0 & & & \\ \vdots & & A^{(n)} & \\ 0 & & & \\ 1 & * & \cdots & * \end{pmatrix} \\
&= a_{11} \det \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{pmatrix} + \cdots + a_{n1} \det \begin{pmatrix} 0 & & & \\ \vdots & & A^{(n)} & \\ 0 & & & \\ 1 & 0 & \cdots & 0 \end{pmatrix} \\
&= a_{11} \det \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{pmatrix} + \cdots + (-1)^{n+1} a_{1n} \det \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & A^{(n)} & \\ 0 & & & \end{pmatrix}.
\end{aligned}$$

And now note that

$$\begin{aligned}
\det \begin{pmatrix} 1 & 0 \\ 0 & A^{(i)} \end{pmatrix} &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)} \\
&= \sum_{\substack{\sigma \in S_n \\ \sigma(1)=1}} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)} \\
&= \sum_{\sigma \in S_{n-1}} \operatorname{sgn}(\sigma) a_{11} \cdots a_{n\sigma(n)} \\
&= \det(A^{(i)}).
\end{aligned}$$

□

Remark 7.3.28. Many of the standard facts of a linear algebra course were proven in homework. For example that a subspace has dimension no larger than the vector space, and the rank–nullity theorem.

8. Fields and polynomials

We now begin our study of polynomials, aiming toward understanding the beautiful statements of Galois Theory, which brings together group theory and the roots of polynomial equations. We will develop quite a bit of theory along the way.

Notation. In this section, F will denote a field unless otherwise stated. In some cases, when the assumption that F be a field is necessary, we will remind ourselves that F is a field.

Definition 8.0.1. The ring of polynomials with coefficients in F is denoted by $F[x]$.

Slogan. $F[x]$ is like $\mathbb{Z}$.

First, $F[x]$ is a ring, we can add and multiply polynomials in the usual way. In $\mathbb{Z}$ we may not always be able to divide, but we can always divide with remainder.

Recall 8.0.2. We'll recall some facts about $\mathbb{Z}$.

Theorem 8.0.3 (Division algorithm in $\mathbb{Z}$)**.** *Given $a, b \in \mathbb{Z}$, there are unique $q, r \in \mathbb{Z}$ such that $b = ra + r$ with $|r| < |a|$ or $r = 0$.*

This gives the Euclidean algorithm for the greatest common divisor.

Definition 8.0.4. The greatest common divisor of a and b is $\gcd(a, b) = d$ if $d \mid a$, $d \mid b$, and if $e \in \mathbb{Z}$ also divides both a and b, then $e \mid d$.

The Euclidean algorithm for computing $\gcd(a, b)$ also shows that $\mathbb{Z}$ is a *principal ideal domain*. That is,

- $\mathbb{Z}$ is a domain (if $ab = 0$, then $a = 0$ or $b = 0$), and
- ideals of $\mathbb{Z}$ are principal (if $I \subseteq \mathbb{Z}$ is an ideal, then $I = (d)$ for some $d \in \mathbb{Z}$. In fact, one can take d to be the smallest non-negative element of I.)

Example 8.0.5. For $a, b \in \mathbb{Z}$, the ideal generated by a and b,

$$(a, b) \coloneqq \{n \in \mathbb{Z} \mid n = ax + by \text{ for some } x, y \in \mathbb{Z}\}$$

is principally generated by $\gcd(a, b)$, i.e., $(a, b) = (\gcd(a, b))$.

The ring of polynomials over a field enjoys similar properties.

Definition 8.0.6. Let $p(x) = a_n x^n + \cdots + a_1 x + a_0 \in F[x]$. The *degree of $p(x)$* is the largest d such that $a_d \neq 0$, and we write $\deg p = d$.

Lemma 8.0.7. *Let F be a domain, and $f, g \in F[x]$. We have $\deg(fg) = \deg(f) + \deg(g)$ and $\deg(f + g) \leq \max\{\deg(f), \deg(g)\}$.*

Remark 8.0.8. This is no longer true if F is replaced with a ring that is not a domain! Namely in $\mathbb{Z}/6\mathbb{Z}[x]$ we have

$$(2x + 1)(3x) = 6x^2 + 3x = 3x.$$

Theorem 8.0.9 (Division algorithm for polynomials)**.** *Given $a, b \in F[x]$, with $a \neq 0$, there are unique $q, r \in F[x]$ such that*

$$b = qa + r$$

with $\deg(r) < \deg(a)$ or $r = 0$.

Proof. We first show that q and r exist.

We'll fix some notation, let

$$a = a_n x^n + \cdots + a_0, \ a_n \neq 0,$$

and

$$b = b_k x^k + \cdots + b_0, \ b_k \neq 0.$$

If $k < n$, then

$$b = 0 \cdot a + b$$

is of the required form. So we may suppose $k \geq n$. We proceed by induction on k.

Note that

$$b - \frac{b_k}{a_n} x^{k-n} a$$

has degree $< k = \deg(b)$. So by induction, there are $\widetilde{q}, r \in F[x]$ such that $\deg(r) < \deg(a)$ and

$$b = \underbrace{(\frac{b_k}{a_n}x^{k-n} + \widetilde{q})a}_{q} + r,$$

and letting $q = (\frac{b_k}{a_n}x^{k-n} + \widetilde{q})a$, we have

$$b = qa + r$$

with $\deg(r) < \deg(a)$, as desired.

To prove that q and r are unique, suppose that

$$b = qa + r + \widetilde{q}a + \widetilde{r}$$

with $\deg(r), \deg(\widetilde{r}) < \deg(a)$. Then

$$a(q - \widetilde{q}) = \widetilde{r} - r,$$

and note that $\deg(\widetilde{r} - r) < \deg(a)$.

If $q - \widetilde{q} \neq 0$, then $\deg(a(q - \widetilde{q})) \geq \deg(a)$, contradicting the fact that

$$\deg(a(q - \widetilde{q})) = \deg(\widetilde{r} - r) < a.$$

Thus $q - \widetilde{q} = 0$, hence $\widetilde{r} - r = 0$, and thus

$$q = \widetilde{q},\ r = \widetilde{r},$$

as desired. □

Definition 8.0.10. For $a, d \in F[x]$, we write $d \mid a$ and say "d divides a" if $a = qd$ for some $q \in F[x]$.

As in $\mathbb{Z}$, the division algorithm gives us a Euclidean algorithm.

Theorem 8.0.11. *For $a, b \in F[x]$, a and b have a greatest common divisor $d \in F[x]$ such that*

- *$d \mid a$, $d \mid b$, and*
- *if $e \mid a$ and $e \mid b$, then $e \mid d$.*

Moreover, $\gcd(a, b) = fa + gb$ for some $f, g \in F[x]$.

Proof. Observe first that

$$\{\text{common divisors } e \text{ of } a \text{ and } b\} = \{\text{common divisors } e \text{ of } a \text{ and } b + ka \text{ for any } k \in F[x]\}.$$

Indeed, if e is a common divisor of a and b, then certainly e divides a and $b + ka$. Conversely, if e divides a and $b + ka$, then e divides a and $b = (b + ka) - ka$.

Now we apply the division algorithm to find $d = \gcd(a, b)$, as

$$\{\text{common divisors of } a \text{ and } b\} = \{\text{common divisors of } a \text{ and } \underbrace{b - qa}_{=r}\}.$$

If $\deg(a) \leq \deg(b)$, then

$$\min\{\deg(r), \deg(a)\} \leq \min\{\deg(a), \deg(b)\},$$

so using the division algorithm to find smaller degree common divisors must terminate, and it terminates when $r = 0$. So we transform the pair (a, b) to $(d, 0)$ with $d = \gcd(a, b)$.

As we found d by replacing a and b with $b - ka$ or vice versa, it is clear that d is obtained in the form $d = fa + gb$ for some $f, g \in F[x]$. □

Recall 8.0.12.

Definition 8.0.13. Let A be a commutative ring (with 1). A is called a *principal ideal domain*, written PID, if A is a domain, and every ideal $I \subseteq A$ is principal, i.e. $I = (a) \coloneqq \{ra \mid r \in A\}$.

Proposition 8.0.14. *$\mathbb{Z}$ is a PID.*

Proof. Let $I \subseteq \mathbb{Z}$ be an ideal. If $I = 0$, then I is principal. So we may assume $I \neq 0$ and there is a nonzero $n \in I$. Let $d \in I$ be the smallest nonzero integer (which exists by the Well-Ordering Principle). We claim that $I = (d) = d\mathbb{Z}$. Clearly since $d \in I$, we have $d\mathbb{Z} \subseteq I$. For the other containment, let $a \in I$ and write $a = qd + r$ with $0 \leq r < d$. Then $a - qd = r \in I$, which is smaller than d, hence $r = 0$ as d is the smallest non-zero integer in I. Thus $a = qd \in d\mathbb{Z}$. Hence $I = (d)$, as claimed. □

Theorem 8.0.15. *Let F be a field. Then $F[x]$ is a PID.*

Proof. Since F is a field, if $f, g \in F[x]$ are non-zero, then fg is also non-zero as the top degree coefficients cannot cancel.

It remains to show that all ideals of $F[x]$ are principal. This is analogous to the proof that $\mathbb{Z}$ is a PID, replacing d with a polynomial of smallest degree in I. We give the details for completeness.

Let $I \subseteq F[x]$ be an ideal. If $I = 0$, then I is principal. So we may assume $I \neq 0$, and let $d \in I$ be a non-zero polynomial of lowest degree. We claim that $I = (d)$. Since $d \in I$, it is clear that $(d) \subseteq I$. Conversely, suppose that $f \in I$, and write $f = qd + r$ with $\deg(r) < \deg(d)$ or $r = 0$. Then $f - qd = r \in I$ and so $r = 0$. Thus $f = qd$, and so $I = (d)$. □

8.1. **Facts about polynomials over fields.** We collect some facts about polynomials over fields that will be useful.

Remark 8.1.1. Let $a \in F$, then there is a ring homomorphism $\mathrm{ev}_a : F[x] \to F, p(x) \mapsto p(a)$, obtained by evaluating polynomials at a.

Definition 8.1.2. For $p(x) \in F[x]$, we call $a \in F$ a *root of* $p(x)$ or a *zero of* $p(x)$ if $\mathrm{ev}_a(p) = p(a) = 0$.

Proposition 8.1.3. *Let F be a field. Then $a \in F$ is a root of $p(x) \in F[x]$ if and only if $x - a$ divides $p(x)$.*

Proof. Let a be a root of p. By the division algorithm, we have

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

with $\deg(r) < \deg(x - a) = 1$. So $r = b \in F$. Evaluating at a, we have

$$0 = p(a) = (a - a)q(a) + r = r,$$

and so $r = 0$ and $x - a$ divides p.

Conversely, if $p = (x - a)q$, then evaluating at a shows that $p(a) = 0$. □

Corollary 8.1.4. *Let F be a field. A non-zero polynomial of degree n has at most n roots in F.*

Proof. Let $p \in F[x]$ with $\deg(p) = n$. if $a_1, \ldots, a_k$ are all the roots of p in F, then by Proposition 8.1.3

$$p = (x - a_1) \cdots (x - a_k)q,$$

for some non-zero $q \in F[x]$ of degree $\deg(q) \geq 0$. Taking degrees gives

$$n = \deg(p) = \deg\left((x - a_1) \cdots (x - a_k)q\right) = k + \deg(q) \geq k,$$

thus $n \geq k$. □

8.2. **Factorization in (polynomial) rings.** Much like the prime factorization of integers in $\mathbb{Z}$, we seek to understand how we can break up polynomials into small pieces multiplicatively. We first develop some terminology, as factorization in abstract rings is more technical than in nice rings like $\mathbb{Z}$. Some facts may be familiar from a first course in abstract algebra in which rings are defined, so we state some results without proof.

Recall 8.2.1. For a ring R, we write

$$R^\times := \{r \in R \mid \text{ there exists } s \in R \text{ such that } rs = sr = 1_R\},$$

to denote the invertible elements of R. We call $R^\times$ the *units* of R.

Definition 8.2.2. Let R be a domain. An element $a \neq 0$ is called *irreducible* if $a \notin R^\times$, and whenever $a = bc$ then $b \in R^\times$ or $c \in R^\times$ (not both).

Irreducible elements play the role of prime numbers in factorization.

Exercise 8.2.3. The irreducible elements of $\mathbb{Z}$ are the prime numbers.

Example 8.2.4.
- $x^2 + 1$ is irreducible in $\mathbb{Q}[x]$.
- $x^2 + 1 = (x + i)(x - i)$ is not irreducible in $\mathbb{C}[x]$.

Example 8.2.5. By the Fundamental Theorem of Algebra, the irreducible polynomials in $\mathbb{C}[x]$ are exactly the linear polynomials (polynomials of degree 1).

Notation 8.2.6. Throughout, A will be a commutative ring with 1.

Definition 8.2.7. An ideal $I \subseteq A$ is *prime* if for any $a, b \in A$, if $ab \in I$ then $a \in I$ or $b \in I$ (or both).

Proposition 8.2.8. *An ideal $I \subseteq A$ is prime if and only if A/I is a domain.*

Exercise 8.2.9. Let $a \in A$. Then the following are equivalent:
(1) (a) is prime,
(2) whenever $a \mid bc$, then $a \mid b$ or $a \mid c$.

Proposition 8.2.10. *Let A be a domain and $0 \neq (a) \subset A$ be a non-zero ideal. If (a) is a prime ideal, then a is irreducible.*

Remark 8.2.11. This shows that prime elements are irreducible. The converse is *not* true in general! In fact, the converse is true if and only if A has unique factorization into irreducibles.

Proof. Let $a = bc$. Then $bc \in (a)$, thus $b \in (a)$ or $c \in (a)$. Without loss of generality, suppose that $b \in (a)$, and write $b = ra$ for some $r \in A$. Then

$$a = bc = rac = (rc)a,$$

and so

$$(rc)a - = (rc - 1)a = 0.$$

As A is a domain and $a \neq 0$, it must be the case that $rc - 1 = 0$, and so $rc = 1$ and $r, c \in A^\times$, as needed. □

Definition 8.2.12. Let R be a ring. A proper ideal $\mathfrak{m} \subsetneq R$ is called *maximal* if for any ideal $I \subset R$ if $\mathfrak{m} \subsetneq I \subseteq R$, then $I = R$.

Proposition 8.2.13. *An ideal $\mathfrak{m} \subset A$ is maximal if and only if $A/\mathfrak{m}$ is a field.*

Proposition 8.2.14. *Maximal ideals are prime.*

Proof. As fields are domains, a maximal ideal is prime. □

Proposition 8.2.15. *Let A be a PID and $f \in A$ irreducible. Then (f) is maximal, hence prime.*

Proof. We want to show that if $(f) \subsetneq I \subseteq A$, then $I = A$. Since $(f) \subsetneq I$, there is some $g \in I$ such that $g \notin (f)$. Consider the ideal $(f, g) \subseteq I \subseteq A$. as A is a PID, $(f, g) = (d)$, in particular $f = qd$ and $g = q'd$ for some $q, q' \in A$. As f is irreducible, $f = qd$ implies that $q \in A^\times$ or $d \in A^\times$.

If $q \in A^\times$, then $d = q^{-1}f$ and so $g = q'd = q'q^{-1}f \in (f)$, which contradicts the assumption that $g \notin (f)$. Thus $d \in A^\times$, and so $(d) = A$. Since $(f, g) = (d) = A \subseteq I \subseteq A$, we have $I = A$. □

Example 8.2.16. Quotienting by primes ideals generated by irreducible elements is a new way of getting fields. For example, $x^2 + 1 \in \mathbb{Q}[x]$ is irreducible. Thus

$$\mathbb{Q}[x]/(x^2 + 1)$$

is a field. In fact, in this field we have $x^2 = -1$, so x behaves like $\sqrt{-1}$.

We'll explore these ideas in depth soon.

8.3. **Irreducibility tests for polynomials.** We will frequently deal with polynomials over $\mathbb{Q}$ and want to quotient by irreducible polynomials to obtain new fields So we state some ways to show that polynomials with rational coefficients are irreducible. Unfortunately, time constraints did not allow for proofs.

Theorem 8.3.1 (Rational Root Theorem)**.** *Let $p(x) = a_n x^n + \cdots + a_0 \in \mathbb{Z}[x]$. If $a \in \mathbb{Q}$ is a root of p, then*

$$a = \frac{\pm\{\text{ factor of } a_0\}}{\{\text{factor of } a_n\}}.$$

Theorem 8.3.2 (Eisenstein's Criterion)**.** *Let $p \in \mathbb{Z}$ be a prime and $f(x) = a_n x^n + \cdots a_0 \in \mathbb{Z}[x]$. If $p \mid a_i$ for $0 \leq i \leq n-1$, $p \nmid a_n$, and $p^2 \nmid a_0$, then $f(x)$ is irreducible in $\mathbb{Q}[x]$.*

Example 8.3.3. $9x^4 - 6x^2 - 62$ is irreducible in $\mathbb{Q}[x]$. Use Eisenstein's Criterion with the prime $p = 2$. Thus in fact $x^4 - \frac{2}{3}x^2 - \frac{62}{9}$ is irreducible in $\mathbb{Q}[x]$.

Theorem 8.3.4. *Let $p \in \mathbb{Z}$ be prime. If $f(x) \in \mathbb{Z}[x]$ has a nontrivial factorization and $p \nmid f(x)$, then $f(x) \mod p \in \mathbb{Z}/p\mathbb{Z}[x]$ has a nontrivial factorization.*

Example 8.3.5. SAGE Demo. Try using these commands in SAGE, available online in a browser window at https://sagecell.sagemath.org/.

```
irreducible test:

R = PolynomialRing(QQ,'x')
x = R.gen()
f=x^2-1
f.is_irreducible()

# need to have the x=R.gen() for is_irreducible()

f.factor()
```

Try also

```
R = PolynomialRing(GF(3), 'x')

# here GF(3) means Z/3Z the finite field with 3 elements.

# GF(q) is the finite field with q elements, and q needs to be a power of a prime!!
```

```
f(x) = x^12 + 3*x^8 - 4*x^6 + 3*x^4 + 12*x^2 + 5

g(x) = x^(3^6)-x

f(x).gcd(g(x))

f.factor()

g.factor()

#try also GF(67) and factoring f to check whether f is irreducible

# define instead of f(x)
# f =  x^12 + 3*x^8 - 4*x^6 + 3*x^4 + 12*x^2 + 5

#add in x = R.gen()

# and use f.factor()
```

8.4. **Unique Factorization (in polynomial rings).** We're familiar with the factorization of integers into prime numbers. In fact, you might have seen this as a theorem in a previous introduction to proofs or algebra class.

Definition 8.4.1 (prime number)**.** An integer $2 \leq p \in \mathbb{Z}$ is called *prime* if whenever $p = mn$ with $m, n \in \mathbb{Z}_{>0}$, then $n = 1$ or $m = 1$.

Theorem 8.4.2 (Fundamental Theorem of Arithmetic)**.** *Let $n \in \mathbb{Z}_{\geq 2}$, then n can be expressed as a product of prime numbers (not necessarily distinct)*

$$n = p_1 \cdots p_k$$

uniquely determined up to order.

To generalize this to arbitrary rings, we distill the key ideas we want from a unique factorization: that we can factor every element into a product of irreducible elements, and that this factorization is unique up to order. However, this doesn't actually give only one way to write an element as a product of irreducible elements (even up to order)!

Example 8.4.3. Note that for a prime number $p \in \mathbb{Z}$, both p and $-p$ are irreducible! So we have many ways to express an integer as a product of irreducible elements. For example

$$12 = 2 \cdot 2 \cdot 3 = (-2) \cdot (-2) \cdot 3 = (-2) \cdot 2 \cdot (-3).$$

The key takeaway is that different factorizations into irreducibles may also differ by units. Since if $r \in R$ is irreducible, and $u \in R^\times$, then ur is also irreducible! So, to get a "unique" factorization int irreducibles, we must accept not just reordering, but also differing by units.

Thus, we define the notion of unique factorization as follows.

Definition 8.4.4. A commutative domain A is called a *unique factorization domain* (UFD) if

- for any $0 \neq a \in A$ and $a \notin A^\times$, then $a = r_1 \cdots r_k$ for some irreducible $r_1, \ldots, r_k \in A$, and
- if $a = r_1 \cdots r_k = q_1 \cdots q_s$ for some irreducible $r_i, q_i \in A$ then $k = s$ and there is a permutation $\sigma \in S_k$ such that $r_i = u_i q_{\sigma(i)}$ for some units $u_i \in A^\times$.

Thankfully, for a field F, $F[x]$ is a UFD, which we will now prove after a quick lemma.

Lemma 8.4.5. *Let $f \in F[x]$ be irreducible. Then f is prime.*

Proof. Assume that $f \mid gh$. We wish to show that $f \mid g$ or $f \mid h$. So suppose that $f \nmid g$, and we must show that $f \mid h$. Since $f \nmid g$, $\gcd(f, g) = 1$, hence there exist polynomials $h_1, h_1 \in F[x]$ such that

$$1 = h_1 f + h_2 g.$$

Multiplying by h gives

$$h = hh_1 f + hh_2 g.$$

But since $f \mid gh$, we have $gh = fh_3$ for some $h_3 \in F[x]$, whereby

$$h = hh_1 f + hh_2 g = (hh_1 + h_2 h_3) f,$$

and $f \mid g$, as desired. □

Proposition 8.4.6. *Let F be a field, then $F[x]$ is a UFD.*

Proof. We first show that a factorization into irreducibles exists. This is by induction on the degree.

Let $f \in F[x]$. If $\deg(f) = 1$, then $f = ax + b$, and is irreducible by the division algorithm.

For the inductive step, let $f \in F[x]$. If f is irreducible, then we are done. Else f factors as $f = gh$ with $g, h \notin F[x]^\times$, and $0 \leq \deg(g), \deg(h) < \deg(f)$. Thus by induction we have factorizations

$$g = r_1 \cdots r_i, \ h = r_{i+1} \cdots r_k$$

for irreducible $r_1, \ldots, r_k \in F[x]$. And thus $f = r_1 \cdots r_k$, as desired.

To show uniqueness, suppose that

$$f_1 \cdots f_n = f = g_1 \cdots g_k$$

are two factorizations of f into irreducibles. Since f_1 is irreducible it is prime by Lemma 8.4.5, (f_1) is a prime ideal. Hence as $g_1 \cdots g_k = f_2 \cdots f_n \cdot f_1 \in (f_1)$, one of the $g_i \in (f_1)$. Thus $g_i = u_1 f_1$ for some $u \in F[x]$. As g_i is irreducible, and $f_1 \notin F[x]^\times$, $u \in F[x]^\times$. Thus we have

$$f_1 f_2 \cdots f_n = f = g_1 g_2 \cdots g_i \cdots g_k = u f_1 g_1 g_2 \cdots g_n,$$

and as $F[x]$ is a domain, we may cancel f_1 from both sides, yielding

$$f_2 \cdots f_n = u g_1 \cdots g_{i-1} g_{i+1} \cdots g_k.$$

Continuing, we have, after relabeling,

$$1 = u_1 \cdots u_n g_{k-n+1} \cdots g_k,$$

thus $n = k$ and $g_{\sigma(i)} = u_i f_i$ for some $\sigma \in S_n$. □

8.4.1. *PID implies UFD.* We did not cover the fact that every PID is a UFD in class, as we focused on polynomial rings and were headed directly toward Galois theory. I include a proof in these notes for completeness, and because I wish to include it in future courses.

Definition 8.4.7. A ring R is called a *factorization domain* (FD) if R is a domain and

- for any $0 \neq a \in A$ and $a \notin A^\times$, then $a = r_1 \cdots r_k$ for some irreducible $r_1, \ldots, r_k \in A$.

Proposition 8.4.8. *Let R be a FD. Then R is a UFD if and only if every irreducible element is prime.*

Proof. Suppose first that R is a UFD, and let $p \in R$ be irreducible. We want to show that p is prime. So suppose that $p \mid ab$ for some $a, b \in R$, and we aim to show that $p \mid a$ or$p \mid b$. As $p \mid ab$, there is some $c \in R$ such that $ab = pc$. Since R is a factorization domain, we can factor a, b, c into some product of irreducible elements

$$a = p_1 \cdots p_k, \ b = q_1 \cdots q_\ell, \ c = r_1 \cdots r_m.$$

Thus we have two irreducible factorizations

$$p_1 \cdots p_k \cdot q_1 \cdots q_\ell = ab = pc = p r_1 \cdots r_m.$$

By assumption, R is a UFD and p is irreducible, hence there exist unit $u \in R^\times$ such that $p = up_i$ for some i or $p = uq_j$ for some j. In either case, $p \mid a$ or $p \mid b$, as was to be shown.

For the converse, suppose that every irreducible element of R is prime. Since R is a factorization domain, it remains to show that the two factorizations are related by units and a permutation. Suppose that we have two factorizations

$$p_1 \cdots p_\ell = q_1 \cdots q_m$$

into irreducibles, and assume without loss of generality that $\ell \leq m$. Since p_1 is irreducible, it is by assumption also prime, hence (p_1) is a prime ideal. Thus as $p_1 \cdots p_\ell = q_1 \cdots q_m$, we have $q_1 \dots q_m \in (p_1)$, and thus one of the $q_i \in (p_1)$. Appyling a permutation to the q_i, we may assume that $q_1 \in (p_1)$ and so $q_1 = u_1 p_1$ for some $u_1 \in R$. As q_1 is irreducible, and $p_1 \notin R^\times$, we see that $u_1 \in R^\times$ is a unit. Continuing in this way, we obtain

$$1 = u_1 \cdots u_\ell g_{\ell+1} \cdots g_m,$$

hence $\ell = m$ and the irreducibles $p_1 \cdots p_\ell$ and $q_1 \cdots q_\ell$ are related by a permutation and units, as desired. □

Theorem 8.4.9. *Every PID is a UFD.*

Proof. Suppose that R is a PID. We begin by showing that R is a factorization domain. Suppose for contradiction that there exists $0 \neq a \in R$ with $a \notin R^\times$ that does not factor as a product of irreducible elements. Thus a itself is not irreducible, so we can factor

$$a = a_1 b_1$$

for some $a_1, b_1 \in R$ with $a_1, b_1 \notin R^\times$. If both a_1 and b_1 factor into a product of irreducible elements, then so would a. Hence we may assume, without loss of generality that a_1 does into factor as a product of irreducible elements. Since a_1 itself is not irreducible, we may factor

$$a_1 = a_2 b_2$$

with $a_2, b_2 \notin R^\times$. And, as before, we may assume that a_2 does not factor into irreducibles. Continuing this process, we construct an infinite sequence

$$a = a_0, a_1, a_2, \dots$$

where for every $i \geq 0$ we have $a_i = a_{i+1} b_{i+1}$ for some $b_{i+1} \notin R^\times$.

We observe that the ideals (a_i) fit into a chain of increasing ideals

$$(a_0) \subsetneq (a_1) \subsetneq \cdots (a_n) \subsetneq (a_{n+1}) \subsetneq \cdots .$$

The union of the ideals (a_i) is also an ideal. Since R is a PID, there is some $c \in R$ such that

$$(c) = \bigcup_{i=0}^{\infty} (a_i).$$

Since c is in the union of the ideals, it must be in some ideal (a_n) for some n, which implies that $(c) = (a_k)$ for all $k \geq n$, contradicting the strict containments $(a_n) \subsetneq (a_{n+1})$ of the chain of ideals. Thus our original assumption that no a_1 can be factored into irreducibles must be false, and so a does indeed factor into a product of irreducibles. Thus R is a factorization domain.

We now want to show that R is a UFD. By Proposition 8.4.8, it suffices to show that every irreducible element is prime. Suppose that p is irreducible. To show that p is prime, it suffices to show that (p) is maximal, as maximal ideas are prime.

Suppose that $I \subseteq R$ is an ideal such that $(p) \subseteq I \subseteq R$. We must prove that either $I = (p)$ or $I = R$. Since R is a PID, $I = (r)$. Since $p \in (p) \subseteq (r)$, we have $p = rs$ for some $s \in R$. As p is irreducible, either r or s is a unit. If r is a unit, then $I = (r) = R$, and if a is a unit then $I = (p)$, as desired. □

Remark 8.4.10. Rings where every increasing chain of ideals stabilizes (such as PIDs) are called *Noetherian* in honor of Emmy Noether, a major figure in algebra. Noetherian rings are ubiquitous in algebraic geometry, and are near and dear to me.

9. Fields and Galois Theory

We can now begin the last main section of the course, Galois Theory. The goal is to state and prove the Fundamental Theorem of Galois Theory, which is a statement relating the structure of Galois group (measuring symmetries of roots a polynomial) and fields containing the roots. It will, as usual, take us quite some time to develop the main ideas, but the payoff and applications are truly beautiful.

Remark 9.0.1. Our treatment of Galois Theory is heavily inspired by Dummit and Foote's great book *Abstract Algebra*, as well as the first 3 chapters of Milne's *Fields and Galois Theory*, available here (https://www.jmilne.org/math/CourseNotes/FT.pdf), as well as previous course notes graciously provided by Mike Lipnowski. In our development of Galois Theory, we reproduce some proofs from Dummit and Foote or Milne, and urge the reader to read those books instead.

Notation 9.0.2. When in doubt, L, K, F will usually denote fields.

9.1. Field Extensions.

We begin by studying fields, and working toward fundamental notions in Galois Theory.

Example 9.1.1. We can check that x^2-2 is irreducible in $\mathbb{Q}[x]$. Indeed, if x^2-2 factored into smaller non-unit polynomials, then it would factor as a product of linear polynomials, and would have a root. But the roots of x^2-2 are exactly $\pm\sqrt{2} \notin \mathbb{Q}$. Hence x^2-2 is irreducible.

Thus the ideal $(x^2-2) \subset \mathbb{Q}[x]$ is maximal, by Proposition 8.2.15, thus $\mathbb{Q}[x]/(x^2-2)$ is a field. Denote by $\overline{x}$ the coset $x \pmod{x)^2-2}$, which is the image of x under the natural quotient map $\mathbb{Q}[x] \to \mathbb{Q}[x]/(x^2-2)$. Then elements of $\mathbb{Q}[x]/(x^2-2)$ looks like

$$a_0 + a_1 x + a_2 x^2 + a_3 x^3 \cdots \pmod{x)^2-2},$$

thus we have $x^2 = 2 \pmod{x)^2-2}$, hence elements of $\mathbb{Q}[x]/(x^2-2)$ look like $a\overline{x}+b$ with $a, b \in \mathbb{Q}$, and multiplication is defined by

$$(a\overline{x}+b)(a'\overline{x}+b') = aa'\overline{x}^2 + ab'\overline{x} + a'b\overline{x} + bb' = 2aa' + (ab'+a'b)\overline{x} + bb' = (ab'+a'b)\overline{x} + 2aa' + bb'.$$

Note that we have an inclusion $\mathbb{Q} \hookrightarrow \mathbb{Q}[x]/(x^2-2), b \mapsto b$, thus $\mathbb{Q}[x]/(x^2-2)$ is a larger field containing $\mathbb{Q}$.

Note that x^2-2 has two roots, namely $\pm\sqrt{2}$, and the element $\overline{x} \in \mathbb{Q}[x]/(x^2-2)$ satisfies $\overline{x}^2 = 2$. Thus $\overline{x}$ plays the role of a root of x^2-2. And $\pm\overline{x}$ are both roots of x^2-2 in $\mathbb{Q}[x]/(x^2-2)$.

We observe that there is in fact a ring homomorphism $\mathbb{Q}[x]/(x^2-2) \to \mathbb{Q}[x]/(x^2-2)$ which is defined by sending $\overline{x} \mapsto -\overline{x}$ and leaving $\mathbb{Q}$ fixed. That this is a ring homomorphism may not be immediate, and we should check this. Let

$$\tau : \mathbb{Q}[x]/(x^2-2) \to \mathbb{Q}[x]/(x^2-2),\ a\overline{x}+b \mapsto -a\overline{x}+b.$$

We check that τ is a ring homomorphism. Namely, $\tau(0)=0$, $\tau(1)=1$, and that τ distributes over addition and multiplication. That $\tau(f+g) = \tau(f)+\tau(g)$ is clear, as it simply switches signs on $\overline{x}$.

Exercise 9.1.2. Finish checking that τ is a ring homomorphism.

We note that τ fixes $\mathbb{Q}$. In fact, any ring homomorphism $\mathbb{Q}[x]/(x^2-2) \to \mathbb{Q}[x]/(x^2-2)$ that fixed $\mathbb{Q}$ is either τ or id. Since τ just changes the sign of $\overline{x}$, $\tau \circ \tau = \mathrm{id}$, and we can define a group structure on $\{\mathrm{id}, \tau\}$ with composition, and see that $\{\mathrm{id}, \tau\} \cong \mathbb{Z}/2\mathbb{Z}$. This is no mistake, and we see that the roots $\pm\sqrt{2}$ also have a group action by $\mathbb{Z}/2\mathbb{Z}$. This relationship between field automorphisms of

$\mathbb{Q}[x]/(x^2-2)$ that fix $\mathbb{Q}$ and symmetries of the roots of $x^2-2 \in \mathbb{Q}[x]$ is exactly what Galois Theory investigates.

Definition 9.1.3 (Field Extension)**.** Let L and K be fields. We say that L *is an extension of* K if $K \subset L$ is a subfield. We write L/K to mean that L is an extension of K, and call K the *base field* of the extension.

Remark 9.1.4 (N.B.)**.** The notation L/K does not mean a quotient.

Remark 9.1.5. If L/K is an extension, then L is a vector space over K.

Example 9.1.6. $\mathbb{Q} \subset \mathbb{R}$ is an extension with uncountable basis!

Example 9.1.7. $\mathbb{R} \subset \mathbb{C}$ is an extension with basis $\{1, i\}$

Proposition 9.1.8. *Let K be a field and R a ring. Any ring homomorphism $\varphi : K \to R$ is injective.*

Proof. $\ker \varphi \subseteq K$ is an ideal. Since K is a field, the only ideals are 0 and K. Since $\varphi(1) = 1$, $\ker \varphi \neq K$, hence $\ker \varphi = 0$. □

Remark 9.1.9. If L and K are fields, and $\varphi : K \to L$ is a ring homomorphism, then $\varphi(K) \subset L$ is an extension.

Definition 9.1.10. Let L/K be an extension and $\alpha \in L$.

$$K(\alpha) := \text{ smallest subfield of } L \text{ containing } K \text{ and } \alpha.$$

For L/K an extension and $\alpha \in L$, we have a tower of extensions $K \subseteq K(\alpha) \subseteq L$.

Example 9.1.11. Let $\alpha \in \mathbb{R}$. Then $\mathbb{Q}(\alpha)$ is the smallest subfield containing $\mathbb{Q}$ and α. Since we can add, multiply, and divide (by non-zero elements) in $\mathbb{Q}(\alpha)$, the elements look like

$$\mathbb{Q}(\alpha) = \left\{ \frac{\sum_i c_i \alpha^j}{\sum_j d_j \alpha^j} \;\middle|\; c_i, d_j \in \mathbb{Q}, \sum_j d_j \alpha^j \neq 0 \right\},$$

which can be neatly packaged as

$$\mathbb{Q}(\alpha) = \left\{ \frac{p(\alpha)}{q(\alpha)} \;\middle|\; p, q \in \mathbb{Q}[x], q(\alpha) \neq 0 \right\}.$$

And we see that $\mathbb{Q}(\alpha)$ is just the values of rational functions with rational coefficients evaluated at α.

Definition 9.1.12. Let L/K be a field extension. We say that $\alpha \in L$ is *algebraic over* K if there is a non-zero polynomial $p(x) \in K[x]$ such that $p(\alpha) = 0$. If $\alpha \in L$ is not algebraic over K, we say that α is *transcendental over* K.

Example 9.1.13. $\sqrt{2} \in \mathbb{R}$ is algebraic over $\mathbb{Q}$. Indeed, $\sqrt{2}$ is a zero of $x^2 - 2 \in \mathbb{Q}[x]$. However, $\pi \in \mathbb{R}$ is transcendental.

Question 2. *For $\alpha \in \mathbb{R}$, how big is the extension $\mathbb{Q}(\alpha)/\mathbb{Q}$? More generally, if L/K is an extension and $\alpha \in L$, how big is $K(\alpha)/K$?*

Example 9.1.14. We've seen that

$$\mathbb{Q}(\sqrt{2}) = \left\{ \frac{p(\sqrt{2})}{q(\sqrt{2})} \;\middle|\; p, q \in \mathbb{Q}[x], q(\sqrt{2}) \neq 0 \right\} \subset \mathbb{R}.$$

Thus as $\sqrt{2}^2 = 2$, $p(\sqrt{2}) = a + b\sqrt{2}$ and $q(\sqrt{2}) = c + d\sqrt{2}$ with $a, b, c, d \in \mathbb{Q}$. Hence

$$\mathbb{Q}(\sqrt{2}) = \left\{ \frac{a + b\sqrt{2}}{c + d\sqrt{2}} \;\middle|\; a, b, c, d \in \mathbb{Q} \right\},$$

and since

$$\frac{a+b\sqrt{2}}{c+d\sqrt{2}}\cdot\frac{c-d\sqrt{2}}{c-d\sqrt{2}}=(\frac{ac+2bd}{c^2-2d^2})+(\frac{ac-bd}{c^2-2d^2})\sqrt{2}=a'+b'\sqrt{2},$$

we have

$$\mathbb{Q}(\sqrt{2})=\{a+b\sqrt{2}\mid a,b\in\mathbb{Q}\}.$$

Thus $\mathbb{Q}(\sqrt{2})$ is a $\mathbb{Q}$-vector space with basis $\{1,\sqrt{2}\}$ and so $\dim_{\mathbb{Q}}\mathbb{Q}(\sqrt{2})=2$. Much like $\mathbb{Q}[x]/(x^2-2)$!

This is no accident. In fact, we have a surjective ring homomorphism

$$\varphi:\mathbb{Q}[x]\to\mathbb{Q}(\sqrt{2}),\ p(x)\mapsto p(\sqrt{2}),$$

which has kernel $\ker\varphi=(d(x))$ which is a maximal ideal since $\mathbb{Q}[x]/\ker\varphi\cong\mathbb{Q}(\sqrt{2})$ is a field. Note that $x^2-2\in\ker\varphi$, so $(x^2-2)\subseteq\ker\varphi$. Since $x^2-2\in\mathbb{Q}[x]$ is irreducible, (x^2-2) is a maximal ideal, and so $\ker\varphi=(x^2-2)$, and we have

$$\mathbb{Q}[x]/(x^2-2)\cong\mathbb{Q}(\sqrt{2}).$$

Example 9.1.15. On the other hand, $\mathbb{Q}(\pi)$ is not finite dimensional over $\mathbb{Q}$. We have a ring homomorphism

$$\mathbb{Q}[x]\to\mathbb{Q}(\pi),\ p(x)\mapsto p(\pi),$$

and since π is transcendental over $\mathbb{Q}$, this map is injective. However, it is not surjective, since $\mathbb{Q}(\pi)$ is all ratios of values of polynomials with rational coefficients at π. However, since no polynomial in $\mathbb{Q}[x]$ vanishes at π, we have a ring homomorphism

$$\mathbb{Q}(x)\to\mathbb{Q}(\pi),\ \frac{p(x)}{q(x)}\mapsto\frac{p(\pi)}{q(\pi)},$$

which is surjective be definition. As $\mathbb{Q}(x)$ is a field, it is also injective. Hence

$$\mathbb{Q}(x)\cong\mathbb{Q}(\pi).$$

And $\mathbb{Q}(x)/\mathbb{Q}$ is a HUGE extension! Its basis over $\mathbb{Q}$ includes

$$\{1,x,x^2,\dots,1/x,1/x^2,\dots 1/(x+1),1/(x+1)^2,\dots,1/(x+2),1/(x+2)^2,\dots\}.$$

Think of field extensions L/K where $\dim_K L$ is finite as an extension that adds roots of polynomials to K. We'll soon see that this is indeed the case.

Proposition 9.1.16. *Let F be a field and $p(x)\in F[x]$ a non-constant polynomial. Then there exists a field extension E/F with an element $\alpha\in E$ such that $p(\alpha)=0$.*

Proof. We may assume that $p(x)$ is irreducible, as it suffices to find an extension E/F with $\alpha\in E$ a root of an irreducible factor of $p(x)$. Thus as $p(x)$ is maximal, the ideal $(p(x))\subset F[x]$ is maximal (by Proposition 8.2.15), hence $E=F[x]/(p(x))$ is a field. We have a clear ring homomorphism $F\to F[x]\to E$, hence E/F is an extension. We claim that $\alpha=\overline{x}=x+(p)\in E$ is the desired element.

Let $p(x)=a_0+a_1x+\cdots a_nx^n\in F[x]$. We compute

$$\begin{aligned}p(\alpha)&=a_0+a_1(x+(p))+a_2(x+(p))^2+\cdots+a_n(x+(p))^n\\&=a_0+(a_1x+(p))+(a_2x^2+(p))+\cdots+(a_nx^n+(p))\\&=a_0+a_1x+\cdots a_nx^n+(p)\\&=p+(p)\\&=0+(p)\\&=0\in F[x]/(p).\end{aligned}$$

Thus $\alpha=\overline{x}\in E$ is indeed a root of $p(x)$. □

Remark 9.1.17. Thus for an irreducible polynomial $p(x) \in F[x]$, $F[x]/(p(x))$ is indeed a field adding a root of $p(x)$. We also see that $\dim_F F[x]/(p(x)) = \deg(p)$, with an F-basis given by $\{1, \overline{x}, \overline{x}^2, \dots, \overline{x}^{\deg(p)-1}\}$.

Conversely, if E/F is an extension and $\alpha \in E$ is algebraic over F, then we can find a subfield $F(\alpha) \subset E$ obtained by adding the root of a unique polynomial $p_\alpha(x) \in F[x]$.

Theorem 9.1.18. *Let E/F be an extension and $\alpha \in E$ be algebraic over F. Then there is a unique irreducible monic ($a_n = 1$) polynomial of smallest degree such that $p(\alpha) = 0$. Moreover, if $f \in F[x]$ is another polynomial such that $f(\alpha) = 0$, then $p \mid f$.*

Proof. Consider the ring homomorphism

$$\mathrm{ev}_\alpha F[x] \to E,\ f(x) \mapsto f(\alpha).$$

Since α is algebraic over F, $\ker \mathrm{ev}_\alpha \neq 0$, and since $F[x]$ is a PID, we have $\ker \mathrm{ev}_\alpha = (p(x))$ for some $p(x) \in F[x]$ of degree ≥ 1 as $p(x)$ cannot be constant. If $f \in F[x]$ such that $f(\alpha) = 0$, then $f \in (p(x))$, hence $p \mid f$. Hence $p(x)$ is a polynomial of minimal degree such that $p(\alpha) = 0$.

If $q(x) \in F[x]$ is another polynomial of minimal degree such that $q(\alpha) = 0$, then $q = c \cdot p$ for some $c \in F^\times$. Since scaling by non-zero $c \in F^\times$ does not change (p), choosing $p(x) = x^n + a_{n-1}x^{n-1} + \cdots a_0$ makes the choice of generator of $\ker \mathrm{ev}_\alpha$ unique.

It remains to show that $p(x)$ is irreducible. So suppose that $p = rs$ with $r, s \in F[x]$. If r or s are both not units, then $1 \leq \deg(r), \deg(s) < \deg(p)$, and $0 = p(\alpha) = r(\alpha)s(\alpha)$. Thus as F is a domain, $r(\alpha) = 0$ or $s(\alpha) = 0$, contradicting that p is a polynomial of minimal degree with α as a root. Thus one of r or s must be a unit, and so $p(x)$ is irreducible. □

Definition 9.1.19. Let E/F be an extension, $\alpha \in E$ algebraic over F. The unique irreducible monic polynomial of minimal degree $p_\alpha(x) \in F[x]$ such that $p_\alpha(\alpha) = 0$ is called the *minimal polynomial of α (over F)*. The degree of $p_\alpha(x)$ is called the *degree of α over F*.

Proposition 9.1.20. *Let E/F be an extension, $\alpha \in E$ algebraic over F, and $p_\alpha(x) \in F[x]$ the minimal polynomial of α. Then $F(\alpha) \cong F[x]/(p_\alpha(x))$ and $\dim_F F(\alpha) = \deg(p_\alpha(x))$.*

Proof. Exercise. □

Definition 9.1.21. We call an extension E/F *algebraic* if every $\alpha \in E$ is algebraic over F.

9.2. **Finite Extensions.**

Definition 9.2.1. Let E/F be a field extension. We say E/F is *finite* if E is a finite dimensional F-vector space. The dimension $\dim_F E$ is called the *degree of E/F* and denoted $[E : F]$.

Example 9.2.2. $\mathbb{Q}(\sqrt{2})\mathbb{Q}$ is a finite extension of degree 2, while $\mathbb{Q}(\pi)/\mathbb{Q}$ is not a finite extension.

Definition 9.2.3. An extension E/F is called *simple* if $E = F(\alpha)$ for some $\alpha \in E$.

Example 9.2.4. If E/F is an extension and p_α is the minimal polynomial of $\alpha \in E$, then $[F(\alpha) : F] = \deg(p_\alpha(x))$.

Theorem 9.2.5. *Let E/F be a finite extension. Then E/F is algebraic.*

Proof. Say $[E : F] = n$ and let $\alpha \in E$. Considering E as an F-vector space, the $n + 1$ elements

$$1, \alpha, \alpha^2, \dots, \alpha^n$$

must be linearly dependent over F. This there exist $a_0, \dots, a_n \in F$ not all zero such that

$$a_0 \cdot 1 + a_1\alpha^1 + \cdots + a_n\alpha^n = 0.$$

Let $p(x) = a_0 + a_1x + \cdots a_nx^n \in F[x]$, which is a non-zero polynomial such that $p(\alpha) = 0$. □

Remark 9.2.6. Note that the polynomial $p(x)$ we found above been not be the minimal polynomial $p_\alpha(x) \in F[x]$.

We now turn our attention to some properties of finite extensions.

Proposition 9.2.7 (Tower Law)**.** *Let $A/B/C$ be a tower of finite field extensions, i.e., $C \subset B \subset A$ and A/B and B/C are both finite extensions. Then*

$$[A : C] = [A : B][B : C].$$

Proof. Let $\{a_1, \ldots, a_n\}$ be a basis for A as a B-vector space, and let $\{b_1, \ldots, b_m\}$ be a basis for B as a C-vector space. Thus $[A : B] = n$ and $[B : C] = m$. We'll show that

$$\{a_i b_j\}_{1 \le i \le n,\ 1 \le j \le m}$$

is a basis for A as a C-vector space, whereby $[A : C] = nm$, as claimed.

Let us show that $\{a_i b_j\}$ spans A over C. Let $a \in A$, then

$$a = \sum_{i=1}^{n} \beta_i a_i \text{ for some } \beta_i \in B.$$

And we can write each

$$\beta_i = \sum_{j=1}^{m} c_{ij} b_j \text{ for some } c_{ij} \in C.$$

Thus

$$\begin{aligned} a &= \sum_{i=1}^{n} \beta_i a_i \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} c_{ij} b_j \right) a_i \\ &= \sum_{\substack{1 \le i \le n \\ 1 \le j \le m}} c_{ij} a_i b_j, \end{aligned}$$

and so $\{a_i b_j\}$ span A over C.

To show that the $a_i b_j$ are linearly independent, suppose that $0 = \sum x_{ij} a_i b_j$ for some $x_{ij} \in C$. Then

$$0 = \sum_{i=1}^{n} \underbrace{\left(\sum_{j=1}^{m} x_{ij} b_j \right)}_{\in B} a_i.$$

Since the a_i are linearly independent over B, we have $\sum_{j=1}^{m} x_{ij} b_j = 0$ for each i. Since the b_j are linearly independent over C, we have $x_{ij} = 0$ for each i and each j. Thus $\{a_i b_j\}_{\substack{1 \le i \le n \\ 1 \le j \le m}}$ is indeed a basis for A over C. □

Corollary 9.2.8. *If $F_k/F_{k-1}/\cdots/F_1$ is a tower of field extensions with F_{i+1}/F_i finite, then F_k/F_1 is a finite field extension, and*

$$[F_k : F_1] = [F_k : F_{k-1}] \cdots [F_2 : F_1].$$

Definition 9.2.9. Let E/F be an extension and $\alpha_1 \ldots, \alpha_n \in E$. We define $F(\alpha_1, \ldots, \alpha_n)$ as the smallest subfield of E containing F and $\alpha_1, \ldots, \alpha_n$.

Example 9.2.10. For example, $\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}$.

Theorem 9.2.11. *Let E/F be a field extension. The following are equivalent:*

(1) E/F is finite.

(2) There exist $\alpha_1, \ldots, \alpha_n \in E$ algebraic over F such that $E = F(\alpha_1, \ldots, \alpha_n)$.

(3) There is a sequence of fields

$$F \subset F(\alpha_1) \subset \cdots \subset F(\alpha_1, \ldots, \alpha_{n-1}) \subset F(\alpha_1, \ldots, \alpha_n) = E$$

with each extension algebraic.

Proof. To show that (1) implies (2), suppose that E/F is finite and let $\{\alpha_1, \ldots, \alpha_n\}$ be a basis for E over F. Since E/F is finite, each α_i is algebraic over F. Since $F(\alpha_1, \ldots, \alpha_n) \subseteq E$ is a vector space of dimension $[E : F]$, we must have $E = F(\alpha_1, \ldots, \alpha_n)$.

To show that (2) implies (3), we observe that since each α_i is algebraic over F,

$$F(\alpha_1, \ldots, \alpha_{i-1}) \subset F(\alpha_1, \ldots, \alpha_{i-1})(\alpha_i) = F(\alpha_1, \ldots, \alpha_{i-1}, \alpha_i)$$

is finite, hence algebraic.

Finally, to show that (3) implies (1), let

$$F \subset F(\alpha_1) \subset \cdots \subset F(\alpha_1, \ldots, \alpha_n) = E$$

with $F(\alpha_1, \ldots, \alpha_i)/F(\alpha_1, \ldots, \alpha_{i-1})$ algebraic. Thus $[F(\alpha_1, \ldots, \alpha_i) : F(\alpha_1, \ldots, \alpha_{i-1})]$ is finite of degree $\deg(p_{\alpha_i}(x))$ for $p_{\alpha_i}(x) \in F(\alpha_1, \ldots, \alpha_{i-1})[x]$ the minimal polynomial. Hence $[E : F] = [F(\alpha_1 \ldots, \alpha_n) : F]$ is finite. □

We capture this theorem by saying the following.

Slogan. Every finite extension is a tower if simple extensions.

Theorem 9.2.12. *Let E/F be a field extension. If $a, b \in E$ are algebraic over F, then*

- *$a + b$ is algebraic over F,*
- *ab is algebraic over F, and*
- *if $a \neq 0$, then $\frac{1}{a}$ is algebraic over F.*

Proof. We have a tower of finite extensions $F \subset F(a) \subset F(a, b)$. Hence $F(a, b)/F$ is finite, hence algebraic. Since $a + b, ab \in F(a, b)$, they are algebraic over F. If $a \neq 0$, then $a \in F(a)$ which is a finite (hence algebraic) extension of F, thus $\frac{1}{a} \in F(a)$ is also algebraic over F. □

Corollary 9.2.13. *Let E/F be an extension. Then $F^{alg,E} := \{a \in E \mid a \text{ is algabraic over } F\}$ is a field extension of F.*

Definition 9.2.14. Let E/F be a field extension. The subfield $F^{alg,E} \subseteq E$ is called the *algebraic closure of F in E*.

Definition 9.2.15. A field K is *algebraically closed* if every non-constant polynomial $f(x) \in K[x]$ has a root in K.

Theorem 9.2.16. *A field K is algebraically closed if and only if every non-constant polynomial in $K[x]$ factors into linear factors.*

Proof. If K is algebraically closed, then for $f(x) \in K[x]$, we have $f(x) = (x - a)q(x)$ for a root $a \in K$ and we continue factoring.

Conversely, if $f(x) = (ax - b) \cdots \in K[x]$ is a product of linear factors, we note that not all a's can be zero, otherwise f would be constant, and $\frac{b}{a} \in K$ is a root. □

Remark 9.2.17. The following Theorem was not included in class, but is included in the notes for completeness.

Theorem 9.2.18. *A field K is algebraically closed if and only if the only finite field extensions are of degree 1.*

Proof. If K is algebraically closed, then any finite field extension is a tower of simple extensions each of degree 1 as all irreducible polynomials are degree 1. Hence by the Tower Law, the entire extension has degree 1.

Conversely, suppose that every finite extension is of degree 1. Then for every irreducible polynomial $f \in K[x]$, $K[x]/(f) = K$, hence f is degree 1, and thus as $K[x]$ is a UFD, every polynomial factors into linear factors. □

Theorem 9.2.19 (Algebraic Closures)**.** *Let F be a field. There is a field extension $\overline{F}^{alg}/F$ that is algebraically closed, called the* algebraic closure of F. *$\overline{F}^{alg}$ is unique up to isomorphism.*

Remark 9.2.20. The proof of the existence of algebraic closures is quite technical, and uses Zorn's Lemma (Lemma 7.1.34).

Once we have developed Galois Theory, we will prove the Fundamental Theorem of Algebra.

Theorem 9.2.21 (Fundamental Theorem of Algebra)**.** *$\mathbb{C}$ is algebraically closed.*

Remark 9.2.22. There are no really "pure algebraic" proofs of the Fundamental Theorem of Algebra. The fact that $\mathbb{C}$ is algebraically closed is really a topological or analytic fact, rather than purely algebraic.

Remark 9.2.23. Algebraic closures are usually VERY BIG! For example $\overline{\mathbb{Q}}^{alg}/\mathbb{Q}$ is an infinite extension.

But we can still attach roots one at a time and slowly factor a polynomial.

9.3. **Splitting Fields.**

Definition 9.3.1. Let F be a field, $p(x) \in F[x]$ a non-constant polynomial. For an extension E/F, we say that $p(x)$ *splits in* E if $p(x) \in E[x]$ factors into linear terms. We call an extension E/F a *splitting field for* $p(x)$ if $p(x)$ splits in E and E is the smallest extension of F in which $p(x)$ splits. That is, $p(x)$ does not split in any subfield of E, or equivalently, that E is contained in any extension K/F such that $p(x)$ splits in K.

Example 9.3.2. $x^2 - 2$ splits in $\mathbb{Q}(\sqrt{2})$, namely $x^2 - 2 = (x + \sqrt{2})(x - \sqrt{2})$. Moreover, since the roots of $x^2 - 2$ are exactly $\pm\sqrt{2}$, any field extension of $\mathbb{Q}$ in which $x^2 - 2$ splits must contain $\mathbb{Q}(\sqrt{2})$. Thus $\mathbb{Q}(\sqrt{2})$ is the splitting field of $x^2 - 2$ over $\mathbb{Q}$.

Theorem 9.3.3. *Let $p(x) \in F[x]$ be a non-constant polynomial. There exists a splitting field for $p(x)$.*

Proof. We induct on $\deg(p)$. If $\deg(p) = 1$, then $E = F$ is a splitting field for $p(x)$. We now assume that the statement holds for all polynomials of degree $< n$, and show it holds for $p(x)$ of degree n.

We may further assume that $p(x)$ is irreducible, as we may take a tower of extensions splitting each irreducible factor, which exist by the inductive hypothesis if $p(x)$ factors into lower degree terms.

Since $p(x)$ is irreducible, the ideal $(p) \subseteq F[x]$ is maximal, hence $F[x]/(p)$ is an extension of F in which $p(x)$ has a root α. Thus denoting by $F_1 = F[x]/(p)$, we have

$$p(x) = (x - \alpha)q(x) \in F_1[x],$$

with $\deg(q) < \deg(p)$. Thus by induction, there is an extension E'/F_1 such that $q(x)$ splits in E', and hence $p(x)$ splits in E'.

Let

$$E = \bigcap_{\substack{F \subset K \subset E' \\ p(x) \text{ splits in } K}} K.$$

Then E is a splitting field for $p(x)$. □

Remark 9.3.4. Our proof also shows that we can construct a splitting field for $p(x)$ by successively forming extensions of the form $F[x]/(q(x))$ for irreducible factors $q(x)$ of $p(x)$, thus if E is a splitting field for $p(x)$ over F, then $[E:F] \leq \deg(p)! = \deg(p) \cdot (\deg(p)-1) \cdots 2 \cdot 1$.

In fact, splitting fields are unique (up to isomorphism). We'll need a key lemma first.

Lemma 9.3.5 (Lifting Lemma)**.** *Let $\varphi : F \xrightarrow{\sim} F'$ be an isomorphism of fields. Let $f(x) \in F[x]$ be an irreducible polynomial, and let $f'(x) \in F'[x]$ be the polynomial obtained by applying φ to the coefficients of $f(x)$. In some spitting field, let α be a root of $f(x)$ and let β be a root of $f'(x)$. Then φ extends to an isomorphism $\widetilde{\varphi} : F(\alpha) \to F'(\beta)$ fitting into a commutative diagram*

$$\begin{array}{ccc} F(\alpha) & \xrightarrow{\widetilde{\varphi}} & F'(\beta) \\ \uparrow & & \uparrow \\ F & \xrightarrow[\varphi]{} & F' \end{array}$$

such that $\widetilde{\varphi}|_F = \varphi$. Moreover, $\widetilde{\varphi}$ is unique if we take $\widetilde{\varphi}(\alpha) = \beta$.

Proof. Suppose that $f(x) = a_0 + a_1 x + \cdots + a_n x^n$. we have an isomorphism $\sigma : F[x]/(f) \xrightarrow{\sim} F(\alpha)$. Likewise, we have an isomorphism $\tau : F'[x]/(f') \xrightarrow{\sim} F'(\beta)$. We also have an isomorphism $\varphi' : F[x] \to F'[x]$ by applying φ to the coefficients.

Let $\pi : F'[x] \to F'[x]/(f')$ be the quotient map. Thus $\pi \circ \varphi' : F[x] \to F'[x]/(f')$ is surjective, and $\ker(\pi \circ \varphi') = (f)$, thus the first isomorphism theorem gives an isomorphism

$$F[x]/(f) \xrightarrow{\psi} F'[x]/(f').$$

Thus we have a large commutative diagram

$$\begin{array}{ccc} F[x] & \xrightarrow[\sim]{\varphi'} & F'[x] \\ \downarrow & & \downarrow \pi \\ F[x]/(f) & \xrightarrow[\sim]{\psi} & F'[x]/(f') \\ \sigma \wr \downarrow & & \downarrow \wr\, \tau \\ F(\alpha) & \xrightarrow{\widetilde{\varphi}=\tau\circ\psi\circ\sigma^{-1}} & F'(\beta) \\ \uparrow & & \uparrow \\ F & \xrightarrow[\varphi]{\sim} & F' \end{array}$$

and defining $\widetilde{\varphi} = \tau \circ \psi \circ \sigma^{-1} : F(\alpha) \to F'(\beta)$ gives an isomorphism. We can choose the isomorphisms σ, ψ and τ so that $\widetilde{\varphi}(\alpha) = \beta$, and to get the uniqueness statement, we note that a ring homomorphism $\widetilde{\varphi} : F(\alpha) \to F'(\beta)$ such that $\varphi|_F = \varphi$ is completely determined by $\widetilde{\varphi}(\alpha)$. □

Remark 9.3.6. Lemma 9.3.5 will be crucial in our development of Galois Theory. It is the first glimpse into how symmetries of roots and symmetries of fields interact. It might have seemed tedious to work over an isomorphism $F \to F'$, if we only want to show that splitting fields are isomorphic, but in the near future we'll use this to lift isomorphisms to larger and larger extensions.

Theorem 9.3.7. *Let $\varphi : F \xrightarrow{\sim} F'$ be an isomorphism of fields, $f(x) \in F[x]$ be non-constant and $f'(x) \in F'[x]$ obtained by applying φ to the coefficients of $f(x)$. If E/F is a splitting field for $f(x)$ and E'/F' is a splitting field for $f'(x)$, then there exists an isomorphism $\psi : E \to E'$ extending φ,*

i.e. $\psi|_F = \varphi$, *and the diagram*

$$\begin{array}{ccc} E & \xrightarrow{\psi} & E' \\ \uparrow & & \uparrow \\ F & \xrightarrow[\varphi]{} & F' \end{array}$$

is commutative.

Proof. We induct on the degree of f. As in Lemma 9.3.5, φ extends to an isomorphism $\varphi : F[x] \to F'[x]$, and the irreducible factors of $f(x)$ correspond to the irreducible factors of $f'(x)$. If $f(x)$ splits in F, then f' splits in F', and so $E = F$ and $E' = F'$ and we take $\psi = \varphi$. This proves the case $\deg(f) = 1$ and when all irreducible factors of f are linear.

We now assume the statement holds for all polynomials of degree $< n$, and let $f(x)$ be a polynomial of degree n. Let $p(x)$ be an irreducible factor of $f(x)$ of degree $\deg(p) \geq 2$, and let $p'(x)$ be the corresponding irreducible factor of $f'(x)$. If $\alpha \in E$ is a root of $p(x)$ and $\beta \in E'$ is a root of $p'(x)$, then Lemma 9.3.5 shows we can extend $\varphi : F \; toF'$ to an isomorphism $\sigma : F(\alpha) \to F'(\beta)$ such that $\sigma|_F = \varphi$. Let $F_1 = F(\alpha)$ and $F_1' = F(\beta)$. Now over F_1 and F_1', we have

$$f(x) = (x - \alpha) f_1(x), \;\; f'(x) = (x - \beta) f_1'(x),$$

and as E and E' are splitting fields for f and f', respectively, the fields E and E' are still the splitting fields for f_1 and f_1' over F_1 and F_1', respectively.

Since f_1 and f_1' have degree $< \deg(f)$, we can extend $\sigma : F_1 \to F_1'$ to an isomorphism $\psi : E \to E'$ such that $\psi|_{F_1} = \sigma$ and $\psi|_F = \sigma|_F = \varphi$, giving the commutative diagram

$$\begin{array}{ccc} E & \xrightarrow[\sim]{\psi} & E' \\ \uparrow & & \uparrow \\ F_1 & \xrightarrow[\sim]{\sigma} & F_1' \\ \uparrow & & \uparrow \\ F & \xrightarrow[\sim]{\varphi} & F' \end{array}$$

as desired. □

Corollary 9.3.8. *Any two splitting fields of* $p(x) \in F[x]$ *are isomorphic.*

Proof. Take $F = F'$ and $\varphi = \mathrm{id}_F$ in Theorem 9.3.7. □

9.4. **Automorphisms of fields.**

Definition 9.4.1. Let K be a field. An *automorphism of* K is an isomorphism $\sigma :\xrightarrow{\sim} K$. We denote by $\mathrm{Aut}(K)$ the field automorphisms of K.

Let L/K be a field extension, we define

$$\mathrm{Aut}(L/K) \coloneqq \{\sigma \in \mathrm{Aut}(L) \mid \sigma(k) = k \text{ for all } k \in K\}$$

the subgroup of $\mathrm{Aut}(L)$ that fixes K pointwise.

Exercise 9.4.2. Show that $\mathrm{Aut}(K)$ is a group under composition.

Proposition 9.4.3. *Let* L/K *be an extension, then* $\mathrm{Aut}(L/K)$ *is a subgroup of* $\mathrm{Aut}(L)$.

Proof. Let $\sigma_1, \sigma_2 \in \mathrm{Aut}(L/K)$, and let $k \in K$. Then

$$(\sigma_1 \circ \sigma_2)(k) = \sigma_1(\sigma_2(k)) = \sigma_1(k) = k,$$

and

$$k = \mathrm{id}_L(k) = (\sigma_1^{-1} \circ \sigma_1)(k) = \sigma_1^{-1}(\sigma_1(k)) = \sigma_1^{-1}(k). \qquad \square$$

Example 9.4.4. There is an automorphism of $\mathbb{C}/\mathbb{R}$, namely complex conjugation $x + iy \mapsto x - iy$ that fixes $\mathbb{R}$.

Example 9.4.5. We claimed before that any automorphism of $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ that fixes $\mathbb{Q}$ is either the identity or the map $\tau(a + b\sqrt{2}) = a - b\sqrt{2}$.

Note that any automorphism $\sigma \in \mathrm{Aut}(\mathbb{Q}(\sqrt{2})/\mathbb{Q})$ is determined by where it sends $\sqrt{2}$, by Lemma 9.3.5. but as

$$2 = \sigma(2) = \sigma(\sqrt{2}^2) = \sigma(\sqrt{2})^2,$$

we see that $\sigma(\sqrt{2})$ must be another root of $x^2 - 2 \in \mathbb{Q}[x]$. And indeed the only automorphisms of $\mathbb{Q}(\sqrt{2})$ fixing $\mathbb{Q}$ are either the identity or τ.

Theorem 9.4.6. *Let $\sigma \in \mathrm{Aut}(L/K)$ and $\alpha \in L$ a root of $p(x) \in K[x]$. Then $\sigma(\alpha)$ is a root of $p(x)$.*

Proof. Let $p(x) = a_0 + a_1 x + \cdots a_n x^n \in K[x]$. Since α is a root, we have

$$0 = p(\alpha) = a_0 + a_1\alpha + \cdots + a_n\alpha^n.$$

Hence

$$\begin{aligned} 0 = \sigma(0) = \sigma(p(\alpha)) &= \sigma(a_0 + a_1\alpha + \cdots + a_n\alpha^n) \\ &= \sigma(a_0) + \sigma(a_1\alpha) + \cdots + \sigma(a_n\alpha^n) \\ &= a_0 + a_1\sigma(\alpha) + \cdots + a_n\sigma(\alpha)^n \\ &= p(\sigma(\alpha)), \end{aligned}$$

as claimed. $\square$

Remark 9.4.7. In particular, $\sigma \in \mathrm{Aut}(L/K)$ permutes the roots of polynomials in $K[x]$.

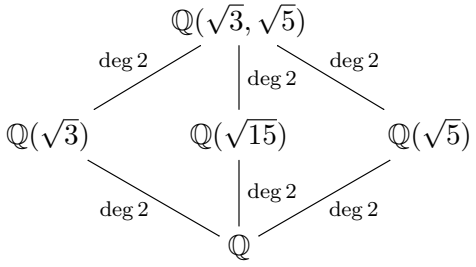
Example 9.4.8. As we've seen, $\mathrm{Aut}(\mathbb{Q}(\sqrt{2})/\mathbb{Q}) = \{\mathrm{id}, \tau\}$ where $\tau(\sqrt{2}) = -\sqrt{2}$, so

$$\mathrm{Aut}(\mathbb{Q}(\sqrt{2})/\mathbb{Q}) \cong \mathbb{Z}/2\mathbb{Z}.$$

Example 9.4.9. We'll consider the extension $\mathbb{Q}(\sqrt{3}, \sqrt{5})/\mathbb{Q}$. On homework, or else an an exercise, you've seen that

$$\mathbb{Q}(\sqrt{3}, \sqrt{5}) = \{a + b\sqrt{3} + c\sqrt{5} + d\sqrt{15} \mid a, b, c, d \in \mathbb{Q}\}.$$

We have a few subfields of $\mathbb{Q}(\sqrt{3}, \sqrt{5})$, namely



where by the Tower Law, we have $[\mathbb{Q}(\sqrt{3}, \sqrt{5}) : \mathbb{Q}] = 4$.

We note that every field extension in sight is a splitting field. Namely $\mathbb{Q}(\sqrt{3}, \sqrt{5})$ is the splitting field of $(x^2 - 3)(x^2 - 5)$ over $\mathbb{Q}$, $\mathbb{Q}(\sqrt{3}, \sqrt{5})$ is the splitting field of $(x^2 - 3)$ over $\mathbb{Q}(\sqrt{5})$ since $\sqrt{3} \notin \mathbb{Q}(\sqrt{5})$, and $\mathbb{Q}(\sqrt{2})$ is the splitting field of $x^2 - 2$ over $\mathbb{Q}$, etc.

If we consider the the extension $\mathbb{Q} \subset \mathbb{Q}(\sqrt{5}) \subset \mathbb{Q}(\sqrt{3}, \sqrt{5})$, we can use Lemma 9.3.5 to lift the automorphism $\mathrm{id}_{\mathbb{Q}}$ to an automorphism

$$\tau : \mathbb{Q}(\sqrt{5}) \to \mathbb{Q}(\sqrt{5}),\ \tau(a + b\sqrt{5} = a - b\sqrt{5}),$$

which we can further extend τ to an automorphism of the splitting field $\mathbb{Q}(\sqrt{3}, \sqrt{5})$ using Lemma 9.3.5 again. In doing this, we have a few choices, as we can send the root $\sqrt{3}$ to itself or to $-\sqrt{3}$, we'll call the first extension still τ and the latter $\sigma\tau$. Thus we find two automorphisms

$$\begin{aligned} \tau, \sigma\tau &\in \mathrm{Aut}(\mathbb{Q}(\sqrt{3}, \sqrt{5})/\mathbb{Q}), \\ \tau(a + b\sqrt{3} + c\sqrt{5} + d\sqrt{15}) &= a + b\sqrt{3} - c\sqrt{5} - d\sqrt{15}, \\ \sigma\tau(a + b\sqrt{3} + c\sqrt{5} + d\sqrt{15}) &= a - b\sqrt{3} - c\sqrt{5} + d\sqrt{15}. \end{aligned}$$

We note that

$$(\sigma\tau \circ \tau)(a + b\sqrt{3} + c\sqrt{5} + d\sqrt{15}) = a - b\sqrt{3} + c\sqrt{5} - d\sqrt{15},$$

and we define

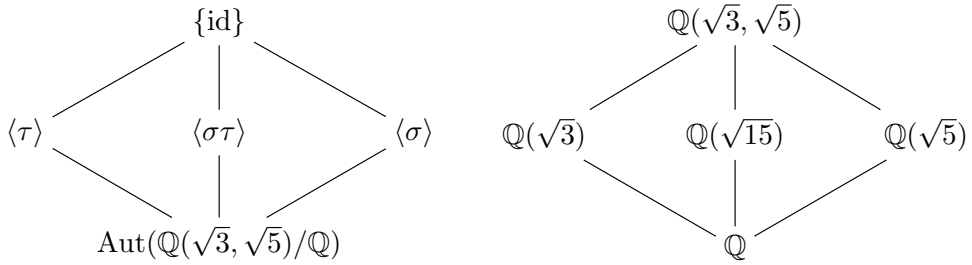
$$\sigma \in \mathrm{Aut}\, \mathbb{Q}(\sqrt{3}, \sqrt{5})/\mathbb{Q},\ \sigma(a + b\sqrt{3} + c\sqrt{5} + d\sqrt{15}) = a - b\sqrt{3} + c\sqrt{5} - d\sqrt{15}.$$

We note that $\sigma^2 = \mathrm{id}$, $\tau^2 = \mathrm{id}$, and $\sigma\tau = \tau\sigma$.

Lifting automorphisms via the other degree 2 extensions of $\mathbb{Q}$, we arrive at the same automorphisms. As any automorphism of $\mathbb{Q}(\sqrt{3}, \sqrt{5})$ permutes the roots of $x^2 - 3$ and $x^2 - 5$, and is completely determined by where those roots go (as they generate $\mathbb{Q}(\sqrt{3}, \sqrt{5})$), we see that in fact

$$\mathrm{Aut}(\mathbb{Q}(\sqrt{3}, \sqrt{5})/\mathbb{Q}) = \{\mathrm{id}, \tau, \sigma, \sigma\tau\} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} = \langle\sigma\rangle \times \langle\tau\rangle.$$

We now observe that σ is the identify on the subfield $\mathbb{Q}(\sqrt{5})$, τ is the identity on the subfield $\mathbb{Q}(\sqrt{3})$, and $\sigma\tau$ is the identity on the subfield $\mathbb{Q}(\sqrt{15})$. Furthermore, these are all the generators of subgroups of $\mathrm{Aut}(\mathbb{Q}(\sqrt{3}, \sqrt{5})/\mathbb{Q})$, and we have two diagrams



where the left shows subgroups of $\mathrm{Aut}(\mathbb{Q}(\sqrt{3}, \sqrt{5})/\mathbb{Q})$ and the right shows subfields of $\mathbb{Q}(\sqrt{3}, \sqrt{5})$ where that subgroup acts trivially (the subfields fixed by that subgroup).

Remark 9.4.10. This is Galois Theory, for an "nice" extension L/K, a relationship

$$\{\text{subgroups of } \mathrm{Aut}(L/K)\} \longleftrightarrow \{\text{intermediate field extensions } K \subset M \subset L\},$$

and a "dictionary" between properties of subgroups and properties of the intermediate extensions.

Proposition 9.4.11. *Let $H \subseteq \mathrm{Aut}(K)$ be a subset. Let*

$$K^H \coloneqq \{k \in K \mid h(k) = k \text{ for all } h \in H\}.$$

Then K^H is a subfield of K.

Proof. Let $h \in H$, $h(0) = 0$ and $h(1) = 1$ as h is a ring homomorphism, thus $0, 1 \in K^H$. Now let $a, b \in K^H$. As h is a ring homomorphism, we have $h(a \pm b) = h(a) \pm h(b) = a \pm b$, $h(ab) = h(a) = h(b) = ab$, and for $a \neq 0$ we have $h(a^{-1}) = h(a)^{-1} = a^{-1}$, thus $a \pm b$, ab and a^{-1} for $a \neq 0$ are all in K^H. Hence K^H is a subfield of K. □

Definition 9.4.12. For a subgroup $H \leq \mathrm{Aut}(K)$, we call K^H the *fixed field of* H.

So we have an association

$$\begin{aligned} \{\text{subgroups of } \mathrm{Aut}(K)\} &\longleftarrow \{\text{subfields } L \subset K\} \\ H \leq \mathrm{Aut}(K) &\longmapsto K^H \\ \mathrm{Aut}(K/L) \leq \mathrm{Aut}(K) &\longmapsfrom L \subset K. \end{aligned}$$

Theorem 9.4.13. *This association is inclusion reversing. That is*

(1) If $L_1 \subset L_2 \subset K$, then $\mathrm{Aut}(K/L_2) \leq \mathrm{Aut}(K/L_1)$, and
(2) if $H_1 \leq H_2$, then $K^{H_2} \subseteq K^{H_1}$.

Proof. To show (1), let $\sigma \in \mathrm{Aut}(K/L_2)$. Then since $L_1 \subset L_2$, and σ fixes L_2, σ fixes L_1 as well. Hence $\sigma \in \mathrm{Aut}(K/L_1)$, and so every element of $\mathrm{Aut}(K/L_1)$ is in $\mathrm{Aut}(K/L_2)$, thus $\mathrm{Aut}(K/L_1) \subseteq \mathrm{Aut}(K/L_2)$.

To show (2), let $k \in K^{H_2}$. For $h \in H_1$, as $h \in H_1 \subset H_2$, $h(k) = k$, hence k is fixed by every element of H_1, and so $K^{H_2} \subseteq K^{H_1}$. □

Example 9.4.14. While there are not too many subgroups, it may be interesting to see this in action in the familiar Example 9.4.9.

Remark 9.4.15 (Important Remark)**.** If L/K is a finite extension, then any $\sigma \in \mathrm{Aut}(L/K)$ is completely determined by where it sends the finite basis of L as a K-vector space. Since L/K is finite, each basis element is algebraic over K and thus has a minimal polynomial, and σ can only send a root of the minimal polynomial to another root, of which there are finitely many. Thus there are only finitely many options for σ. Thus, for L/K a finite extension, $\mathrm{Aut}(L/K)$ is a finite group.

However, not every permutation of roots necessarily gives an automorphism.

Example 9.4.16. Consider the extension $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$. As this is a simple extension, any automorphism $\sigma \in \mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$ is determined by $\sigma(\sqrt[3]{2})$, which must be a root of $x^3 - 2 \in \mathbb{Q}[x]$. The roots are

$$\sqrt[3]{2},\ e^{2\pi i/3}\sqrt[3]{2},\ e^{4\pi i/3}\sqrt[3]{2},$$

though only $\sqrt[3]{2} \in \mathbb{Q}(\sqrt[3]{2})$ as the rest are complex and every element of $\mathbb{Q}(\sqrt[3]{2})$ is real.

Thus $\sigma \in \mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$ must send $\sqrt[3]{2}$ to $\sqrt[3]{2}$, and hence $\sigma \mathrm{id}_{\mathbb{Q}(\sqrt[3]{2})}$. So

$$\mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}) = \{\mathrm{id}\}.$$

In general, we'll show that $|\mathrm{Aut}(L/K)| \leq [L : K]$, and equality holds for "nice" extensions. We'll start with finite simple extensions first, and then extend to any finite extension using a tower of finite simple extensions.

Proposition 9.4.17. *Let $K(\alpha)/K$ be a finite simple extension. Then $|\mathrm{Aut}(L/K)| \leq [L : K]$.*

Proof. Let $p_\alpha \in K[x]$ be the minimal polynomial of α, and so $[K(\alpha) : K] = \deg(p_\alpha)$. Since

$$K[x]/(p_\alpha) \cong K(\alpha) = \{k_0 + k_1\alpha + \cdots + k_{d-1}\alpha^{d-1} \mid k_i \in K,\ d = \deg(p_\alpha)\},$$

we see from Lemma 9.3.5 and Remark 9.4.15 that any $\sigma \in \mathrm{Aut}(K(\alpha)/K)$ is determined completely by $\sigma(\alpha)$, which must be another root of p_α by Theorem 9.4.6. Thus

$$|\mathrm{Aut}(K(\alpha)/K)| \leq \#\text{ root of } p_\alpha \text{ in } K(\alpha) \leq \deg(p_\alpha) = [K(\alpha) : K].$$ □

Lemma 9.4.18. *Let E/F be a finite extension and $\sigma : E \to E$ a ring homomorphism fixing F. Then $\sigma \in \mathrm{Aut}(E/F)$.*

Proof. We just need to show that σ is a bijection, as then σ^{-1} is also a ring homomorphism fixing F.

Since $\ker\sigma \subseteq E$ is an idea, and E is a field, $\ker\sigma = 0$, hence σ is injective.

To show that σ is surjective, note that σ is F-linear, hence is a map $\sigma : E \to E$ of F-vector spaces of the same finite dimension. Hence as σ is injective, it is also surjective. □

We require a fact, which is part of Theorem 9.2.11.

Proposition 9.4.19. *Let E/F be a finite extension. Then there exist intermediate extensions*

$$F = F_0 \subset F_1 \subset \cdots F_{n-1} \subset F_n = E$$

with F_{i+1}/F_i simple.

Proof. See Theorem 9.2.11. We also give a new proof for ease of readability.

Let $\alpha \in E$ with $\alpha \notin F$. Then $F_1 = F(\alpha)/F$ is a simple extension and

$$[E : F] = [E : F_1]\underbrace{[F_1 : F]}_{>1},$$

hence $[E : F_1] < [E : F]$. If $E = F_1$, we are done. Else by induction on $[E : F]$, there exist intermediate extensions

$$F_1 \subset \cdots F_{n-1} \subset F_n = E$$

such that F_{i+1}/F_i is simple, and adding $F = F_0 \subset F_1$ gives the claim. □

The idea is now to lift automorphisms for simple extensions, and then lift in a tower of simple extensions. We show that for simple extensions, the ways we can lift an automorphism corresponds to the roots of the minimal polynomial of the generating element.

Proposition 9.4.20. *Let $F(\alpha)/F$ be a simple extension and let $p_\alpha(x) \in F[x]$ be the minimal polynomial of α, then*

$$|\operatorname{Aut}(F(\alpha)/F)| = \#\text{ roots of } p_\alpha(x) \text{ in } F(\alpha).$$

Proof. We saw $|\operatorname{Aut}(F(\alpha)/F)| \leq \#$ roots of $p_\alpha(x)$ in $F(\alpha)$ in Proposition 9.4.17. For the other inequality, let β be a root of $p_\alpha(x)$ in $F(\alpha)$. Since $F(\beta) \subseteq F(\alpha)$ and they have the same degree over F (namely $\deg(p_\alpha)$), we have $F(\alpha) = F(\beta)$. Thus by Lemma 9.3.5, we can lift $\mathrm{id}_F : F \to F$ uniquely to an isomorphism $F(\alpha) \to F(\beta) = F(\alpha)$. □

Remark 9.4.21. The proof of Proposition 9.4.20 also shows that

$$|\operatorname{Aut}(F(\alpha)/F)| = \#\text{ ways to extend } \mathrm{id}_F : F \to F \text{ to } \sigma : F(\alpha) \to F(\alpha).$$

Theorem 9.4.22. *Let E/F be finite, then*

$$|\operatorname{Aut}(E/F)| \leq [E : F].$$

Proof. Since E/F is finite, byProposition 9.4.19, we have a tower of simple extensions

$$F = F_0 \subset F_1 \subset \cdots F_{n-1} \subset F_n = E.$$

We see also that

$$|\operatorname{Aut}(F(\alpha)/F)| = \#\sigma_n : F_n \to F_n \text{ extending } \mathrm{id}_F : F \to F,$$

and going one simple extension at a time we have

$$\begin{aligned}
\#\sigma_n \text{ extending } \mathrm{id}_F &= (\#\sigma_n \text{ extending } \sigma_{n-1} : F_{n-1} \to F_{n-1}) \cdot (\#\sigma_{n-1} \text{ extending } \mathrm{id}_F) \\
&= (\#\sigma_n \text{ extending } \sigma_{n-1}) \cdot (\#\sigma_{n-1} \text{ extending } \sigma_{n-2}) \cdot (\#\sigma_{n-2} \text{ extending } \mathrm{id}_F) \\
&\vdots \\
&= \underbrace{(\#\sigma_n \text{ extending } \sigma_{n-1})}_{\leq [E:F_{n-1}]} \cdot \underbrace{(\#\sigma_n \text{ extending } \sigma_{n-1})}_{\leq [F_{n-1}:F_{n-2}]} \cdots \underbrace{(\#\sigma_1 \text{ extending } \mathrm{id}_F)}_{\leq [F_1:F]},
\end{aligned}$$

where the inequalities come from Proposition 9.4.17. Hence

$$|\operatorname{Aut}(E/F)| \leq [E : F_{n-1}][F_{n-1} : F_{n-2}] \cdots [F_1 : F] = [E : F]. \qquad \square$$

Example 9.4.23. We've seen the two examples

$$|\operatorname{Aut}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}) = \{1\}| = 1 \leq [\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3,$$

and

$$|\operatorname{Aut}(\mathbb{Q}(\sqrt{2})/\mathbb{Q}) \cong \mathbb{Z}/2\mathbb{Z}| = 2 = [\mathbb{Q}(\sqrt{2}) : \mathbb{Q}].$$

Definition 9.4.24. A finite extension E/F is *Galois* if $|\operatorname{Aut}(E/F)| = [E : F]$.

Example 9.4.25. The extensions $\mathbb{Q}(\sqrt{3}, \sqrt{5})/\mathbb{Q}$, $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ are both Galois, while $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is not. How exactly is $\mathbb{Q}(\sqrt[3]{2})$ different? Well, not every root of $x^3 - 2$ is in $\mathbb{Q}(\sqrt[3]{2})$. And we've seen that the way to get automorphisms is by permuting roots. Hence not having all of the roots prevents an extension from having certain automorphisms.

It turns out that there is another complication that can occur.

Example 9.4.26. Let $\mathbb{F}_2 \coloneqq \mathbb{Z}/2\mathbb{Z}$, and let $F = \mathbb{F}_2(t)$ be the field of rational functions in t with coefficients in $\mathbb{F}_2$. Consider $x^2 - t \in F[x]$, which is irreducible as $\sqrt{t} \notin F$. But, since $1 = -1 \in F$, $x^2 - t$ has a double root $\sqrt{t}$. Thus $x^2 - t$ only has one root!

Thankfully, these are the only complications preventing an extension from being Galois, as we'll soon see. First, we take a detour into understanding when an extension has all the roots we expect.

9.5. **Separable extensions.**

Definition 9.5.1. Let F be a field and $f(x) \in F[x]$ be a non-constant polynomial. In a splitting field of $f(x)$, we can write

$$f(x) = a \prod_{i=1}^{m} (x - \alpha_i)_i^e,$$

where $a \in F$, α_i are the *distinct* roots of $f(x)$, and $e_i \in \mathbb{Z}_{\geq 1}$. We call the exponents e_i the *[hdegree of the root α_i, and we say that a root α_i is a* multiple root *if $e_i \geq 2$.*

Definition 9.5.2. Let F be a field and $f(x) \in F[x]$. We say that $f(x)$ is *separable* if $f(x)$ has no multiple roots. If $f(x)$ has a multiple root, we say $f(x)$ is *inseparable*

Let us explore some ways to check whether a polynomial is separable.

Definition 9.5.3. Let $f(x) = a_n x^n + \cdots + a_1 x + a_0 \in F[x]$. The *formal derivative of $f(x)$* is

$$D_x f(x) \coloneqq a_n n x^{n-1} + a_{n-1}(n-1)x^{n-2} + \cdots a_2 2x + a_1 \in F[x].$$

The usual rules of derivatives apply, namely for $f, g \in F[x]$, we have

$$D_x(f + g) = D_x f + D_x g,$$

and

$$D_x(fg) = f(D_x g) + (D_x f)g.$$

Proposition 9.5.4. *Let F be a field, and $f(x) \in F[x]$ with splitting field E/F. Then $\alpha \in E$ is a multiple root of $f(x)$ if and only if α is also a root of $D_x f$. Let $p_\alpha(x) \in F[x]$ be the minimal polynomial of α. Then α is a multiple root of $f(x)$ if and only if*

$$p_\alpha \mid f \text{ and } p_\alpha \mid D_X f.$$

Remark 9.5.5. In particular, $f(x)$ is separable if and only if $\gcd(f, D_x f) = 1$, which is the contrapositive of Proposition 9.5.4.

Proof. If α is a multiple root of $f(x)$, then over a splitting field,

$$f(x) = (x-\alpha)^n g(x),\ n \geq 2,$$

thus

$$D_x f = n(x-\alpha)^{n-1} g(x) + (x-\alpha)^n D_x g.$$

Thus α is a root of $D_x f$.

Conversely, suppose α is a root of f and $D_x f$. Then in a splitting field of f and $D_x f$, we have

$$f(x) = (x-\alpha)h(x),$$

and

$$D_x f = h(x) + (x-\alpha)D_x h.$$

Thus

$$0 = D_X f(\alpha) = h(\alpha),$$

and so $h(x) = (x-\alpha)h_1(x)$, and hence $f(x) = (x-\alpha)^2 h_1(x)$, and α is a multiple root of $f(x)$. To get the statement with p_α, recall that α is a root of $f(x)$ if and only if $p_\alpha \mid f$. □

Proposition 9.5.6. *If E/F is the splitting field of a separable polynomial $f(x) \in F[x]$, then*

$$|\operatorname{Aut}(E/F)| = [E:F],$$

i.e. E/F is Galois.

Proof. We have seen that $|\operatorname{Aut}(E/F)| \leq [E:F]$ in Theorem 9.4.22. Hence it remains to show that $|\operatorname{Aut}(E/F)| \geq [E:F]$.

Recall that by Theorem 9.3.7, automorphisms of E/F can be obtained by extending id_F. Thus we show that id_F can be extended to an automorphism of E in at least $[E:F]$ ways.

In particular, Theorem 9.3.7 states that an isomorphism $\varphi : F \to F'$ can be extended to an isomorphism $\psi : E \to E'$ of splitting fields E of f and E' of $f'(x) = \varphi(f(x)) \in F'[x]$. We show that for a separable polynomial $f(x)$ the number of ways of extending φ to ψ is exactly $[E:F]$.

We proceed by induction on $[E:F]$. If $f(x)$ factors into linear factors over F, then $E = F$, and $|\operatorname{Aut}(E/F)| = [E:F] = 1$, and we are done. We now assume that the statement holds for all extensions of degree $< [E:F]$.

Let $p \in F[x]$ be an irreducible factor of $f(x)$ of degree $d = \deg(p) \geq 2$, with corresponding irreducible factor p' of f'. Let α be a root of $p(x)$. If ψ is any extension of φ to E, then $\psi|_{F(\alpha)}$ is an isomorphism of $F(\alpha)$ with another subfield $F'(\beta) \subseteq E'$ with β a root of $p'(x)$. This gives a diagram

$$\begin{array}{ccc} E & \xrightarrow{\psi} & E' \\ \uparrow & & \uparrow \\ F(\alpha) & \xrightarrow{\widetilde{\varphi}} & F'(\beta) \\ \uparrow & & \uparrow \\ F & \xrightarrow{\varphi} & F' \end{array}$$

Conversely, by Lemma 9.3.5, for any root β of $p'(x)$, there is an extension of φ to $\widetilde{\varphi} : F(\alpha) \to F'(\beta)$ and by Theorem 9.3.7, there is an extension ψ of $\widetilde{\varphi}$.

The number of extensions of φ to $\widetilde{\varphi}$ is equal to the number of distinct roots β of $p'(x)$. Since f is separable, p and p' are separable, thus there are exactly $[F(\alpha):F]$ distinct extensions $\widetilde{\varphi}$ of φ.

Since E is the splitting field of f over $F(\alpha)$, and E' is the splitting field of f' over $F'(\beta)$, and $[E:F(\alpha)] < [E:F]$, we can apply the induction hypothesis to these field extensions and see that the number of extensions of $\widetilde{\varphi}$ to ψ is exactly $[E:F(\alpha)]$. Hence as $[E:F(\alpha)][F(\alpha):F] = [E:F]$, there are exactly $[E:F]$ ways to extend $\varphi : F \to F'$ to $\psi : E \to E'$, as desired. □

Definition 9.5.7. An extension E/F is *separable* if every $\alpha \in E$ is the root of a separable polynomial over F. Equivalently, for all $\alpha \in E$, the minimal polynomial of α over F is separable.

Definition 9.5.8. Let F be a field. The smallest $p \in \mathbb{Z}_{>0}$ such that $p \cdot 1 = 0$ is called the *characteristic of* F. If no such p exists, we say F has *characteristic* 0. We write $\operatorname{char} F = p$ when F has characteristic p.

Exercise 9.5.9. Let F be a field.

(1) Show that there is a unique ring homomorphism $\mathbb{Z} \to F$.
(2) What can the kernel be? This number is called the *characteristic of* F. Hint: A field is a domain, how does that restrict what the kernel can be?

    The subfield generated by the image of the unique map $\mathbb{Z} \to F$ is called the *prime subfield of* F, and it is the smallest subfield containing 1.
(3) Show that any field homomorphism $\sigma : F_1 \to F_2$ induces an isomorphism on the prime subfields of F_1 and F_2.
(4) What is the prime subfield of $\mathbb{R}$?
(5) Show that if F is a **finite** field, then it is a finite dimensional vector space over its prime subfield ($\mathbb{Z}_p$ for some prime p, which is the characteristic of F). Argue that therefore $|F| = p^n$ for some n.

Example 9.5.10. The characteristic of $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ is 0. The characteristic of $\mathbb{Z}/p\mathbb{Z}$ is p, as is the characteristic of $\mathbb{Z}/p\mathbb{Z}(t)$.

Proposition 9.5.11. *If* $\operatorname{char} F = 0$ *and* $f(x)$ *is irreducible, then* $f(x)$ *is separable.*

Proof. Let $f(x) \in F[x]$ be irreducible of degree n. Since $\operatorname{char} F = 0$, $D_x f$ has degree $n-1$. Thus

$$\deg(\gcd(f, D_x f)) \leq n - 1.$$

By definition, $\gcd(f, D_x f)g = f$ for some $g \in F[x]$. But as f is irreducible, either $\gcd(F, D_x f) \in F^\times$ or $g \in F^\times$. If $g \in F^\times$, then

$$\deg(\gcd(f, D_x f)) = \deg f = n > n - 1,$$

which is a contradiction. Thus $\gcd(f, D_x f) = 1$, and by Proposition 9.5.4, f is separable. □

Remark 9.5.12. This proof does not work when $\operatorname{char} F = p > 0$. Indeed, $D_x(x^p - 1) = px^{p-1} = 0$.

Nevertheless, the same fact is true for a finite field. We'll need a few lemmas first.

Lemma 9.5.13 (Freshperson's Dream)**.** *Let* F *be a ring of characteristic* $p > 0$*. Then*

$$(a + b)^p = a^p + b^p.$$

Proof. We use the binomial theorem, noting that for $1 \leq k \leq p-1$, $\binom{p}{k} = \frac{p!}{k!(p-k)!}$ is divisible by p, as the only primes appearing int he denominator are smaller than p.

Now expand

$$(a+b)^p = \sum_{k=1}^{p} \binom{p}{k} a^k b^{p-k} = a^p + b^p + \underbrace{\sum_{k=1}^{p} -1\binom{p}{k} a^k b^{p-k}}_{=0} = a^p + b^p.$$ □

Lemma 9.5.14. *Let* $\mathbb{F}$ *be a field of characteristic* $p > 0$*. The map* $\operatorname{Frob} : \mathbb{F} \to \mathbb{F}$, $a \mapsto a^p$ *is a field isomorphism.*

Proof. We first check that Frob is a ring homomorphism. Clearly $\text{Frob}(0) = 0$ and $\text{Frob}(1) = 1$. By Lemma 9.5.13, for $a, b \in \mathbb{F}$, we have

$$\text{Frob}(a+b) = (a+b)^p = a^p + b^p = \text{Frob}(a) + \text{Frob}(b),$$

and clearly $\text{Frob}(ab) = (ab)^p = a^p b^p = \text{Frob}(a)\,\text{Frob}(b)$. Since Frob is a homomorphism of fields, Frob is injective. As $|\mathbb{F}|$ is finite, Frob is surjective as well. □

Corollary 9.5.15. *Let $\mathbb{F}$ be a finite field of characteristic $p > 0$, and let $a \in \mathbb{F}$. Then there is an element $b \in \mathbb{F}$ such that $a = b^p$.*

Proof. Take $b = \text{Frob}^{-1}(a)$. □

Theorem 9.5.16. *Let $\mathbb{F}$ be a finite field of characteristic $p > 0$. If $f(x) \in \mathbb{F}[x]$ is irreducible, then $f(x)$ is separable.*

Proof. Suppose for contradiction that $f(x)$ is irreducible but is not separable. Then by Proposition 9.5.4, $\gcd(f, D_x f) \neq 1$. But since $f(x)$ is irreducible, it must be the case that $D_x f = 0$. Thus

$$f(x) = a_n x^{np} + a_{n-1} x^{(n-1)p} + \cdots + a_1 x^p + a_0,$$

hence $f(x) = f_1(x^p)$ with

$$f_1(x) = a_n x^n + \cdots a_1 x + a_0.$$

By Corollary 9.5.15, there exist $b_i \in \mathbb{F}$ such that $a_i = b_i^p$, and so we have

$$\begin{aligned} f(x) = f_1(x^p) &= a_n x^{np} + a_{n-1} x^{(n-1)p} + \cdots + a_1 x^p + a_0 \\ &= b_n^p (x^n)^p + \cdots + b_1^p (x)^p + b_0^p \\ &= (b_n x^n)^p + \cdots + (b_1 x)^p + (b_0)^p \\ &= (b_n x^n + \cdots + b_1 x + b_0)^p, \end{aligned}$$

where the last equality comes fromLemma 9.5.13. But this is a contradiction, as $f(x)$ is irreducible. Thus $f(x)$ must be separable, as desired. □

9.6. **Cylotomic Extensions.** Let $\mu_n \in \mathbb{C}$ be an n^{th} roots of unity, i.e. $\mu_n^n = 1$, so μ_n is a root of $x^n - 1$.

Example 9.6.1. the third roots of unity are $1, e^{2\pi i/3}, e^{4\pi i/3}$.

Exercise 9.6.2. Let p be prime and μ_p a primitive p^{th} root of unity, i.e. $\text{ord}(\mu_p) = p$. Then the minimal polynomial of μ_p over $\mathbb{Q}$ is

$$\Phi_p := x^{p-1} + x^{p-2} + \cdots + x + 1.$$

Corollary 9.6.3. $[\mathbb{Q}(\mu_p)/\mathbb{Q}] = p - 1$.

More generally, for a primitive n^{th} root of unity, the minimal polynomial over $\mathbb{Q}$ is

$$\Phi_n = \prod_{\substack{1 \leq k \leq n-1 \\ \gcd(k,n)=1}} (x - e^{2\pi i k/n}).$$

And we have a relation

$$\prod_{d|n} \Phi_d = x^n - 1.$$

Definition 9.6.4. An extension of the form $F(\mu_n)/F$ where n is a primitive n^{th} root of unity is called a *cyclotomic extension*. The polynomials Φ_n are called *cyclotomic polynomials*.

Proposition 9.6.5. *Let p be prime and μ_p a primitive p^{th} root of unity. Then*

$$\mathrm{Aut}(\mathbb{Q}(\mu_p)/\mathbb{Q}) \cong \mathbb{Z}/p\mathbb{Z}^\times,$$

in particular $\mathbb{Q}(\mu_p)/\mathbb{Q}$ is a Galois extension.

Proof. For $a \in \mathbb{Z}/p\mathbb{Z}^\times$, the map

$$\mathbb{Q}(\mu_p) \cong \mathbb{Q}[x]/(\Phi_p) \to \mathbb{Q}[x]/(\Phi_p),\ \overline{x} \mapsto \overline{x}^a$$

is an automorphism, since $(\mu_p)^a$ is another primitive p^{th} root of unity.

This defines a group homomorphism

$$\mathbb{Z}/p\mathbb{Z}^\times \to \mathrm{Aut}(\mathbb{Q}(\mu_p)/\mathbb{Q}), a \mapsto (\bullet \mapsto (\bullet)^a),$$

which is clearly injective. By Theorem 9.4.22, $|\mathrm{Aut}(\mathbb{Q}(\mu_p)/\mathbb{Q})| \leq p-1 = [\mathbb{Q}(\mu_p) : \mathbb{Q}]$. Thus as $|\mathbb{Z}/p\mathbb{Z}^\times| = p-1$, this group homomorphism is also surjective, whereby $\mathrm{Aut}(\mathbb{Q}(\mu_p)/\mathbb{Q}) \cong \mathbb{Z}/p\mathbb{Z}^\times$ and $\mathbb{Q}(\mu_p)/\mathbb{Q}$ is Galois. □

9.7. **Galois Correspondence.** We now earnestly move toward proving the Fundamental Theorem of Galois Theory, spelling out the dictionary between subgroups of automorphisms and subextensions.

Theorem 9.7.1 (Artin)**.** *Let E be a field, and $G \leq \mathrm{Aut}(E)$ a finite subgroup. Then $[E : E^G] \leq |G|$.*

Proof. Let $F = E^G$, and let $G = \{\sigma_1 = \mathrm{id}, \sigma_2, \dots, \sigma_m\}$. If suffices to show that if $\{\alpha_1, \dots, \alpha_n\}$ with $n > m$, then the $\alpha_1 \dots, \alpha_n$ are linearly dependent over F. Consider the linear map

$$T : E^n \to E^m$$

given by

$$\begin{pmatrix} \sigma_1(\alpha_1) \cdots \sigma_1(\alpha_n) \\ \vdots \\ \sigma_m(\alpha_1) \cdots \sigma_m(\alpha_n) \end{pmatrix}.$$

Since $n > m$, $\ker T \neq 0$, hence there is some non-zero vector $(c_1, \dots, c_n) \in \ker T$, say with the fewest non-zero elements. After reordering the α_i, we may assume that $0 \neq c_1 \in E$. Since T is linear, $\frac{1}{c_1}(c_1, \dots, c_n) = (1, d_2, \dots, d_n) \in \ker T$, thus we may also assume that $c_1 \in F$.

Now

$$\sigma_1(c_1)\alpha_1 + \cdots + \sigma_1(c_n)\alpha_n = c_1\alpha_1 + \cdots + c_n\alpha_n = 0$$

is a linear relation among the α_i. Thus if all the $c_i \in F$, we are done.

Else, suppose that there is some $c_i \notin F$, so there is some $\sigma_k \in G$ such that $\sigma_k(c_i) \neq c_i$. Precomposing T with σ_k in each factor clearly preserves the kernel, as σ_k simply permutes the elements of G which acts on the matrix

$$\begin{pmatrix} \sigma_1(\alpha_1) \cdots \sigma_1(\alpha_n) \\ \vdots \\ \sigma_m(\alpha_1) \cdots \sigma_m(\alpha_n) \end{pmatrix}$$

by permuting the rows, which does not change the kernel. Hence if $(c_1, \dots, c_n) \in \ker T$, then $(\sigma_k(c_1), \dots, \sigma_k(c_n)) \in \ker T$. As we may assume $c_1 \in F$, and we have

$$(c_1, \sigma_k(c_2), \dots, \sigma_k(c_n)) \in \ker T,$$

hence

$$(c_1, c_2 \dots, c_n) - (c_1, \sigma_k(c_2), \dots, \sigma_k(c_n)) = (0, c_2 - \sigma_k(c_2), \dots, c_n - \sigma_k(c_n)) \in \ker T.$$

But this has at least one more zero entry that $(c_1 \dots, c_n)$, which is a contradiction.

Thus all of the $c_i \in F$, and $\{\alpha_1, \dots, \alpha_n\}$ are linearly dependent over F, as desired. □

Corollary 9.7.2. *Let $G \leq \mathrm{Aut}(E)$ be a finite group. Then $G = \mathrm{Aut}(E/E^G)$.*

Proof. We clearly have $G \leq \mathrm{Aut}(E/E^G)$, thus we have the inequalities

$$[E : E^G] \overset{\text{Theorem 9.7.1}}{\leq} |G| \overset{G \leq \mathrm{Aut}(E/E^G)}{\leq} |\mathrm{Aut}(E/E^G)| \overset{\text{Theorem 9.4.22}}{\leq} [E : E^G]. \qquad \square$$

Definition 9.7.3. An extension E/F is *normal* if it is algebraic and for every $\alpha \in E$ the minimal polynomial $p_\alpha(x) \in F[x]$ splits in $E[x]$.

Example 9.7.4. The extension $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ is normal.

Remark 9.7.5. For an extension E/F and $\alpha \in E$, we have the implications

$$\begin{aligned} E/F \text{ normal} &\implies p_\alpha(x) \text{ splits} \\ E/F \text{ separable} &\implies p_\alpha(x) \text{ has distinct roots.} \end{aligned}$$

Thus if E/F is normal and separable, then $p_\alpha(x)$ has exactly $\deg(p_\alpha)$ distinct roots in E.

Remark 9.7.6. An extension E/F is normal and separable if and only if every irreducible polynomial $f(x) \in F[x]$ with a root in E splits completely and has $\deg(f)$ distinct roots in E.

Theorem 9.7.7. *A finite extension E/F is Galois if and only if it is normal and separable.*

Proof. Suppose that E/F is normal and separable. Since E/F is finite, by Proposition 9.4.19, there is a tower

$$F = F_0 \subset F_1 \subset \cdots \subset F_n = E$$

of simple normal separable finite extensions, and lifting automorphisms at each step as in the Proposition 9.5.6 using Lemma 9.3.5 shows that at each stage, since E/F is normal, we have exactly $[F_{i+1} : F_i]$ extensions since we have that many roots of the corresponding simple generator. Hence by the Tower Law, $|\mathrm{Aut}(E/F)| = [E : F]$.

Conversely, suppose that E/F is Galois, and let $G = \mathrm{Aut}(E/F)$. Since $|G| = [E : F]$, $|G|$ is finite. By definition, we have $F \subset E^G$, thus as $G = \mathrm{Aut}(E/E^G)$ by Corollary 9.7.2, we have $G = \mathrm{Aut}(E/E^G) \leq \mathrm{Aut}(E/F) = G$, whereby $F = E^G$ by Theorem 9.4.13. Now let $\alpha \in E$ with minimal polynomial $p_\alpha \in F[x]$. Let $\{\alpha_1 = \alpha, \alpha_2, \ldots, \alpha_m\} \subset E$ be the distinct elements of the G-orbit of α. Let

$$g(x) = \prod_{i=1}^{m} (x - \alpha_i) = x^m + a_1 x^{m-1} + \cdots + a_m \in E[x].$$

The coefficients a_j are invariant under the action of G, as $g(x)$ is invariant under the action of G. Hence $g(x) \in F[x]$. Since $g(\alpha) = 0$, it is divisible by $p_\alpha(x)$. By definition, for α_i, there is some $\sigma \in G$ such that $\alpha_i = \sigma(\alpha)$. Hence

$$p_\alpha(\alpha_i) = p_\alpha(\sigma(\alpha)) = \sigma(p_\alpha(\alpha)) = 0,$$

thus each α_i is a root of p_α. Hence $g(x)$ divides p_α, and thus $p_\alpha = g(x)$ splits completely with distinct roots over E. Thus E is normal and separable, as observed in Remark 9.7.6. $\square$

Theorem 9.7.8. *For an extension E/F, the following are equivalent:*

(a) *E is the splitting field of a separable polynomial $f(x) \in F[x]$.*
(b) *E/F is finite and $F = E^{\mathrm{Aut}(E/F)}$.*
(c) *$F = E^G$ for a finite subgroup $G \leq \mathrm{Aut}(E/F)$.*
(d) *E/F is Galois.*

Proof. We first show (a) $\implies$ (b). Since f has finite degree, E/F is finite, as observed in Remark 9.3.4. Let $F' = E^{\mathrm{Aut}(E/F)} \supset F$. We want to show that $[F' : F] = 1$. By the Tower Law, this is equivalent to showing that $[E : F] = [E : F']$. Consider $f \in F'[x]$, whose splitting field is still E, so

$$|\mathrm{Aut}(E/F')| = [E : F'] \leq [E : F] = |\mathrm{Aut}(E/F)|.$$

By Corollary 9.7.2,

$$\mathrm{Aut}(E/F') \overset{def}{=} \mathrm{Aut}(E/E^{\mathrm{Aut}(E/F)}) = \mathrm{Aut}(E/F),$$

and thus $[E : F'] = [E : F]$.

To show $(b) \implies (c)$. Let $G = \mathrm{Aut}(E/F)$. we are given $F = E^G$ and E/F finite.

To show $(c) \implies (d)$, this is just the forward direction of Theorem 9.7.7.

Finally, we show $(d) \implies (a)$. As E/F is finite, by Theorem 9.2.11 there are elements $\alpha_1, \dots, \alpha_n \in E$ algebraic over F such that $E = F(\alpha_1, \dots, \alpha_n)$. Let $p_{\alpha_i} \in F[x]$ be their corresponding minimal polynomials, and let

$$p(x) = \prod_i p_{\alpha_i}(x).$$

By Theorem 9.7.7, E/F is normal and separable, so $p(x)$ splits in E, and has distinct roots, thus $p(x)$ is separable. Hence as E is generated by the roots of $p(x)$, it is the splitting field of $p(x)$, a separable polynomial. □

Corollary 9.7.9 (Artin's Theorem)**.** *Let $G \leq \mathrm{Aut}(E/F)$ be a finite subgroup, and $F = E^G$. Then E/F is Galois and $[E : F] = |G|$.*

Proof. The fact that E/F is Galois is $(c) \implies (d)$ in the proof of Theorem 9.7.8. We have

$$G \overset{\text{Corollary 9.7.2}}{=} \mathrm{Aut}(E/E^G) \overset{E^G=F}{=} \mathrm{Aut}(E/F) \overset{E/F \text{ Galois}}{=} [E : F]$$ □

Definition 9.7.10. For E/G Galois, we write

$$\mathrm{Gal}(E/F) \coloneqq \mathrm{Aut}(E/F),$$

called the *Galois group of E/F*. For a separable polynomial $f \in F[x]$ and E/F a splitting field for f, we write

$$\mathrm{Gal}(f) \coloneqq \mathrm{Gal}(E/F)$$

called the *Galois group of $f(x)$*.

Corollary 9.7.11. *Every finite separable extension E/F is contained in some Galois extension.*

Proof. Since E/F is finite, Theorem 9.2.11 shows that $E = F(\alpha_1, \dots, \alpha_n)$ with $\alpha_i \in E$ algebraic over F, and denote their minimal polynomials $p_i \in F[x]$. Let $p(x) = p_1 \cdots p_n$. By Theorem 9.7.8, the splitting field of p is Galois and contains E. □

Corollary 9.7.12. *Let $E/K/F$ be a tower of extensions with E/F Galois. Then E/K is Galois.*

Proof. By Theorem 9.7.8, E/F is the splitting field of a separable polynomial $f \in F[x]$. Considering $f \in K[x]$, E is still the splitting field, and thus again byTheorem 9.7.8, E/K is Galois. □

Definition 9.7.13. Let E/F be an extension. A *subextension* is a tower of of extensions $E/K/F$.

We are now ready to state and prove the Fundamental Theorem of Galois Theory.

Theorem 9.7.14 (Fundamental Theorem of Galois Theory)**.** *Let E/F be a Galois extension and call $G = \mathrm{Gal}(E/F)$. The map*

$$\{\textit{subgroups } H \leq G\} \longrightarrow \{\textit{subextensions } F \subset K \subset E\}$$

$$H \overset{f}{\longmapsto} E^H$$

is a bijection with inverse

$$\mathrm{Gal}(E/K) \overset{g}{\longmapsfrom} K.$$

For a subgroup $H \leq H$, E/E^H is Galois. Moreover

(a) $H_1 \supset H_2 \iff E^{H_1} \subset E^{H_2}$ *(order reversing)*
(b) $[H_1 : H_2] = [E^{H_2} : E^{H_1}]$

(c) *For $\sigma \in G$ and $H \leq G$, $\sigma H \sigma^{-1} \leftrightarrow \sigma(E^H)$. I.e., $E^{\sigma H \sigma^{-1}} = \sigma(E^H)$ and*

$$\mathrm{Gal}(E/\sigma(K)) = \sigma \,\mathrm{Gal}(E/K)\sigma^{-1}.$$

(d) *$H \leq G$ is normal if and only if E^H/F is normal. In which case E^H/F is Galois, and*

$$\mathrm{Gal}(E^H/F) \cong G/H = \frac{\mathrm{Gal}(E/F)}{\mathrm{Gal}(E/E^H)}.$$

Proof. Corollary 9.7.9 shows that E/E^H is Galois. We need to show that

$$H \overset{f}{\longmapsto} E^H$$
$$\mathrm{Gal}(E/K) \overset{g}{\longmapsto} K$$

are inverses. By Corollary 9.7.2, $\mathrm{Gal}(E/E^H) = H$, so

$$g(f(H)) = g(E^H) = \mathrm{Gal}(E/E^H) = H.$$

For the other direction, Let $E/K/F$ be a subextension. By Corollary 9.7.12, E/K is Galois, thus by Corollary 9.7.9,

$$f(g(K)) = E^{\mathrm{Gal}(E/K)} \overset{\textit{Corollary 9.7.9}}{=} K.$$

Hence f and g are indeed inverse bijections.

To show (a), we note that this is exactly Theorem 9.4.13.

To show (b), let $H \leq G$. Then by Corollary 9.7.9,

$$[\mathrm{Gal}(E/E^H) : 1] = |\,\mathrm{Gal}(E/E^H)| = [E : E^H].$$

This proves (b) when $H_2 = 1$. In general we have

$$[E : E^{H_2}][E^{H_2} : E^{H_1}] = [E : E^{H_1}] = [H_1 : 1] = [H_1 : H_2][H_2 : 1],$$

and we've just shown that $[H_2 : 1] = [E : E^{H_2}]$, hence $[H_1 : H_2] = [E^{H_2} : E^{H_1}]$, as claimed.

To show (c), let $\sigma, \tau \in G$ and $\alpha \in E$. Then

$$\tau\alpha = \alpha \iff \sigma\tau\sigma^{-1}(\sigma\alpha) = \sigma\alpha.$$

Thus $\tau \in G$ fixed K if and only if $\sigma\tau\sigma^{-1}$ fixes $\sigma(K)$. So $\mathrm{Gal}(E/\sigma(K)) = \sigma\,\mathrm{Gal}(E/K)\sigma^{-1}$, and thus

$$\sigma\,\mathrm{Gal}(E/K)\sigma^{-1} \longleftrightarrow \sigma(K).$$

To show (d), let $H \trianglelefteq G$ be a normal subgroup. Since $\sigma H \sigma^{-1} = H$ for all $\sigma \in G$, $\sigma E^H = E^H$. This gives a homomorphism

$$\varphi : G \to \mathrm{Aut}(E^H/F),\ \sigma \mapsto \sigma|_{E^H}.$$

We have $H \subset \ker\varphi$, as H fixes E^H, and if $\sigma \in \ker\varphi$ then $\sigma \in \mathrm{Gal}(E/E^H) = H$. Thus $\ker = H$, giving us an injective homomorphism

$$\widetilde{\varphi} : G/H \to \mathrm{Aut}(E^H/F).$$

As G/H is finite, and $(E^H)^{G/H} = E^G = F$, Theorem 9.7.8 shows that E^H/F is Galois (hence normal) with $[E^H : F] = |G/H|$. Hence $\widetilde{\varphi}$ is surjective, hence an isomorphism. Conversely, suppose that K/F is normal and $H = \mathrm{Gal}(E/K)$. Let $\alpha_1, \ldots, \alpha_n$ generate K over F,a nd let $\sigma \in G$. Then $\sigma(\alpha_i)$ is a root of the minimal polynomial $p_{\alpha_i}(x) \in F[x]$. Since $\alpha_i \in K$, and K/F is normal, $p_{\alpha_i}(x)$ splits over K, hence $\sigma(\alpha_i) \in K$. Thus $\sigma K = K$, and by part (c), $\sigma H \sigma^{-1} = H$. Thus $H \trianglelefteq G$ is normal. □

9.7.1. *S_3 Galois Group Example.* We find the Galois group of x^3-5 over $\mathbb{Q}$, in perhaps a roundabout way, but illustrating how some group theory can shed light on the roots of x^3-5 without even knowing them!

First note that $x^3-5 \in \mathbb{Q}[x]$ is irreducible, you can apply Eisenstein's criterion with $p=5$. Thus x^3-5 is separable, as $\text{char}(\mathbb{Q})=0$. Let E be the splitting field of x^3-5 over $\mathbb{Q}$. Thus

$$[E:\mathbb{Q}] \leq 3! = 6.$$

Since $E/\mathbb{Q}$ is the splitting field of the separable polynomial x^3-5, it is Galois, and so

$$\text{Gal}(x^3-5) = \text{Aut}(E/\mathbb{Q}).$$

Note, however, that x^3-5 has only one real root, which can be seen in a few ways.

One way to see this is to factor x^3-5 in $\mathbb{Q}(\sqrt[3]{5}[x])$ and note that

$$x^3-5 = (x-\sqrt[3]{5})(x^2+\sqrt[3]{5}x+\sqrt[3]{5}^2),$$

and noting that $x^2+\sqrt[3]{5}x+\sqrt[3]{5}^2$ has no real roots using the quadratic formula. Another way is to note that x^3-5 only one critical point ($x=0$), and thus if x^3-5 had two distinct real roots the mean value theorem would imply it had two distinct critical points.

In any case, we have $\mathbb{Q}(\sqrt[3]{5}) \subsetneq E$ as the splitting field of x^3-5 cannot be $\mathbb{Q}(\sqrt[3]{5})$, and thus

$$3 = [\mathbb{Q}(\sqrt[3]{5}):\mathbb{Q}] < [E:\mathbb{Q}] \leq 6.$$

By the tower law, $[\mathbb{Q}(\sqrt[3]{5}):\mathbb{Q}]=3$ divides $[E:\mathbb{Q}]$, hence $[E:\mathbb{Q}]=6$.

Thus as $\#\,\text{Gal}(x^3-5) = [E:\mathbb{Q}]$, $\text{Gal}(x^3-5)$ is a group of order 6. There are precisely two groups of order 6, namely $\mathbb{Z}/6\mathbb{Z}$ and S_3. Now the question is which of these two groups is it?

Well, in E, x^3-5 has exactly 3 distinct roots

$$\{\alpha_1,\alpha_2,\alpha_3\},$$

and $\text{Gal}(x^3-5)$ permutes these roots. Thus, considering the group action

$$\text{Gal}(x^3-5) \curvearrowright \{\alpha_1,\alpha_2,\alpha_3\},$$

we obtain a group homomorphism

$$\text{Gal}(x^3-5) \to S_3.$$

There are now many ways to conclude.

- One way (the fastest) is to note that the map $\text{Gal}(x^3-5) \to S_3$ is just the injective inclusion of the subgroup $\text{Gal}(x^3-5)$ that permutes the roots, and there are no injective group homomorphisms $\mathbb{Z}/6\mathbb{Z} \to S_3$ (as otherwise they would be isomorphic), so $\text{Gal}(x^3-5) \cong S_3$.
- Yet another way is to note that since we can send any root α_i to any other root α_j and thus obtain an element of $\text{Aut}(E/\mathbb{Q})$ (by the proof of Theorem 2 with $F=F'=\mathbb{Q}$, $E=E'$, and $\varphi = \text{id}_{\mathbb{Q}}$ in the notes) we see that $\text{Gal}(x^3-5) \leq S_3$ is a *transitive* subgroup (see below) of S_3. The transitive subgroups of S_3 are isomorphic to $\mathbb{Z}/3\mathbb{Z}$ or S_3, and only the latter has order 6.

In any case, we see that $\text{Gal}(x^3-5) \cong S_3$.

Definition 9.7.15. A subgroup $H \leq S_n$ is called *transitive* if the induced action of H on $\{1,\dots,n\}$ is transitive, i.e., for $x,y \in \{1,\dots,n\}$ there is some $h \in H$ such that $h.x = y$.

Example 9.7.16. If $H \leq S_3$ is transitive, then it cannot be generated by a single transposition, as then it cannot send the third number to either of the two in the transposition. The subgroups of S_3 are:

- the trivial subgroup $\{1\}$,
- $\langle (ij) \rangle$ for a transposition (ij),
- $A_3 = \langle (123) \rangle$ generated by the 3-cycle, and
- S_3.

Hence there are two options for transitive subgroups $H \leq S_3$, either $H = A_3 \cong \mathbb{Z}/3\mathbb{Z}$ or $H = S_3$.

Remark 9.7.17. One important thing to note is that the fact that $\mathrm{Gal}(x^3-5)$ was transitive is no fluke! Indeed, by Lemma 9.3.5 and Theorem 9.3.7, if we take any splitting field E of an irreducible separable $f(x) \in F[x]$, then there is an automorphism taking every root of $f(x)$ to any other root, hence $\mathrm{Gal}(f)$ acts transitively on the roots of $f(x)$.

Example 9.7.18. The Galois group does not necessarily act transitively on the roots of a reducible separable polynomial. Indeed, consider the Galois extension $\mathbb{Q}(\sqrt{2},\sqrt{3})/\mathbb{Q}$, which is the splitting field of the separable polynomial $(x^2-2)(x^2-3)$. We've shown in class that $\mathrm{Gal}(\mathbb{Q}(\sqrt{2},\sqrt{3})/\mathbb{Q}) \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, and the action is be independently acting on the roots of x^2-2 and x^2-3.

In fact, one truly needs irreducibility to get a transitive action on the roots.

Proposition 9.7.19. *Let $f \in F[x]$ be a separable polynomial of degree d. Then f is irreducible in $F[x]$ if and only if $\mathrm{Gal}(f)$ is a transitive subgroup of S_d.*

Remark 9.7.20. Example 9.7.18 hints at the possibility the Galois group of a separable polynomial and the factorization of a polynomial are closely related, and this is true, see the end of Chapter 4 of Milne's book or Section 14.6 and 14.8 of Dummit and Foote for more.

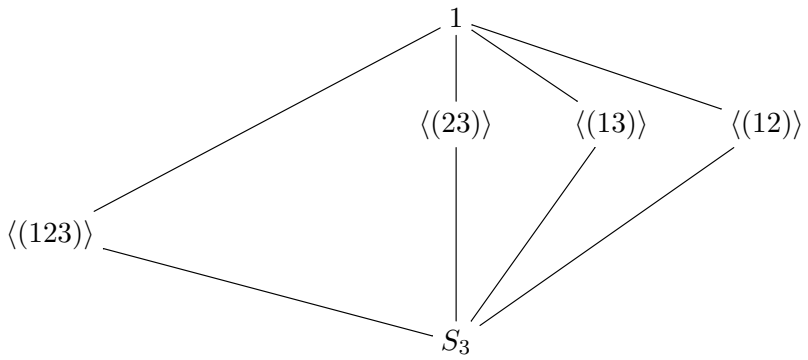
9.7.2. *Galois Correspondence for the splitting field of x^3-5.* We've seen that $\mathrm{Gal}(x^3-5) \cong S_3$. Let $\zeta = e^{2\pi i/3}$ be a primitive third root of unity. Let's write out the subgroups of S_3, and the corresponding fixed fields of $\mathbb{Q}(\sqrt[3]{5},\zeta)$, the splitting field of x^3-5 over $\mathbb{Q}$.

We'll first set some notation. We'll write the three roots of x^3-5 as

$$\alpha_1 = \sqrt[3]{5}, \alpha_2 = \zeta\sqrt[3]{5}, \alpha_3 = \zeta^2\sqrt[3]{5},$$

and let $E \coloneqq \mathbb{Q}(\sqrt[3]{5},\zeta)$ be the splitting field of x^3-5 over $\mathbb{Q}$. And we let $S_3 \cong \mathrm{Gal}(E/\mathbb{Q})$ act by permuting the indices of the roots α_i in the natural way.

The subgroup diagram of S_3 is



and we now determine the corresponding fixed fields.

The field fixed by S_3 is $\mathbb{Q}$.

The to find the field fixed by $\langle(123)\rangle$, we do some exploration. Namely, we note that as

$$(123) \in \mathrm{Gal}(E/\mathbb{Q})$$

is a field homomorphism, it splits up over multiplication, so we have

$$(123)\zeta = (123)\frac{\alpha_2}{\alpha_1} = \frac{(123)\alpha_2}{(123)\alpha_1} = \frac{\alpha_3}{\alpha_2} = \frac{\zeta^2\sqrt[3]{5}}{\zeta\sqrt[3]{5}} = \zeta,$$

thus $\zeta \in E^{\langle(123)\rangle}$. Hence we also have $\zeta^2 = \zeta^{-1} \in E^{\langle(123)\rangle}$.

Thus $\mathbb{Q}(\zeta) \subseteq E^{\langle(123)\rangle}$. By the fundamental theorem of Galois theory, $\mathbb{Q}(\zeta) = E^H$ for some subgroup $H \leq \mathrm{Gal}(E/\mathbb{Q})$. This subgroup must contain $\langle(123)\rangle$, as the correspondence is order-reversing. Thus as there are no subgroup other than S_3 containing A_3, we have $\mathbb{Q}(\zeta) = E^{\langle(123)\rangle}$. We

can independently compute $\mathrm{Gal}(E/\mathbb{Q}(\zeta))$, noting that E is the splitting field of $x^3-5 \in \mathbb{Q}(\zeta)[x]$, and find that $\mathrm{Gal}(E/\mathbb{Q}(\zeta)) = \langle (123)\rangle = A_3 \cong \mathbb{Z}/3\mathbb{Z}$. Indeed, as $\sigma \in \mathrm{Gal}(E/\mathbb{Q}(\zeta))$ preserves $\zeta = \frac{\alpha_2}{\alpha_1}$, we must have

$$\zeta = \sigma\frac{\alpha_2}{\alpha_1} = \frac{\sigma(\alpha_2)}{\sigma(\alpha_1)} = \frac{\alpha_i}{\alpha_j} = \zeta,$$

giving the possibilities

$$\frac{\alpha_i}{\alpha_j} = \frac{\alpha_3}{\alpha_2} \text{ or } \frac{\alpha_2}{\alpha_1},$$

but the latter only occurs when $\sigma = \mathrm{id}$, as then σ must also fix α_3. Hence $\mathrm{Gal}(E/\mathbb{Q}(\zeta)) \cong \mathbb{Z}/3\mathbb{Z}$ with generator (123).

We note that $A_3 \trianglelefteq S_3$ is normal, we have $\mathbb{Q}(\zeta)/\mathbb{Q}$ Galois with $\mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q}) \cong S_3/A_3 \cong \mathbb{Z}/2\mathbb{Z}$. And indeed, we have an automorphism of $\mathbb{Q}(\zeta)$ fixing $\mathbb{Q}$ given by complex conjugation

$$\zeta \mapsto \zeta^2 = \overline{\zeta}.$$

The other fixed fields are found in a similar way to each other similarly. For example, we compute that

$$(23)\sqrt[3]{5} = \sqrt[3]{5},$$

and so $\mathbb{Q}(\sqrt[3]{5}) \subset E^{\langle (12)\rangle}$. As before, there are no subgroups between S_3 and $\langle (23)\rangle$, so

$$E^{\langle (23)\rangle} = \mathbb{Q}(\sqrt[3]{5}).$$

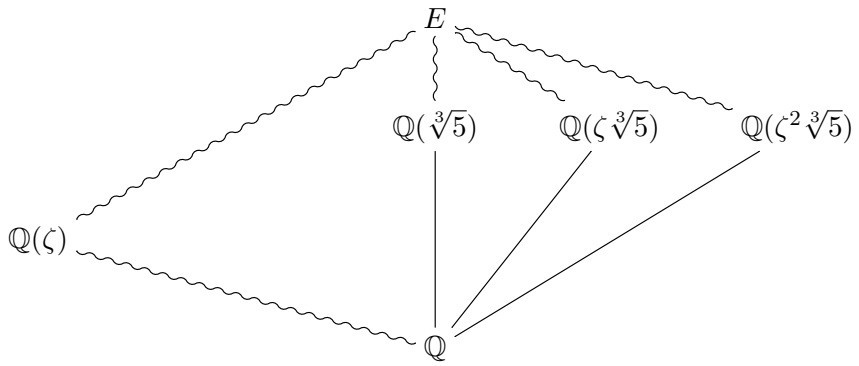
Similarly, we have

$$(12)\alpha_3 = \alpha_3,\ (13)\alpha_2 = \alpha_2$$

so

$$E^{\langle (12)\rangle} = \mathbb{Q}(\alpha_3),\ E^{\langle (13)\rangle} = \mathbb{Q}(\alpha_2).$$

Thus the corresponding diagram of subfields looks like



where we denote Galois extensions with wavy arrows.

Remark 9.7.21. Without knowing the Galois group $\mathrm{Gal}(E/\mathbb{Q})$, we would first need to find the Galois group before we could fully fill in the Galois correspondence. However, some of the subfields (and thus sub groups) can be readily seen from the polynomial (or the roots), and so we can gain some insight into what the Galois group could be without computing the whole group!

If we did not know the Galois group, we could instead try to find it by writing down automorphism that are apparent from the roots. For example, we know the roots of x^3-5 are $\alpha_1, \alpha_2, \alpha_3$, and so we could see how we could lift the isomorphism

$$\varphi : \mathbb{Q}(\alpha_1) \to \mathbb{Q}(\alpha_2),\ \alpha_1 \to \alpha_2$$

all the way to E. In doing so, we see that there is a little freedom, there are two lifts of φ to an automorphism σ of E and they are specified by what they do to α_2. Namely, we know that

$\sigma(\alpha_2) = \alpha_i$ must be another root of $x^3 - 5$, but it cannot be α_2, as then σ would not be injective. We can also check that both choices

$$\sigma(\alpha_2) = \alpha_1 \text{ or } \sigma(\alpha_2) = \alpha_3$$

define automorphisms of E fixing $\mathbb{Q}$, and in the latter case we must have $\sigma(\alpha_3) = \alpha_1$ as it must send α_3 to another root and must be injective. Hence we have constructed the automorphisms (12) and (123) of $\mathrm{Gal}(E/\mathbb{Q})$.

To obtain the automorphisms (23) and (13) we can instead lift the isomorphism

$$\mathbb{Q}(\alpha_3) \to \mathbb{Q}(\alpha_i)$$

with $i = 1, 2$.

10. Applications of Galois Theory

10.1. **Constructible Numbers.** Historically, much of mathematics was concerned with engineering or physics problems. A question of interest to the ancients is to precisely draw figures, for example in building plans or other engineering plans. In particular, what angles, lengths, and figures can you precisely define? The tools the ancients had available were quite limited, and at least in western Europe, constructions with a straightedge and compass were fairly common. In addition, since drawings can be made to scale, one is allowed a "unit length". The first question we would like to answer is which numbers can we define with just a unit length, and straightedge, and a compass?

Definition 10.1.1. A real number $a \in \mathbb{R}$ is *constructible* if a can be constructed as a length using only

- a unit length called 1,
- lines through two previously constructed points, and
- circles with center a previously constructed point and radius a previously constructed length.

Example 10.1.2. In Euclid's *Elements*, Book 1 Proposition 46, one constructs a square with side length 1. Interestingly Proposition 47 is the Pythagorean Theorem. As a corollary, $\sqrt{2}$ is constructible, being the length of a diagonal of the square with unit side length.

With the advent of analytic geometry, and coordinates and equations, since circles are given by

$$(x-a)^2 + (y-b)^2 = c^2$$

for some previously constructed lengths a, b, c, and lines are given by

$$ab + by = c,$$

again with a, b, c previously constructed, we see that constructible numbers can only be obtained as iterated solutions to quadratic equations. Thus we have the following characterization of constructible numbers.

Theorem 10.1.3. *$a \in \mathbb{R}$ is constructible if and only if $a \in \mathbb{Q}(\sqrt{a_1}, \ldots, \sqrt{a_n})$ with $a_i \in \mathbb{Q}(\sqrt{a_1}, \ldots, \sqrt{a_{i-1}})$.*

Corollary 10.1.4. *If $a \in \mathbb{R}$ is constructible, then $[\mathbb{Q}(a) : \mathbb{Q}] = 2^r$ for some r.*

Proof. Since a is constructible, $a \in \mathbb{Q}(\sqrt{a_1}, \ldots, \sqrt{a_n})$, and

$$[\mathbb{Q}(\sqrt{a_1}, \ldots, \sqrt{a_n}) : \mathbb{Q}(a)][\mathbb{Q}(a) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{a_1}, \ldots, \sqrt{a_n}) : \mathbb{Q}] = 2^k. \qquad \square$$

Corollary 10.1.5. *You cannot trisect an arbitrary angle with straightedge and compass.*

Proof. To trisect an angle θ means that $\cos(\theta/3)$ is constructible. But

$$4\cos(\theta/3)^3 - 3\cos(\theta/3) - \cos(\theta) = 0.$$

Thus $\cos(\theta)$ is a root of $4x^3 - 3x - \cos(\theta) \in \mathbb{Q}(\cos(\theta))[x]$. But for $\theta = \frac{2\pi}{6}$, $\cos(\theta) = \frac{1}{2}$. And so $\cos(2\pi/18)$ is a root of $4x^3 - 3x - \frac{1}{2} \in \mathbb{Q}[x]$, or equivalently of $8x^3 - 6x - 1 \in \mathbb{Q}[x]$, this is irreducible by the rational root theorem. Thus $[\mathbb{Q}(\cos(\theta/3)) : \mathbb{Q}] = 3 \neq 2^r$, and by Corollary 10.1.4, $\cos(\theta/3)$ is not constructible. □

Let us now see which angles are constructible.

Since

$$\cos^2 + \sin^2 = 1,$$

we see that $\cos(\theta)$ is constructible if and only if $\sin(\theta)$ is constructible, and from from Euler's identity

$$e^{i\theta = \cos(\theta) + i\sin(\theta)},$$

we see that μ_n is constructible if and only if $\cos(2\pi/n)$ is constructible. Note that

$$\mu_n^2 - 2\cos(2\pi/n)\mu_n + 1 = 0,$$

hence by the quadratic formula, $[\mathbb{Q}(\mu_n) : \mathbb{Q}(\cos(2\pi/n))] = 2$.

Lemma 10.1.6. *Let p be prime. If $\cos(2\pi/p)$ is constructible, then $p = 2^k + 1$.*

Proof. By Euler's identity $e^{i\theta} = cos(\theta) + i\sin(\theta)$, we see that $\mathbb{Q}(\cos(2\pi/p)) \subset \mathbb{Q}(\mu_p)$ for a primitive p^{th} root of unity say $\mu_p = e^{2\pi i/p}$. As $\mathbb{Q}(\mu_p)/\mathbb{Q}$ is Galois, and $[\mathbb{Q}(\mu_p) : \mathbb{Q}(\cos(2\pi/p))] = 2$, we have

$$[\mathbb{Q}(\mu_p) : \mathbb{Q}] = [\mathbb{Q}(\mu_p) : \mathbb{Q}(\cos(2\pi/p))][\mathbb{Q}(\cos(2\pi/p)) : \mathbb{Q}] = 2 \cdot 2^r = 2^{r+1},$$

by Corollary 10.1.4. As $\mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q}) = \mathbb{Z}/p\mathbb{Z}^\times$, we have

$$|\mathbb{Z}/p\mathbb{Z}^\times| = p - 1 = 2^{r+1},$$

thus $p = 2^k + 1$, as claimed. □

Exercise 10.1.7. If $p = 2^r + 1$ is prime, then $r = 2^k$.

Definition 10.1.8. A *Fermat prime* is a prime of the form $2^{2^k} + 1$.

Remark 10.1.9. The history of Fermat primes is quite interesting. I encourage you to also watch the great Numberphile video on YouTube with a construction of a 17-gon.

It turns out that these primes are exactly the primes with $\cos(2\pi/p)$ constructible. We'll need a few facts first.

Lemma 10.1.10. *Let E/F be a degree 2 extension, and suppose $\mathrm{char}\, F \neq 2$. Then $E = F(\sqrt{d})$ for some $d \in F$.*

Proof. Let $\alpha \in E \setminus F$. The minimal polynomial is of degree 2, say $x^2 + bx + c$. And by the quadratic formula

$$\alpha = \frac{-b \pm \sqrt{b^2 - 4c}}{2},$$

so $E = F(\sqrt{b^2 - 4c})$. □

Note that $\mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q}) \cong \mathbb{Z}/p\mathbb{Z}^\times$ is abelian. Thus,

$$\mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q}(\cos(2\pi/p))) \leq \mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q}),$$

being a subgroup of an abelian group, is a normal subgroup. Moreover, the quotient

$$\mathrm{Gal}(\mathbb{Q}(\cos(2\pi/p))/\mathbb{Q}) \cong \frac{\mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q})}{\mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q}(\cos(2\pi/p)))}$$

is abelian.

Now if $p = 2^{2^r} + 1$ is a Fermat prime, then

$$\mathbb{Z}/p\mathbb{Z}^\times \cong (p-1)\mathbb{Z} = \mathbb{Z}/2^{2^r}\mathbb{Z}.$$

Since $[\mathbb{Q}(\mu_p) : \mathbb{Q}(\cos(2\pi/p))] = 2$, we have $\mathrm{Gal}(\mathbb{Q}(\cos(2\pi/p))/\mathbb{Q}) \cong \mathbb{Z}/2$, and thus

$$\mathrm{Gal}(\mathbb{Q}(\cos(2\pi/p))/\mathbb{Q}) \cong \frac{\mathbb{Z}/2^{2^r}\mathbb{Z}}{\mathbb{Z}/2\mathbb{Z}} \cong \mathbb{Z}/(2^{2^r-1})\mathbb{Z}.$$

We have a sequence of normal subgroups

$$1 \trianglelefteq \mathbb{Z}/2\mathbb{Z} \trianglelefteq \mathbb{Z}/4\mathbb{Z} \trianglelefteq \cdots \trianglelefteq \mathbb{Z}/(2^{2^r-1})\mathbb{Z} \cong \mathrm{Gal}(\mathbb{Q}(\cos(2\pi/p))/\mathbb{Q}),$$

giving intermediate extensions

$$\mathbb{Q}(\cos(2\pi/p)) \supset F_{n-1} \supset \cdots F_1 \supset F_0 = \mathbb{Q},$$

with $[F_{i+1} : F_i] = 2$, by Theorem 9.7.14. Since $\mathbb{Q} \subseteq F_i$, $\mathrm{char}\, F_i = 0$, and Lemma 10.1.6 gives $F_{i+1} = F(\sqrt{d})$ for some $d \in F_i$. Hence

$$\mathbb{Q}(\cos(2\pi/p)) = \mathbb{Q}(\sqrt{d_1}, \ldots, \sqrt{d_n})$$

and Theorem 10.1.3 shows that $\cos(2\pi/p)$ is constructible.

Summarizing, we have proven the following.

Proposition 10.1.11. *If $p = 2^{2^r} + 1$ is a Fermat prime, then $\cos(2\pi/p)$ is constructible.*

In total, we have shown the following.

Theorem 10.1.12. *Let p be a prime. The following are equivalent.*

(1) A regular p-gon is constructible.
(2) $\cos(2\pi/p)$ is constructible.
(3) $p = 2^{2^r} + 1$ is a Fermat prime.

Remark 10.1.13. As of the writing of these notes, there are only 5 Fermat primes known. Namely, $3, 5, 17, 257, 65537$. Any updates will likely quickly appear here https://oeis.org/A019434.

Remark 10.1.14. While we did not prove this, since the n^{th} cyclotomic polynomial $\Phi_n \in \mathbb{Q}[x]$ has degree

$$\varphi(n) = \#\{1 \le k \le n, \gcd(k, n) = 1\}$$

where φ is Euler's totient function, we see that $[\mathbb{Q}(\mu_n) : \mathbb{Q}] = \varphi(n)$. Thus by Corollary 10.1.4 and the argument involving $\mathrm{Gal}(\mathbb{Q}(\cos(2\pi/p))/\mathbb{Q})$ above, μ_n is constructible if and only if $\varphi(n)$ is a power of 2. However, for primes $p_1, \ldots, p_k$, we have

$$\varphi(p_1^{e_1} \cdots p_k^{e_k}) = p_1^{e_1-1}(p_1 - 1) \cdots p_k^{e_k-1}(p_k - 1).$$

Thus $\varphi(n)$ is a power of 2 exactly when $n = 2^r \cdot p_1 \cdots p_k$ where the primes p_i are *distinct* Fermat primes. And since $\varphi(n)$ is a power of 2 if and only if $\varphi(n)/2$ is a power of two, recalling that $[\mathbb{Q}(\mu_n) : \mathbb{Q}(\cos(2\pi/n))] = 2$, we immediately see that $\cos(2\pi/n)$ is constructible if and only if n satisfies the same condition with Fermat primes. For more details, we refer the reader to Dummit and Foote §14.5.

Remark 10.1.15. Thus a regular 3-gon (the equilateral triangle) is constructible by straightedge and compass, but a regular 7-gon is not.

Remark 10.1.16. Using origami, one can in fact trisect an angle, and the new numbers that are constructible using a straightedge and compass is a larger subset of algebraic numbers. For more on this, the interested reader is invited to read https://arxiv.org/abs/math/9912039.

10.2. **Fundamental Theorem of Algebra.** Our aim in this section is to prove that the field of complex numbers $\mathbb{C}$, which is the splitting field of $x^2+1 \in \mathbb{R}[x]$ is algebraically closed, a fact known as the Fundamental Theorem of Algebra.

Remark 10.2.1. This is really a "topological fact". At least as far as I have read, I haven't seen a "purely algebraic" proof of the Fundamental Theorem of Algebra.

We'll need two facts about $\mathbb{R}$, and a fact about p-groups.

Lemma 10.2.2 (Fact 1)**.** *If $\alpha \in \mathbb{R}_{>0}$, then $\sqrt{\alpha} \in \mathbb{R}$.*

"Proof". The function $x^2 - \alpha : \mathbb{R} \to \mathbb{R}$ is continuous. By the Intermediate Value Theorem, since at 0 it is negative, and at $b \gg \alpha$ it is positive, there exists some $\beta \in [0, b]$ such that $\beta^2 - \alpha = 0$. □

Lemma 10.2.3 (Fact 2)**.** *If $f(x) \in \mathbb{R}[x]$ has odd degree, then $f(x)$ has a zero in $\mathbb{R}$.*

"Proof". For $b \ll 0$, $f(x)$ is negative. For $c \gg 0$, $f(x)$ is positive. (or vice versa, depending on the sign of the leading coefficient of $f(x)$) By the Intermediate Value Theorem, there exists $\beta \in [b, c]$ such that $f(\beta) = 0$. □

We'll also need one fact about p-groups.

Definition 10.2.4. A group G is *solvable* if there exists a sequence of subgroups

$$P_0 = \{\mathrm{id}_G\} \trianglelefteq P_1 \trianglelefteq P_2 \trianglelefteq \cdots \trianglelefteq P_{k-1} \trianglelefteq P_k = G$$

such that for all i, $P_i \trianglelefteq P_{i+1}$ is a normal subgroup and P_{i+1}/P_i is abelian.

Remark 10.2.5. Abelian groups are solvable, just take $\{\mathrm{id}_G\} \trianglelefteq G$. The group S_5 is not solvable.

Theorem 10.2.6. *Let p be prime, and G a finite $p-$group. Then G is solvable.*

Proof. We've shown (when proving the Sylow Theorems) that $Z(G) \neq \{\mathrm{id}_G\}$ for a p-group (Theorem 5.2.5). Consider the sequence

$$\{\mathrm{id}_G\} \trianglelefteq Z(G) \trianglelefteq G.$$

Now $G/Z(G)$ is also a p-group of smaller order. Bu induction (the base case being the abelian group $\mathbb{Z}/p\mathbb{Z}$), we may assume that we have a sequence

$$Q_0 = \{1\} \trianglelefteq Q_1 \trianglelefteq \cdots \trianglelefteq Q_j = G/Z(G)$$

with Q_{i+1}/Q_i abelian. Let $\pi : G \to G/Z(G)$ be the quotient map, and let $P_i = \pi^{-1}(Q_i)$ be the preimage of Q_i. Then we have

$$P_0 = \{\mathrm{id}_G\} \trianglelefteq Z(G) \trianglelefteq P_1 \trianglelefteq P_2 \trianglelefteq \cdots \trianglelefteq P_{k-1} \trianglelefteq P_k = G$$

Exercise 10.2.7. Show that $P_i \trianglelefteq P_{i+1}$ is indeed normal.

and $P_{i+1}/P_i \cong Q_{i+1}/Q_i$ or $Z(G)$ is abelian. Thus G is solvable. □

Corollary 10.2.8. *Let G be a finite p-group. Then there is a sequence*

$$P_0 = \{\mathrm{id}_G\} \trianglelefteq P_1 \trianglelefteq P_2 \trianglelefteq \cdots \trianglelefteq P_{k-1} \trianglelefteq P_k = G$$

such that $P_{i+1}/P_i \cong \mathbb{Z}/p\mathbb{Z}$.

Proof. Since G is solvable, we have a sequence

$$P_0 = \{\mathrm{id}_G\} \trianglelefteq P_1 \trianglelefteq P_2 \trianglelefteq \cdots \trianglelefteq P_{k-1} \trianglelefteq P_k = G$$

with P_{i+1}/P_i a finite abelian p-group. By the Structure Theorem for finite abelian groups (Theorem 4.0.6), we have

$$P_{i+1}/P_i \cong \mathbb{Z}/p^{e_1}\mathbb{Z} \times \cdots \mathbb{Z}/p^{e_j}\mathbb{Z}.$$

For each of these factors, we have a chain of normal subgroups with quotient $\mathbb{Z}/p\mathbb{Z}$ given by

$$0 \trianglelefteq p^{e_i-1}\mathbb{Z}/p^{e_i}\mathbb{Z} \trianglelefteq \cdots \trianglelefteq p\mathbb{Z}/p^{e_i}\mathbb{Z} \trianglelefteq \mathbb{Z}/p^{e_i}\mathbb{Z},$$

and lifting all of these to expand the original sequence, we obtain a sequence

$$P_0 = \{\mathrm{id}_G\} \trianglelefteq P_1 \trianglelefteq P_2 \trianglelefteq \cdots \trianglelefteq P_{k-1} \trianglelefteq P_k = G$$

with $P_{i+1}/P_i \cong \mathbb{Z}/p\mathbb{Z}$. □

Theorem 10.2.9 (Fundamental Theorem of Algebra)**.** $\mathbb{C}$ *is algebraically closed.*

Proof. $\mathbb{C}$ is the splitting field of $x^2+1 \in \mathbb{R}[x]$, calling a root of x^2+1 $i \in \mathbb{C}$, we have $\mathbb{C} = \mathbb{R}(i)$, a degree 2 extension of $\mathbb{R}$. Recall that there is an automorphism $\bar{\cdot} \in \mathrm{Gal}(\mathbb{C}/\mathbb{R})$ given by complex conjugation, namely $\overline{x+iy} = x - iy$.

We want to show that if $f(x) \in \mathbb{C}[x]$ is non-constant, then $f(x)$ splits over $\mathbb{C}$. Consider $f(x)\overline{f(x)} \in \mathbb{R}[x]$. This has the same roots as $f(x)$, so $f(x) \in \mathbb{C}[x]$ splits over $\mathbb{C}$ if and only if $f(x)\overline{f(x)} \in \mathbb{R}[x]$ splits over $\mathbb{C}$.

Thus we may assume that $f(x) \in \mathbb{R}[x]$ is non-constant, and we aim to show that $f(x) \in \mathbb{C}[x]$ splits into linear factors. We may further suppose that $f(x)$ is monic and irreducible (as a factorization of $f(x)$ gives one of $f(x)\overline{f(x)}$), and that $f(x) \neq x^2+1$.

We first show that every $\alpha \in \mathbb{C}$ has a square-root $\sqrt{\alpha} \in \mathbb{C}$. We can write $\alpha = a+bi$ with $a, b \in \mathbb{R}$, a nd we want to solve $x^2 - \alpha^2 = 0$. Say $x = c + di$ with $c, d \in \mathbb{R}$, so

$$x^2 = c^2 - d^2 + 2cdi.$$

equating real and imaginary parts, we need

$$c^2 - d^2 = a,\ 2cd = b.$$

Now we define

$$c^2 = \frac{a + \sqrt{a^2+b^2}}{2},\ d^2 = \frac{-a + \sqrt{a^2+b^2}}{2}.$$

Since $a^2 + b^2 > 0$, $\sqrt{a^2+b^2} \in \mathbb{R}$ by Lemma 10.2.2. And since $c^2, d^2 > 0$, we also have $c, d \in \mathbb{R}$. Choosing signs of c and d, so that cd has the same sign as b, we have

$$c^2 - d^2 = a,\ 2cd = b,$$

as needed.

Now let E be the splitting field of $(x^2+1)f(x) \in \mathbb{R}[x]$. Clearly $\mathbb{C} \subseteq E$. Since $\mathrm{char}\,\mathbb{R} = 0$, Proposition 9.5.11 shows that $f(x)$ is separable. Thus $E/\mathbb{R}$ is Galois, and we let $G = \mathrm{Gal}(E/\mathbb{R})$, and we let $H \leq G$ be a Sylow 2-subgroup. Then

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

so $[G : H]$ is odd, and $[E^H : \mathbb{R}] = [G : H]$ by Theorem 9.7.14(b).

Thus, for any $\alpha \in E^H$, the minimal polynomial of α over $\mathbb{R}$ has odd degree. But then it has a real root! Hence as the minimal polynomial is irreducible, it has degree 1, and thus $\alpha \in \mathbb{R}$. Hence $E^H = \mathbb{R}$, and so $\mathrm{Gal}(E/\mathbb{R})$ is a 2-group.

We want to show that $E = \mathbb{C}$. Suppose for contradiction that $\mathrm{Gal}(E/\mathbb{C}) \neq \{1\}$, then by Corollary 10.2.8, $\mathrm{Gal}(E/\mathbb{C})$ has a subgroup N of index 2. Since 2 is the smallest prime dividing $|\mathrm{Gal}(E/\mathbb{C})|$, $N \trianglelefteq \mathrm{Gal}(E/\mathbb{C})$ is normal. Thus, by Theorem 9.7.14, $[E^N : \mathbb{C}] = [\mathrm{Gal}(E/\mathbb{C}) : N] = 2$, and by Lemma 10.1.10, we see that $E^N = \mathbb{C}(\sqrt{d})$ for some $d \in \mathbb{C}$. But we just showed that $\sqrt{d} \in \mathbb{C}$, contradicting $[E^N : \mathbb{C}] = 2$. Thus $\mathrm{Gal}(E/\mathbb{C}) = \{1\}$, and $E = \mathbb{C}$. □

10.3. **Loose Ends.** There are a few topics we did not get to during the course of the semester. I wanted to end with some cool and modern developments in Galois Theory, as well as state the whole reason Galois Theory was developed.

Definition 10.3.1. Let $f(x) \in F[x]$. We say that $f(x)$ is *solvable in radicals* if there exists a tower of extensions

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_m$$

such that

- for all i, $F_i = F_{i-1}(\alpha_i)$ with $\alpha_i^{m_i} \in F_{i-1}$,
- F_m contains a splitting field for $f(x)$.

That is, the roots of $f(x)$ are obtained by successive addition, multiplication, division, or taking roots $\sqrt[k]{\ }$.

Theorem 10.3.2 (Galois, 1832)**.** *Let F be a field of characteristic zero, and $f(x) \in F[x]$. Then $f(x)$ is solvable in radicals if and only if the Galois group of f, $\mathrm{Gal}(f)$, is solvable.*

Theorem 10.3.3 (Abel–Ruffini, 1824)**.** *$x^5 - x - 1 \in \mathbb{Q}[x]$ cannot be solved by radicals. ($\mathrm{Gal}(x^5 - x - 1) \cong S_5$)*

In fact, a "general degree ≥ 5 polynomial" in $\mathbb{Q}[x]$ cannot be solved by radicals.

The proofs of these facts are available in both Dummit and Foot and in Milne, and I encourage the interested reader to read the proofs there (if they have not yet taken my previous advice and gone to read those books already).

10.3.1. *Questions in Galois Theory.* We've seen some groups that appear as Galois groups. A natural question to ask is

Question 3. *Which groups are Galois groups?*

Historically, this has been asked over many fields, and is still open over many familiar fields. Much of this exposition is taken from https://arxiv.org/abs/1512.08708, as well as from *Inverse Galois Theory* by Gunter Malle and B. Heinrich Matzat (https://link.springer.com/book/10.1007/978-3-662-55420-3).

Problem 1 (Inverse Galois Problem)**.** Let G be a finite group. Does G appear as a Galois group of some Galois extension over $\mathbb{Q}$? That is, is there a Galois extension $K/\mathbb{Q}$ such that $G \cong \mathrm{Gal}(K/\mathbb{Q})$?

Much is known about this problem, but it is still unsolved. The goal of this section is to prove the following theorem, whose proof is a whirlwind tour through the course.

Theorem 10.3.4. *Let G be a finite abelian group. Then G is a Galois group over $\mathbb{Q}$.*

Before giving the proof, we'll give a brief summary of some known results.

Theorem 10.3.5 (Kronecker–Weber, early 1800's)**.** *Let G be a finite abelian group. Then $G \cong \mathrm{Gal}(K/\mathbb{Q})$ for some extension $K/\mathbb{Q}$ with $K \subseteq \mathbb{Q}(\mu_n)$ for some n depending on G. Moreover, any Galois extension of $\mathbb{Q}$ with abelian Galois group is of this form.*

Remark 10.3.6. The second part of the previous theorem is the hard part! We'll actually prove the first part!

Theorem 10.3.7 (Hilbert, 1892)**.** *For any $n \geq 1$, S_n and A_n are Galois groups over $\mathbb{Q}$.*

Theorem 10.3.8 (Scholz, Reichardt, 1937)**.** *For p an odd prime, every finite p-group is a Galois group over $\mathbb{Q}$.*

Theorem 10.3.9 (Shafarevich, 1958)**.** *Let G be a finite solvable group. Then G is a Galois group over $\mathbb{Q}$.*

Theorem 10.3.10 (1970's)**.** *Let G be a finite group. Then G appears as a Galois group over $\mathbb{C}(t)$, the field of rational functions with complex coefficients.*

Let us now prove Theorem 10.3.4, that finite abelian groups are Galois groups over $\mathbb{Q}$. As usual, we'll need a a few facts first.

Lemma 10.3.11. *Let $m \in \mathbb{Z}_{>0}$. Then there are infinitely many primes such that $p \equiv 1 \mod m$.*

Remark 10.3.12. There is a proof of Lemma 10.3.11 at the end of §13.6 of Dummit and Foote in the exercises using cyclotomic extensions. This also follows from a much more general theorem of Dirichlet on primes in arithmetic progressions.

Theorem 10.3.13. *Let μ_n be a primitive n^{th} root of unity. Then $\mathrm{Gal}(\mathbb{Q}(\mu_n)/\mathbb{Q}) \cong \mathbb{Z}/n\mathbb{Z}^\times$ given by*

$$\begin{aligned} \mathbb{Z}/n\mathbb{Z}^\times &\longrightarrow \mathrm{Gal}(\mathbb{Q}(\mu_n)/\mathbb{Q}) \\ a \mod n &\longmapsto \sigma_a := (\mu_n \mapsto \mu_n^a). \end{aligned}$$

Proof. as $a \in \mathbb{Z}/n\mathbb{Z}^\times$, it is relatively prime to n, and so μ_n^a is another primitive n^{th} root of unity. Thus σ_a is an automorphism obtained from Lemma 9.3.5 by sending μ_n to another root μ_n^a of the irreducible polynomial $\Phi_n \in \mathbb{Q}[x]$. Since

$$\sigma_b(\sigma_a(\mu_n)) = \sigma_b(\mu_n^a) = \mu_n^{ab} = \sigma_{ab}(\mu_n),$$

the map $a \mapsto \sigma_a$ is a group homomorphism. It is bijective as every automorphism of $\mathbb{Q}(\mu_n)$ is obtained by sending μ_n to another root of Φ_n all of the form μ_n^a for some $a \in \mathbb{Z}/n\mathbb{Z}^\times$. □

Theorem 10.3.14. *Let G be a finite abelian group. Then there is a subfield $K \subseteq \mathbb{Q}(\mu_n)$ for some n depending on G such that $K/\mathbb{Q}$ is Galois and $\mathrm{Gal}(K/\mathbb{Q}) \cong G$.*

Proof. From the Structure Theorem for Finitely Generated Abelian Groups, Theorem 4.0.6, we see that

$$G \cong \mathbb{Z}/n_1\mathbb{Z} \times \cdots \mathbb{Z}/n_k\mathbb{Z}.$$

Let $p_1, \ldots, p_k$ be distinct primes such that $p_i \equiv 1 \mod n_i$, which exist by Lemma 10.3.11. Let $n = p_1 \cdots p_k$. By the Chinese Remainder Theorem, Theorem 4.5.12 or Theorem 4.5.4, we have

$$\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/p_1\mathbb{Z} \times \cdots \times \mathbb{Z}/p_k\mathbb{Z},$$

thus

$$\mathbb{Z}/n\mathbb{Z}^\times \cong \mathbb{Z}/p_1\mathbb{Z}^\times \times \cdots \times \mathbb{Z}/p_k\mathbb{Z}^\times \cong \mathbb{Z}/(p_1-1)\mathbb{Z} \times \cdots \times \mathbb{Z}/(p_k-1)\mathbb{Z}.$$

By construction, $p_i - 1 \equiv 0 \mod n_i$, so n_i divides $p_i - 1$. Hence, as you should show, $\mathbb{Z}/(p_i-1)\mathbb{Z}$ has a subgroup $H_i \trianglelefteq \mathbb{Z}/(p_i-1)\mathbb{Z}$ of order $|H_i| = \frac{p_i-1}{n_i}$.

Exercise 10.3.15. Let A be an abelian group of order m. If k divides m, then A has a subgroup of order k. This is not true when A is not assumed to be abelian.

And moreover, as $\mathbb{Z}/(p_i-1)\mathbb{Z}$ is cyclic, $(\mathbb{Z}/(p_i-1)\mathbb{Z})/H_i$ is cyclic, of order n_i. Thus

$$\frac{\mathbb{Z}/(p_1-1)\mathbb{Z} \times \cdots \times \mathbb{Z}/(p_k-1)\mathbb{Z}}{H_1 \times \cdots \times H_k} \cong \mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_k\mathbb{Z} \cong G.$$

By Theorem 10.3.13 and the Fundamental Theorem of Galois Theory, Theorem 9.7.14, there is a subextension $\mathbb{Q} \subset K \subset \mathbb{Q}(\mu_n)$ corresponding to the subgroup $H_1 \times \cdots \times H_k$. Thus

$$\mathrm{Gal}(\mathbb{Q}(\mu_n)/K) \cong H_1 \times \cdots \times H_k \trianglelefteq \mathbb{Z}/n\mathbb{Z}^\times \cong \mathrm{Gal}(\mathbb{Q}(\mu_n)/\mathbb{Q})$$

is a normal subgroup (as $\mathbb{Z}/n\mathbb{Z}^\times$ is abelian), and so by Theorem 9.7.14, $K/\mathbb{Q}$ is Galois with Galois group

$$\mathrm{Gal}(K/\mathbb{Q}) \cong \frac{\mathrm{Gal}(\mathbb{Q}(\mu_n)/\mathbb{Q})}{\mathrm{Gal}(\mathbb{Q}(\mu_n)/K)} \cong \frac{\mathbb{Z}/n\mathbb{Z}^\times}{H_1 \times \cdots \times H_k} \cong G.$$

□

Remark 10.3.16. The story of the Inverse Galois Problem is still being written. For some recent results, we direct the reader to https://arxiv.org/abs/2411.07857.

Remark 10.3.17. The abelian extensions of a field are in some sense the "easiest" to understand, and the study of abelian extensions of fields has been developed under the name "Class Field Theory" between the mid 1800s and the mid 1900s. These results have been significantly extended to a "standard approach" around the 1950s and 1980s, and is still a significant area of research today in studying non-abelian extensions and their relations to geometry and number theory.