

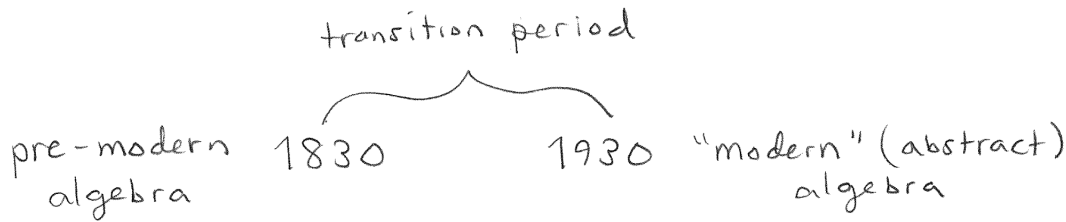
Contents

Week 1	
Quadratic and Cubic Equations, Lagrange's Method	1
Week 2	
Permutations, Definition of Groups and Subgroups, Main Examples, Statement of Galois' Theorem	10
Problem Set 1	18
Week 3	
Subgroup Generated by a Subset, Intersection and Join of Subgroups, Cyclic Groups, The Circle Group	28
Week 4	
Orthogonal and Unitary Groups, Dihedral Groups, Definition of Isomorphism	34
Problem Set 2	42
Week 5	
Posets and Lattices, Galois Connections, The Correspondence Theorem, The Fundamental Theorem of Cyclic Groups	52
Week 6	
Equivalence Modulo a Subgroup, Lagrange's Theorem, Quotients of Abelian Groups	64
Problem Set 3	74
Week 7	
Normal Subgroups, Quotient Groups in General, The First Isomorphism Theorem	81
Week 8	
Definition of Automorphisms, Definition of Group Actions, Cayley's Theorem, The Center and Inner Automorphisms	92
Problem Set 4	99

Week 9		
	Direct and Semidirect Products of Groups, Isometries of Euclidean Space	108
Week 10		
	Simple Groups, The Jordan-Hölder Theorem, Proof that S_n is not Solvable	
	The Orbit-Stabilizer Theorem, Free and Transitive Actions	119
Problem Set 5		132
Week 11		
	The Class Equation, Groups of Size p and p^2 , The Sylow Theorems	142
Week 12		
	Proof That A_5 is Simple, Classification of Finite Simple Groups	148
Problem Set 6		156

Week 1

What is “algebra?” The meaning of the word has changed over time. Here’s a historical sketch:



In this course we will discuss “modern” (also known as abstract) algebra. But in order to motivate this, I will first talk about pre-modern algebra. Prior to 1830 algebra was understood at the study of polynomial equations.

Example. Let a, b, c be any numbers. Find all numbers x such that

$$ax^2 + bx + c = 0.$$

I assume that you all learned about quadratic equations in high school. If $a = 0$ then there is nothing interesting to do, so let us assume that $a \neq 0$ and divide both sides by a to get

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

$$x^2 + \frac{b}{a}x = -\frac{c}{a}.$$

Now there a famous trick called “completing the square.” If we add the quantity $(b/2a)^2$ to both sides then it turns out that the left side factors:

$$\begin{aligned} x^2 + \frac{b}{a}x &= -\frac{c}{a} \\ x^2 + \frac{b}{a}x + \left(\frac{b}{2a}\right)^2 &= -\frac{c}{a} + \left(\frac{b}{2a}\right)^2 \\ \left(x + \frac{b}{2a}\right)\left(x + \frac{b}{2a}\right) &= -\frac{c}{a} + \frac{b^2}{4a^2} \\ \left(x + \frac{b}{2a}\right)^2 &= \frac{b^2 - 4ac}{4a^2}. \end{aligned}$$

Finally, we take “the” square root of both sides and then solve for x :

$$\begin{aligned} \left(x + \frac{b}{2a}\right)^2 &= \frac{b^2 - 4ac}{4a^2} \\ x + \frac{b}{2a} &= \frac{\sqrt{b^2 - 4ac}}{2a} \\ x &= -\frac{b}{2a} + \frac{\sqrt{b^2 - 4ac}}{2a} \\ &= \frac{-b + \sqrt{b^2 - 4ac}}{2a}. \end{aligned}$$

Wait, I lied. There is no such thing as “the” square root of a number. Actually every number (except 0) has **two** different square roots. So the “quadratic formula”

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

is not really a formula at all, but more of a “recipe” that tells us how to compute the two roots of the equation. First, let $\sqrt{b^2 - 4ac}$ denote **one** of the two square roots of the number $b^2 - 4ac$. (I don’t care which one; you can choose your favorite.) Then the **other** square root is just the negative: $-\sqrt{b^2 - 4ac}$. Thus we obtain (in general) two different solutions to the quadratic equation:

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{or} \quad x = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

///

This algorithm was known to ancient civilizations and was slowly put into symbolic form over time. The advantage of the symbolic form is that it encapsulates many different situations and

applications. For example: The equation $ax^2 + bx + c = 0$ might represent the intersection of two shapes in the plane; say, a line and a circle. If the quantity $b^2 - 4ac$ (called the “discriminant”) is negative then its two square roots are imaginary, which means that the line and the circle don’t intersect.

So far so good, but now we come to a subject that was not known to ancient civilizations. During the Italian Renaissance of the 1500s, a group of mathematicians discovered similar (but more complicated) formulas for the cubic and quartic equations.

Example. Let a, b, c, d be any numbers. Find all numbers x such that

$$ax^3 + bx^2 + cx + d = 0.$$

There is a clever change of variables (never mind the details) that can remove the x^2 term. Also, we might as well assume that the leading coefficient is $a = 1$, otherwise we can just divide both sides by a . Thus it is enough to solve the so-called “depressed cubic:”

$$x^3 + px + q = 0.$$

Here’s the solution, which is called “Cardano’s formula:”

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Again, this is not really a formula, but a “recipe” for finding the solutions. Just like “the” square root, there is no such thing as “the” cube root, since every nonzero number actually has **three** cube roots. Unfortunately this makes Cardano’s formula difficult to interpret, since there might be **nine** ways to choose the cube roots but the cubic equation $x^3 - px + q = 0$ has only **three** solutions. We’ll discuss a more comprehensive method below.

///

The difficulty of interpreting Cardano’s formula is the historical reason why people finally accepted the concept of complex numbers. For example, we know that the equation $x^3 - 15x - 4 = 0$ has the real root $x = 4$, but Cardano’s formula gives

$$x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}.$$

The only way to get from this expression to $x = 4$ is to accept the fact that $2 + \sqrt{-1}$ is a cube root of $2 + \sqrt{-121}$ and $2 - \sqrt{-1}$ is a cube root of $2 - \sqrt{-121}$. That is, sometimes the only way to get to a real solution is by going through the imaginary numbers. In modern terms we write

$$\mathbb{C} = \{a + ib : a, b \in \mathbb{R}\},$$

where $i = \sqrt{-1}$ is one of the two square roots of -1 . (I don’t care which one; you can pick your favorite.)

After a burst of activity in the 1500s people got stuck for hundreds of years on the following question.

Question. Does there exist a formula for the quintic? Given any numbers a, b, c, d, e, f we consider the equation

$$ax^5 + bx^4 + cx^3 + dx^2 + ex + f = 0.$$

Is it possible to write down a recipe for the roots x in terms of the coefficients a, b, c, d, e, f and the “algebraic” operations $+, -, \times, \div, \sqrt{}, \sqrt[3]{}, \sqrt[4]{}, \sqrt[5]{}, \dots$?

[Jargon: Is the quintic equation “solvable by radicals?”]

In 1770 Joseph-Louis Lagrange wrote an influential book summarizing the state of knowledge of this problem. He had very sophisticated methods that led him to believe that the problem was likely impossible. In 1799 Paolo Ruffini claimed to have a proof of impossibility, but it was flawed. Then in 1823 the young Norwegian mathematician Niels Henrik Abel (1802-1829) finally gave an air-tight proof of the following theorem.

The Abel-Ruffini Theorem. Let $n \geq 5$. It is **impossible** in general to write down the roots of an n -th degree polynomial equation in terms of the coefficients and the algebraic operations

$$+, -, \times, \div, \sqrt{}, \sqrt[3]{}, \sqrt[4]{}, \sqrt[5]{}, \dots$$

In other words, there exist polynomial equations of all degrees ≥ 5 that are **not** solvable by radicals.

This was a historic achievement but Abel’s proof was long and complicated. Furthermore, there are certain special kinds of polynomials that **can** be solved by radicals.

Example. The quintic equation $ax^5 + b = 0$ is easily solvable. Here’s the recipe:

$$\begin{aligned} ax^5 + b &= 0 \\ ax^5 &= -b \\ x^5 &= -\frac{b}{a} \\ x &= \sqrt[5]{-\frac{b}{a}}. \end{aligned}$$

Now one just has to compute the five 5th roots of the number $-b/a$.

[Remark: How to do this? In general one can use numerical methods (like Newton’s method) to find one particular fifth root, say $\alpha \in \mathbb{C}$ where $\alpha^5 = -b/a$. Then the complete list of 5th roots is

$$\alpha, \quad \omega\alpha, \quad \omega^2\alpha, \quad \omega^3\alpha, \quad \omega^4\alpha,$$

where $\omega = e^{2\pi i/5}$ is called a “primitive 5th root of 1.”]

That was sort of obvious, but there are more complicated examples.

Example. De Moivre’s quintic has the form

$$x^5 + ax^3 + \frac{1}{5}a^2x + b = 0.$$

It turns out (you don’t want to see the details) that the roots of this equation r_1, r_2, r_3, r_4, r_5 are given by the formula

$$r_j = \omega^j u_1 + \omega^{-j} u_2,$$

where ω is any primitive 5th root of 1 (e.g. $\omega = e^{2\pi i/5}$) and where u_1 and u_2 are any two numbers satisfying

$$\begin{cases} u_1^5 + u_2^5 &= -b, \\ u_1^5 \cdot u_2^5 &= (-a/5)^5. \end{cases}$$

[Exercise: Use the quadratic formula to solve for the numbers u_1^5 and u_2^5 .]

Long story short: De Moivre’s quintics **are** solvable by radicals. So the next order of business was to clean up the details and understand the distinction between solvable and non-solvable polynomials. Abel intended to do this.

Abel’s Research Program. Find a method to determine which polynomials are solvable by radicals, and which are not.

But then he died suddenly (of tuberculosis) at the age of 26. This brings us to the year 1829. At this time, a young frenchman named Évariste Galois was working in obscurity on similar problems. Galois had some brilliant and visionary ideas, but he also had a volatile personality and he died in a duel in 1832 at the age of 20, before he could gain any recognition for these ideas.

Thus we have reached the year 1832. The two greatest algebraists of the age are dead; one was famous (Abel) and one was completely unknown (Galois). Galois had left behind some hastily written manuscripts, but no one had really read or understood them yet.

However, Galois’ work was not lost and it slowly influenced others until by 1930 it had completely changed the definition of the word “algebra.” The final steps of this transition were completed at Göttingen in the lectures of Emil Artin and Emmy Noether, which were incorporated by Bartel van der Waedern into his famous textbook *Moderne Algebra* (1930). This new “modern” style of algebra is the topic of our course.

Galois completely changed the rules of the game. Instead of “numbers” and “polynomials,” he decided to focus on “symmetries” and relationships between symmetries. Galois used the word “group” to refer to a “collection of symmetries.” Let me motivate this by showing you Lagrange’s (1770) approach to the quadratic and cubic equations.

Lagrange’s Solution of the Quadratic. Instead of writing

$$ax^2 + bx + c = 0,$$

we will assume that $a \neq 0$ and divide both sides by a to get

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0.$$

To clean this up a bit, we will also rename the coefficients as $e_1 := -b/a$ and $e_2 := c/a$, so that

$$x^2 - e_1x + e_2 = 0.$$

Now we are looking for two numbers r_1 and r_2 (the “roots” of the equation) such that

$$x^2 - e_1x + e_2 = (x - r_1)(x - r_2).$$

[Remark: We know that these are the roots because $(x - r_1)(x - r_2) = 0$ if and only if $x = r_1$ or $x = r_2$.]

Our goal is to solve for the unknown roots r_1, r_2 in terms of the given coefficients e_1, e_2 . To begin, we expand the right hand side:

$$\begin{aligned} x^2 - e_1x + e_2 &= (x - r_1)(x - r_2) \\ x^2 - e_1x + e_2 &= x^2 - (r_1 + r_2)x + r_1r_2. \end{aligned}$$

And then we compare coefficients to obtain a system of two equations in two unknowns.

$$\begin{cases} e_1 = r_1 + r_2, \\ e_2 = r_1r_2. \end{cases}$$

In this way, we can think of $e_1(r_1, r_2) = r_1 + r_2$ and $e_2(r_1, r_2) = r_1r_2$ as “functions” of the unknown roots r_1, r_2 . Furthermore, let me observe that each of these functions is “symmetric” under “permuting” the two roots:

$$\begin{aligned} e_1(r_1, r_2) &= r_1 + r_2 = r_2 + r_1 = e_1(r_2, r_1) \\ e_2(r_1, r_2) &= r_1r_2 = r_2r_1 = e_2(r_2, r_1). \end{aligned}$$

We would like to “invert” the system of two equations to obtain

$$\begin{cases} r_1 = \text{some function of } e_1, e_2, \\ r_2 = \text{some other function of } e_1, e_2. \end{cases}$$

But unfortunately this is **impossible**. Indeed, since e_1 and e_2 are each symmetric under permuting $r_1 \leftrightarrow r_2$, any function of e_1 and e_2 must also be symmetric. But the simple functions $f(r_1, r_2) = r_1$ and $g(r_1, r_2) = r_2$ are very much **not symmetric**. QED

Thus, Lagrange's problem is to break down the given "symmetric functions" $e_1(r_1, r_2) = r_1 + r_2$ and $e_2(r_1, r_2) = r_1 r_2$ into the unknown "non-symmetric functions" $f(r_1, r_2) = r_1$ and $g(r_1, r_2) = r_2$.

First, Lagrange made a change of variables. (We now call this a "Lagrange resolvent.") He defined two new functions $s_1(r_1, r_2)$ and $s_2(r_1, r_2)$ by

$$\begin{cases} s_1 = r_1 + r_2, \\ s_2 = r_1 - r_2. \end{cases}$$

Note that this (linear) system is easily invertible:

$$\begin{cases} r_1 = (s_1 + s_2)/2, \\ r_2 = (s_1 - s_2)/2. \end{cases}$$

Thus we will be done if we can solve for s_1 and s_2 in terms of e_1 and e_2 :

$$\begin{cases} s_1 = \text{some function of } e_1, e_2 ? \\ s_2 = \text{some other function of } e_1, e_2 ? \end{cases}$$

The first one is easy:

$$s_1 = r_1 + r_2 = e_1.$$

But the second equation is still impossible because any function of e_1, e_2 is symmetric in r_1, r_2 , while s_2 is still **not** symmetric:

$$s_2(r_1, r_2) = r_1 - r_2 \neq r_2 - r_1 = s_2(r_2, r_1).$$

[Jargon: We say that $s_2(r_1, r_2) = r_1 - r_2$ is an "alternating function" of r_1, r_2 because switching $r_1 \leftrightarrow r_2$ multiplies the function by -1 .]

So it all comes down to this:

How can we convert the alternating function s_2 into a symmetric function?

This is easy; we can just square it:

$$s_2(r_1, r_2)^2 = (r_1 - r_2)^2 = r_1^2 - 2r_1 r_2 + r_2^2.$$

Since s_2^2 is now a symmetric function, an old theorem of Isaac Newton guarantees that we can express it as a polynomial in the "elementary" symmetric functions e_1 and e_2 . (This is secretly why I used the letter "e" for the coefficients.) There is an algorithm, but trial-and-error works just as well in such a small case:

$$e_1^2 = (r_1 + r_2)^2$$

$$\begin{aligned}
e_1^2 &= r_1^2 + 2r_1r_2 + r_2^2 \\
e_1^2 - 4e_2 &= (r_1^2 + 2r_1r_2 + r_2^2) - 4r_1r_2 \\
e_1^2 - 4e_2 &= r_1^2 - 2r_1r_2 + r_2^2 \\
e_1^2 - 4e_2 &= s_2^2.
\end{aligned}$$

Finally, let $s_2 = \sqrt{e_1^2 - 4e_2}$ be either one of the two square roots. (I don't care which one; pick your favorite.) This is precisely where we "break the symmetry." Then the final answer is

$$\begin{cases} r_1 = (s_1 + s_2)/2 &= (e_1 + \sqrt{e_1^2 - 4e_2})/2 \\ r_2 = (s_1 - s_2)/2 &= (e_1 - \sqrt{e_1^2 - 4e_2})/2. \end{cases}$$

Do you recognize this as the quadratic formula?

///

Continuing with our discussion of Lagrange's method:

Lagrange's Solution of the Cubic. Let's cut to the chase. For any three given coefficients e_1, e_2, e_3 , we want to find three roots r_1, r_2, r_3 such that

$$x^3 - e_1x^2 + e_2x - e_3 = (x - r_1)(x - r_2)(x - r_3).$$

By expanding the right hand side and equating coefficients, this is equivalent to the following system of three (nonlinear) equations in three unknowns:

$$\begin{cases} e_1 = r_1 + r_2 + r_3, \\ e_2 = r_1r_2 + r_1r_3 + r_2r_3, \\ e_3 = r_1r_2r_3. \end{cases}$$

Please observe that each of the functions e_1, e_2, e_3 is symmetric under any permutation of the input r_1, r_2, r_3 . For example,

$$e_2(r_3, r_1, r_2) = r_3r_1 + r_3r_2 + r_1r_2 = r_1r_2 + r_1r_3 + r_2r_3 = e_2(r_1, r_2, r_3).$$

[Jargon: These e_1, e_2, e_3 are called the "elementary symmetric functions" of r_1, r_2, r_3 .]

Our job is to "break the symmetry" in a controlled way. Step 1 is the "Lagrange resolvent." We define three new functions s_1, s_2, s_3 by the linear system

$$\begin{cases} s_1 = r_1 + r_2 + r_3, \\ s_2 = r_1 + \omega r_2 + \omega^2 r_3 \\ s_3 = r_1 + \omega^2 r_2 + \omega r_3, \end{cases}$$

where ω is any primitive third root of unity, say $\omega = e^{2\pi i/3}$. This system is easily invertible:

$$\begin{cases} r_1 = (s_1 + s_2 + s_3)/3, \\ r_2 = (s_1 + \omega^2 s_2 + \omega s_3)/3 \\ r_3 = (s_1 + \omega s_2 + \omega^2 s_3)/3. \end{cases}$$

Therefore our new problem is to solve for s_1, s_2, s_3 in terms of e_1, e_2, e_3 . The first one is always easy:

$$s_1 = r_1 + r_2 + r_3 = e_1.$$

But the next two are **impossible** because s_2 and s_3 are not symmetric functions.

So here's the problem:

How can we convert the non-symmetric functions s_2 and s_3 into symmetric functions of the roots, and hence express them in terms of the elementary symmetric functions e_1, e_2, e_3 ?

This is very tricky (indeed, if there were some general algorithm then all polynomials would be solvable), so I'll just tell you the answer. After a bit of trial-and-error you will find that $s_2 s_3$ is a symmetric function and after a **lot** of trial-and-error you will find that $s_2^3 + s_3^3$ is a symmetric function. Thus by Newton's theorem each of these can be expressed as a polynomial in e_1, e_2, e_3 . Here are the results:

$$\begin{cases} s_2 s_3 &= e_1^2 - 3e_2 \\ s_2^3 + s_3^3 &= 2e_1^3 - 9e_1 e_2 + 27e_3. \end{cases}$$

The last step is to "solve" for s_2 and s_3 individually. To make the notation cleaner let us define

$$A := s_2^3 + s_3^3 = 2e_1^3 - 9e_1 e_2 + 27e_3 \quad \text{and} \quad B := s_2 s_3 = e_1^2 - 3e_2.$$

Then we observe that

$$(y - s_2^3)(y - s_3^3) = y^2 - (s_2^3 + s_3^3)y + s_2^3 s_3^3 = y^2 - Ay + B^3.$$

On the one hand we know that the roots of this quadratic are $y = s_2^3$ and $y = s_3^3$. On the other hand, we can use the quadratic formula to find that

$$y = \frac{1}{2} \left(A \pm \sqrt{A^2 - 4B^3} \right).$$

Thus by "breaking the symmetry" we can write

$$s_2^3 = \frac{1}{2} \left(A + \sqrt{A^2 - 4B^3} \right) \quad \text{and} \quad s_3^3 = \frac{1}{2} \left(A - \sqrt{A^2 - 4B^3} \right)$$

And then by "breaking the symmetry" again we can write

$$s_2 = \sqrt[3]{\frac{1}{2} \left(A + \sqrt{A^2 - 4B^3} \right)} \quad \text{and} \quad s_3 = \sqrt[3]{\frac{1}{2} \left(A - \sqrt{A^2 - 4B^3} \right)}.$$

Now we're done. I won't bother to write the full formulas for r_1, r_2, r_3 in terms of e_1, e_2, e_3 because they are ugly and they don't fit horizontally on the page. Instead I'll have you work out an example on the homework. ///

In general, let e_1, e_2, \dots, e_n be any given numbers. Then we want to find n roots r_1, r_2, \dots, r_n such that

$$x^n - e_1x^{n-1} + e_2x^{n-2} - \dots + (-1)^n e_n = (x - r_1)(x - r_2) \cdots (x - r_n).$$

By expanding and equating coefficients this is equivalent to a system of n nonlinear equations in n unknowns:

$$\begin{cases} e_1 &= \sum_i r_i \\ e_2 &= \sum_{i < j} r_i r_j \\ e_3 &= \sum_{i < j < k} r_i r_j r_k \\ &\vdots \\ e_n &= r_1 r_2 \cdots r_n. \end{cases}$$

Since the e_i are “symmetric functions” and the r_i are not, it is impossible to invert this system without “breaking the symmetry.” As we have seen, Lagrange’s method succeeds if we can break down the symmetry in a sequence of controlled steps, where each step is just “choosing an arbitrary k -th root” of some polynomial expression. Ruffini and Abel eventually proved that this problem is impossible for $n \geq 5$.

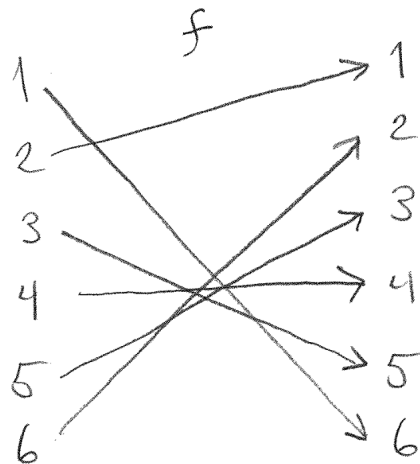
Week 2

Then Galois changed the rules of the game. Specifically, he decided to ignore the “symmetric functions” and to focus instead on the “symmetries” in themselves.

Permutations. A *permutation* is an invertible function from a finite set to itself. Since all sets of the same size are basically the same we will usually consider the set $\{1, 2, \dots, n\}$. Let S_n denote the set of all permutations

$$f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}.$$

For example, here is a typical element of S_6 :



It is cumbersome to draw the full diagram every time, so we have two more concise notations.

Word Notation. Given $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ we will prefer to write f_i instead of $f(i)$. Then to specify the function f it is enough to give the list of values f_1, f_2, \dots, f_n . To save as much space as possible (if $n \leq 9$) we will even omit the commas and write

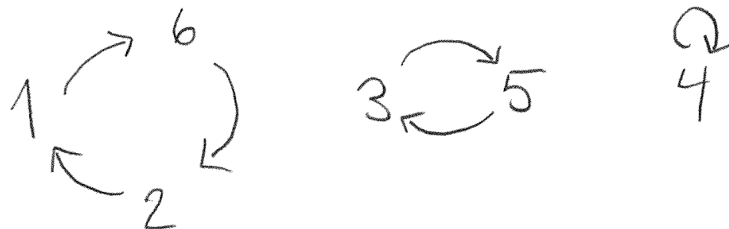
$$f = f_1 f_2 \cdots f_n.$$

For example, the permutation above is

$$f = 615432.$$

[Exercise: Use word notation to prove that $\#S_n = n!$.]

Cycle Notation. Word notation is the most concise way to express permutations, but cycle notation is the most meaningful way. To compute the cycle notation we write down just **one** copy of the symbols and then we draw the arrows. Here is our example:



Note that the symbols break up into “oriented cycles.” To express these cycles concisely we just put them inside parentheses, like so:

$$f = (162)(35)(4).$$

The only drawback of this notation is that it is not unique. For example, we can record a cycle starting from any point:

$$(162) = (621) = (216).$$

And the ordering among the cycles is irrelevant:

$$f = (162)(35)(4) = (4)(162)(35) = (53)(4)(621).$$

Another quirk of notation is that we typically omit the “singleton cycles” from the notation. In our example this means omitting the (4):

$$f = (162)(35).$$

We will see that the most important kinds of permutations are the *transpositions*, which switch one pair of symbols $i \leftrightarrow j$ and send every other symbol to itself. Transpositions are particularly simple when expressed in cycle notation:

$$(ij) \in S_n.$$

[Exercise: Show that the set S_n contains $n(n-1)/2$ transpositions.]

///

Here is Galois’ big contribution to mathematics.

Galois’ Theorem. Consider a positive integer $n \geq 1$ and let S_n be the set of all permutations of the set $\{1, 2, \dots, n\}$. Let $\text{id} : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ be the “identity permutation” that sends each element to itself. Then the general n -th degree equation is solvable by radicals if and only if there exists a chain of subsets

$$S_n = G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_k = \{\text{id}\}$$

in which each pair $G_i \supseteq G_{i+1}$ satisfies some technical conditions.

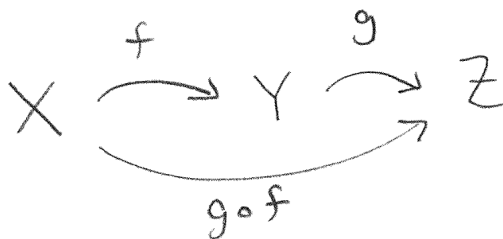
///

In terms of Lagrange’s method, the technical conditions on $G_i \supseteq G_{i+1}$ correspond to “breaking the symmetry” by choosing an arbitrary root of some function. The advantage of Galois’

reformulation is that it will eventually allow us to give a short proof of unsolvability for $n \geq 5$, without even mentioning “equations” or “roots.”

So what are the technical conditions?

Composition of Permutations. Let X, Y, Z be sets and let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be functions. Since the target set of f equals the domain set of g , we may *compose* them to obtain a function from X to Z :



The function $g \circ f$ is called “ g composed with f ” or “ g follows f .” The reason we write g on the left is because on this side of the Atlantic we always write functions to the left of their arguments:

$$(g \circ f)(x) := g(f(x)) \text{ for all } x \in X.$$

Now suppose that $X = Y = Z = \{1, 2, \dots, n\}$ and suppose that each of f and g is invertible. In other words, suppose that $f, g \in S_n$. Then the composition $g \circ f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ is also invertible.

Proof. Suppose there exist functions $f^{-1}, g^{-1} : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ such that

$$f(f^{-1}(i)) = f^{-1}(f(i)) = i \quad \text{and} \quad g(g^{-1}(i)) = g^{-1}(g(i)) = i$$

for all $i \in \{1, 2, \dots, n\}$. Then I claim that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. Indeed, for all $i \in \{1, 2, \dots, n\}$ we have

$$(g \circ f)((f^{-1} \circ g^{-1})(i)) = g(\cancel{f(f^{-1}(g^{-1}(i)))}) = g(g^{-1}(i)) = i$$

and

$$(f^{-1} \circ g^{-1})((g \circ f)(i)) = f^{-1}(\cancel{g^{-1}(g(f(i)))}) = f^{-1}(f(i)) = i.$$

□

Better Proof. The inverses satisfy $f \circ f^{-1} = f^{-1} \circ f = \text{id}$ and $g \circ g^{-1} = g^{-1} \circ g = \text{id}$. Then since functional composition is “associative” we have

$$(g \circ f) \circ (f^{-1} \circ g^{-1}) = g \circ (f \circ f^{-1}) \circ g^{-1} = g \circ \text{id} \circ g^{-1} = g \circ g^{-1} = \text{id}$$

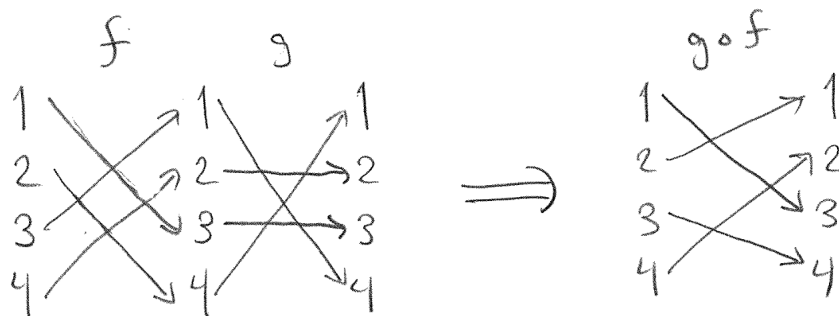
and

$$(f^{-1} \circ g^{-1}) \circ (g \circ f) = f^{-1} \circ (g^{-1} \circ g) \circ f = f^{-1} \circ \text{id} \circ f = f^{-1} \circ f = \text{id}.$$

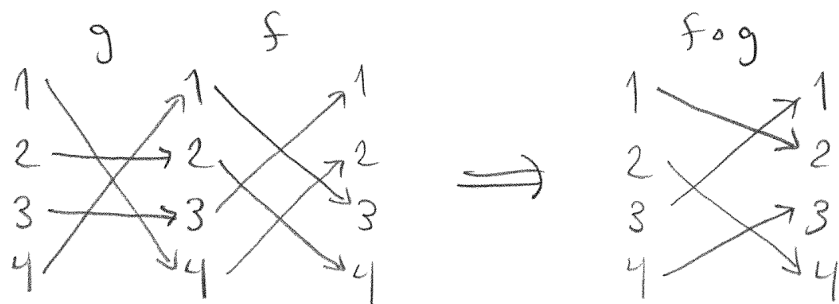
□

Example. Consider the permutations $f = 3412$ and $g = 4231$ in word notation, or $f = (13)(24)$ and $g = (14)(2)(3) = (14)$ in cycle notation. Compute $f \circ g$ and $g \circ f$.

Here is a picture showing that $g \circ f = 3124 = (1342)$:



And here is a picture showing that $f \circ g = 2413 = (1243)$:



It is important to note that $f \circ g \neq g \circ f$. In other words, the composition of permutations is not always “commutative.”

[Exercise: But sometimes it is. Check that the transpositions $(12) \in S_4$ and $(34) \in S_4$ commute with each other. More generally, any two “disjoint” cycles commute.]

///

Thus the set S_n is equipped with the binary operation of composition $\circ : S_n \times S_n \rightarrow S_n$, which is associative but not necessarily commutative. Furthermore, every element $f \in S_n$ has a compositional inverse $f^{-1} \in S_n$, and there exists a special element $\text{id} \in S_n$ satisfying $f \circ \text{id} = \text{id} \circ f = f$ for all $f \in S_n$. Galois used the word “group” to encapsulate these three properties. Here is the modern formulation.

Definition of Groups and Subgroups. Let G be a set equipped with an abstract binary operation $*$: $G \times G \rightarrow G$, which we will write as $(a, b) \mapsto a * b$. We say that the pair $(G, *)$ is a *group* if the following three¹ axioms hold:

(G0) *Substitution.* For all $a, b, c \in G$ we have that

$$a = b \quad \text{implies} \quad a * c = b * c \quad \text{and} \quad c * a = c * b.$$

(G1) The operation $*$ is *associative*. In other words, we have

$$a * (b * c) = (a * b) * c \quad \text{for all } a, b, c \in G.$$

(G2) There exists a *two-sided identity element* $\varepsilon \in G$ satisfying

$$a * \varepsilon = \varepsilon * a = a \quad \text{for all } a \in G.$$

(G3) For each $a \in G$ there exists a *two-sided inverse* $a^{-1} \in G$ satisfying

$$a * a^{-1} = a^{-1} * a = \varepsilon.$$

Note that we do not require the operation $*$ to be commutative. If it **is** commutative (i.e., if $a * b = b * a$ for all $a, b \in G$), then we say that the group is *abelian* (after Niels Henrik Abel).

Now let $H \subseteq G$ be any subset. We say that H is a *subgroup* if the following properties hold:

- For all $a, b \in H$ we have $a * b \in H$.
- The identity ε is in H .
- For all $a \in H$, the inverse a^{-1} is in H .

[Exercise: Actually there is a shorter definition. Prove that $H \subseteq G$ is a subgroup if and only if for all $a, b \in H$ we have $a * b^{-1} \in H$. Now there's only one property to check instead of three.] In other words: A subgroup is a subset that is also a group with respect to the same operation $*$ and identity ε .

///

Remarks:

- On the homework you will show that the identity element ε is unique. In other words, if there exist two elements $\varepsilon, \varepsilon'$ satisfying axiom (G2) then we must have $\varepsilon = \varepsilon'$. This is why we are allowed to talk about “the” identity element of the group.
- You will also show that any two inverses for $a \in G$ must be equal, therefore we are allowed to talk about “the” inverse of the element $a \in G$, and refer to it with the special notation a^{-1} . On the homework you will generalize this notation to define a^n for all $n \in \mathbb{Z}$.

¹Some authors think that axiom (G0) is unnecessary because it follows from general logical principles. I'm not so sure about that.

What led Galois to the definition of “groups?” Recall from Lagrange’s method that certain functions $f : \mathbb{C}^n \rightarrow \mathbb{C}$ are “invariant” or “symmetric” under certain permutations of their input. Here is the key fact that Galois noticed.

Exercise: Let $f : \mathbb{C}^n \rightarrow \mathbb{C}$ be any function with n inputs and 1 output and define $H \subseteq S_n$ as the subset of permutations of the input that leave f invariant. Then in fact H is a subgroup of S_n .

And here is Galois’ theorem with the technical conditions filled in.

Galois’ Theorem Again. The general n -th degree equation is solvable by radicals if and only if there exists a chain of subgroups

$$S_n = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_k = \{\text{id}\}$$

satisfying the following condition:

for each pair $G_i \supseteq G_{i+1}$ the quotient group G_i/G_{i+1} exists and is abelian.

///

We already know what “abelian” means (it means “commutative”) but it will take a while before we get to “quotient groups.”



Last time I defined abstract groups. Here are some key examples.

Example: The Symmetric Group. For all integers $n \geq 1$, the structure (S_n, \circ, id) is a group. It is called the *symmetric group* on n letters.

Why “the symmetric group?” This comes from the fact that Galois considered functions $f : \mathbb{C}^n \rightarrow \mathbb{C}$ that were “symmetric” under certain permutations of their input. Today we have a much broader view of symmetry. Here is a geometric example.

Example: The Icosahedral Group. Consider a regular icosahedron living in \mathbb{R}^3 . By a “symmetry” of the icosahedron we mean any **distance preserving** function $\mathbb{R}^3 \rightarrow \mathbb{R}^3$ that sends points of the icosahedron to points of the icosahedron. For example, let

- $f =$ rotation by $2\pi/5$ around the line through two opposite vertices,
- $g =$ rotation by $2\pi/3$ around the line through two opposite triangles.

It is immediate that the composition of two symmetries is again a symmetry, even though it might not be obvious how to describe $f \circ g$ and $g \circ f$ geometrically. It is also immediate that f^{-1} and g^{-1} are symmetries obtained by rotating in the opposite direction, and it is true (though harder to show) that any symmetry can be inverted. Finally, we note that the identity function $\text{id} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is a symmetry. Thus if G is the set of all icosahedral symmetries then we conclude that (G, \circ, id) is a group.

[Remark: We will learn later that this group has 120 elements, and is closely related to the unsolvability of the quintic.]

Wherever we find an associative binary operation there is probably a group. The prototype is functional composition, but there are also more basic examples.

Additive Groups. Consider the chain of integers, rational numbers, real numbers and complex numbers, each of which contains the number zero:

$$0 \in \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}.$$

Each of these sets is closed under the associate and commutative operation of “addition,” and therefore we obtain four abelian groups:

$$\mathbb{Z}^+ = (\mathbb{Z}, +, 0), \quad \mathbb{Q}^+ = (\mathbb{Q}, +, 0) \quad \mathbb{R}^+ = (\mathbb{R}, +, 0), \quad \mathbb{C}^+ = (\mathbb{C}, +, 0).$$

The set $\mathbb{N} = \{0, 1, 2, \dots\}$ of natural numbers is also closed under addition and contains the additive identity 0. However, the set \mathbb{N} does not contain “additive inverses,” so the structure $(\mathbb{N}, +, 0)$ is not a group. (You can call it a “monoid” or a “semigroup” if you want.)

Additive groups are always abelian.

Multiplicative Groups. Multiplication is also an associative operation, whether of numbers or matrices. Multiplication of numbers is generally commutative, whereas multiplication of matrices is generally not.

You may have heard that the structures $(\mathbb{Q}, +, \times, 0, 1)$, $(\mathbb{R}, +, \times, 0, 1)$ and $(\mathbb{C}, +, \times, 0, 1)$ are called *fields*, meaning that every non-zero element α has a multiplicative inverse $\alpha^{-1} = 1/\alpha$. If we delete zero then we obtain three (abelian) groups:

$$\mathbb{Q}^\times = (\mathbb{Q} - \{0\}, \times, 1), \quad \mathbb{R}^\times = (\mathbb{R} - \{0\}, \times, 1), \quad \mathbb{C}^\times = (\mathbb{C} - \{0\}, \times, 1).$$

It’s harder to squeeze a multiplicative group out of the *ring*² $(\mathbb{Z}, +, \times, 0, 1)$ because most of its elements do not have multiplicative inverses. For example, the equation $2x = 1$ has no integer solution $x \in \mathbb{Z}$. In fact, the only invertible integers are ± 1 . Thus we obtain a very small (abelian) group with only two elements:

$$\mathbb{Z}^\times = (\{+1, -1\}, \times, +1).$$

²Never mind the formal definition right now. I’m sure you can guess the important details.

On the other extreme, suppose that $(R, +, \times, 0, 1)$ is any ring. Then the set of $n \times n$ matrices $\text{Mat}_n(R)$ with entries from R is closed under matrix multiplication, which is an associative operation. (Reason: Because matrix multiplication is the composition of linear functions in disguise.) If $GL_n(R) \subseteq \text{Mat}_n(R)$ is the subset of all invertible matrices then we obtain the *general linear group*

$$(GL_n(R), \times, I),$$

where I is the $n \times n$ identity matrix:

$$I = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}.$$

The group $GL_n(R)$ is never abelian unless $n = 1$. In the case $n = 1$ we use a special notation

$$R^\times := GL_1(R),$$

and we call this the *group of units* of the ring R . You have seen four examples above.

Problem Set 1

1. An Example Cubic. Consider the cubic equation

$$x^3 - 6x - 6 = 0.$$

- (a) Apply Cardano's formula to find **one specific root** of the equation.
- (b) Now apply Lagrange's method to find **all three roots**. [Hint: Follow the steps in the course notes. There will be a lot of simplification.]

(a) We let $p = -6$ and $q = -6$ so our equation has the form $x^3 + px + q = 0$. Then Cardano's formula says

$$\begin{aligned} x &= \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ &= \sqrt[3]{3 + \sqrt{9 - 8}} + \sqrt[3]{3 - \sqrt{9 - 8}} \\ &= \sqrt[3]{3 - 1} + \sqrt[3]{3 + 1} \\ &= \sqrt[3]{2} + \sqrt[3]{4}. \end{aligned}$$

Let's choose the real cube roots $\sqrt[3]{2} \approx 1.26$ and $\sqrt[3]{4} \approx 1.59$, so that

$$x \approx 2.85.$$

One can check that this is indeed a root.

(b) To use Lagrange's method we let $e_1 = 0$, $e_2 = 6$ and $e_3 = -6$, so our equation has the form $x^3 - e_1x^2 + e_2x - e_3 = 0$. Then to find all three roots r_1, r_2, r_3 we let $\omega = e^{2\pi i/3}$ and make the change of variables

$$\begin{cases} s_1 = r_1 + r_2 + r_3 \\ s_2 = r_1 + \omega r_2 + \omega^2 r_3 \\ s_3 = r_1 + \omega^2 r_2 + \omega r_3 \end{cases} \iff \begin{cases} r_1 = (s_1 + s_2 + s_3)/3 \\ r_2 = (s_1 + \omega^2 s_2 + \omega s_3)/3 \\ r_3 = (s_1 + \omega s_2 + \omega^2 s_3)/3 \end{cases}$$

It is immediate that $s_1 = e_1 = 0$. Then from the course notes we see that

$$s_2 \cdot s_3 = e_1^2 - 3e_2 = 0 - 3(-6) = 18$$

and

$$s_2^3 + s_3^3 = 2e_1^3 - 9e_1e_2 + 27e_3 = 0 - 0 + 27(6) = 162.$$

To solve for s_2 and s_3 we first consider the quadratic polynomial with roots s_2^3 and s_3^3 :

$$\begin{aligned} (y - s_2^3)(y - s_3^3) &= 0 \\ y^2 - (s_2^3 + s_3^3)y + (s_2s_3)^3 &= 0 \\ y^2 - 162y + 18^3 &= 0 \\ y^2 - 162y + 5832 &= 0 \end{aligned}$$

Then the quadratic formula tells us that

$$\begin{aligned} s_2^3, s_3^3 &= \frac{1}{2} \left(162 \pm \sqrt{162^2 - 4 \cdot 18^3} \right) \\ &= \frac{1}{2} (162 \pm 54) \\ &= 54, 108. \end{aligned}$$

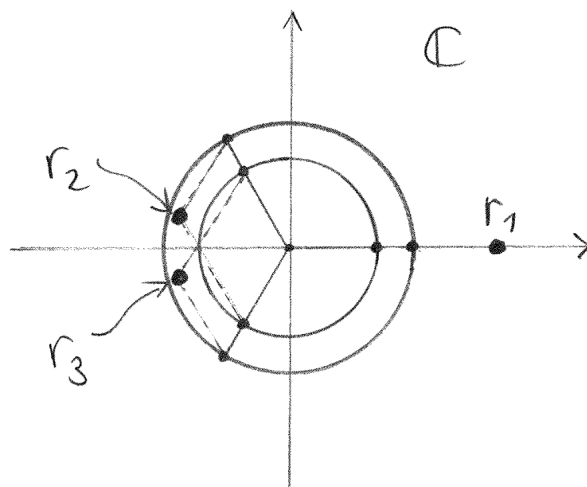
By breaking the symmetry, let us say that $s_2^3 = 54$ and $s_3^3 = 108$ so that

$$s_2 = \sqrt[3]{54} = 3 \cdot \sqrt[3]{2} \quad \text{and} \quad s_3 = \sqrt[3]{108} = 3 \cdot \sqrt[3]{4}.$$

Finally, we obtain

$$\begin{cases} r_1 = (s_1 + s_2 + s_3)/3 &= \sqrt[3]{2} + \sqrt[3]{4} \\ r_2 = (s_1 + \omega^2 s_2 + \omega s_3)/3 &= \omega^2 \cdot \sqrt[3]{2} + \omega \cdot \sqrt[3]{4} \\ r_3 = (s_1 + \omega s_2 + \omega^2 s_3)/3 &= \omega \cdot \sqrt[3]{2} + \omega^2 \cdot \sqrt[3]{4} \end{cases}$$

Here is a picture of the three roots in the complex plane:



2. Working With Permutations. Let S_3 be the set of all permutations of the set $\{1, 2, 3\}$, i.e., all invertible functions

$$f : \{1, 2, 3\} \rightarrow \{1, 2, 3\}.$$

- (a) List all 6 elements of the set. [I recommend using cycle notation.]
 - (b) We can think of (S_3, \circ, id) as a group, where \circ is functional composition and id is the identity function. Write out the full 6×6 group table.
 - (c) Let S_n be the group of permutations of $\{1, 2, \dots, n\}$. An element of S_n is called a *transposition* if it switches two elements of the set and sends every other element to itself. We denote the transposition that switches $i \leftrightarrow j$ by $(i, j) \in S_n$. Prove that every element of S_n can be expressed as a composition of transpositions.
 - (d) Let $A_n \subseteq S_n$ be the subset of permutations that can be expressed as a composition of an **even number** of transpositions. Prove that $A_n \subseteq S_n$ is a subgroup.
 - (e) List all elements of the subgroup $A_3 \subseteq S_3$ and draw its group table.
- (a) Here are the six permutations of $\{1, 2, 3\}$ in word notation and cycle notation:

word notation	cycle notation
123	ε
132	(23)
213	(12)
231	(123)
312	(132)
321	(13)

(b) Here is the group table:

\circ	ε	(12)	(13)	(23)	(123)	(132)
ε	ε	(12)	(13)	(23)	(123)	(132)
(12)	(12)	ε	(132)	(123)	(23)	(13)
(13)	(13)	(123)	ε	(132)	(12)	(23)
(23)	(23)	(132)	(123)	ε	(13)	(12)
(123)	(123)	(13)	(23)	(12)	(132)	ε
(132)	(132)	(23)	(12)	(13)	ε	(123)

(c) By the notation $(i_1, i_2, \dots, i_k) \in S_n$, I mean the permutation that sends i_j to i_{j+1} for all $1 \leq j < k$, sends i_k to i_1 , and sends every other element of $\{1, 2, \dots, n\}$ to itself. We call this kind of permutation a *k-cycle*. [Example: Transpositions are 2-cycles.] The cycle notation tells us that every element of S_n can be expressed as a composition of (commuting) cycles. Thus we will be done if we can show that every cycle is a composition of transpositions.

Here is the proof:

$$(i_1, i_2, \dots, i_k) = (i_1, i_2) \circ (i_2, i_3) \circ \dots \circ (i_{k-1}, i_k).$$

[Example: The permutation $f = 615432$ in word notation can be expressed as $f = (162)(35) = (162) \circ (35)$ in cycle notation, hence we have $f = (16) \circ (62) \circ (35)$.]

(d) Let $A_n \subseteq S_n$ be the subset consisting of permutations which can be expressed as a composition of an **even number** of transpositions. I claim that this is a subgroup. *Proof.*

- **Closure.** Suppose that $f, g \in A_n$. Then by definition we can write

$$f = s_1 \circ s_2 \circ \dots \circ s_k \quad \text{and} \quad g = t_1 \circ t_2 \circ \dots \circ t_\ell,$$

for some transpositions s_i and t_i , where k, ℓ are even numbers. But then

$$f \circ g = s_1 \circ s_2 \circ \dots \circ s_k \circ t_1 \circ t_2 \circ \dots \circ t_\ell$$

is a composition of $k + \ell$ transpositions, where $k + \ell$ is an even number.

- **Identity.** By convention we will say that the identity ε is a composition of zero transpositions. Since zero is an even number this means that $\varepsilon \in A_n$. If you don't buy that, let $t \in S_n$ be **any** transposition. Then we have

$$\varepsilon = t \circ t,$$

which is in A_n because 2 is an even number.

- **Inverses.** For any transposition $t \in S_n$ we have $t^2 = t \circ t = \varepsilon$ and hence $t^{-1} = t$. More generally, if $f = t_1 \circ t_2 \circ \cdots \circ t_k$ is any composition of transpositions then we have

$$f^{-1} = t_k \circ t_{k-1} \circ \cdots \circ t_2 \circ t_1.$$

It follows that $f \in A_n$ implies $f^{-1} \in A_n$. □

[Jargon: The subgroup $A_n \subseteq S_n$ is called the *alternating subgroup* of S_n .]

(e) Note that $(123) = (12) \circ (23)$ and $(132) = (12) \circ (13)$ are both in A_3 . It is a bit harder to check that the elements (12) , (13) , (23) are **not** in A_3 . *Check.* Let's write $c = (123)$ so that $c^2 = c^{-1} = (132)$. Now assume for contradiction that (12) **can** be expressed as a composition of evenly many transpositions:

$$(12) = (t_1 \circ t_2) \circ \cdots \circ (t_{2k-1} \circ t_{2k}).$$

But from the group table we see that any two transpositions compose to ε , $c = (123)$ or $c^{-1} = (132)$. This implies that (12) is a power of c . Contradiction. /// We conclude that

$$A_3 = \{\varepsilon, (123), (132)\}.$$

Here is the group table:

\circ	ε	(123)	(132)
ε	ε	(123)	(132)
(123)	(123)	(132)	ε
(132)	(132)	ε	(123)

[Exercise: In general we have $\#A_n = n!/2$. Can you prove this? It's possible to give a bijective proof right now but I prefer to wait until we can give a very slick proof.]

3. Working With Axioms. Let G be a set with a binary operation $(a, b) \mapsto a * b$. Consider the following four possible axioms:

- (G1) For all $a, b, c \in G$ we have $a * (b * c) = (a * b) * c$.
- (G2) There exists some $\varepsilon \in G$ such that $a * \varepsilon = \varepsilon * a = a$ for all $a \in G$.
- (G3) For each $a \in G$ there exists some $b \in G$ such that $a * b = b * a = \varepsilon$.
- (G4) For each $a \in G$ there exists some $c \in G$ such that $a * c = \varepsilon$.

The element ε in (G2) is called a *two-sided identity*. The element b in (G3) is called a *two-sided inverse* for a and the element c in (G4) is called a *right inverse* for a .

- (a) If (G1) and (G2) hold, prove that the two-sided identity element is unique.
- (b) If (G1), (G2) and (G3) hold, prove that the two-sided inverse is unique.

(c) Assuming that (G1) and (G2) hold, prove that (G3) and (G4) are equivalent. [Hint: One direction is obvious. The hard part is to prove that the existence of right inverses implies the existence of two-sided inverses.]

(a) *Proof.* Assume that (G1) and (G2) hold and suppose that the elements $\varepsilon, \varepsilon' \in G$ both satisfy (G2). Then we have

$$\varepsilon = \varepsilon * \varepsilon' = \varepsilon'.$$

[Remark: Actually I didn't need to use (G1).]

(b) *Proof.* Assume that (G1), (G2) and (G3) hold and suppose that the elements $b, b' \in G$ both satisfy (G3). Then we have

$$b = b * \text{id} = b * (a * b') = (b * a) * b' = \text{id} * b' = b'.$$

(c) *Proof.* Assume that (G1) and (G2) hold. Then (G3) clearly implies (G4). On the other hand, suppose that (G4) holds. Then for all $a \in G$ there exists some $c \in G$ such that $a * c = \varepsilon$. But we can also apply (G4) to this c to obtain some $d \in G$ such that $c * d = \varepsilon$. Putting these together gives

$$d = \text{id} * d = (a * c) * d = a * (c * d) = a * \text{id} = a,$$

so that $c * d = c * a = \varepsilon$ and hence c is a two-sided inverse for a . Finally, since $a \in G$ was arbitrary we conclude that (G3) holds.

4. Groups of Matrices. Matrix multiplication is associative because it corresponds to composition of linear functions. You may recall from linear algebra that a real $n \times n$ matrix $A \in \text{Mat}_n(\mathbb{R})$ has a (unique) two-sided inverse precisely when $\det A \neq 0$. Now consider the following sets of matrices:

$$GL_n(\mathbb{R}) = \{A \in \text{Mat}_n(\mathbb{R}) : \det A \neq 0\},$$

$$SL_n(\mathbb{R}) = \{A \in \text{Mat}_n(\mathbb{R}) : \det A = 1\},$$

$$O_n(\mathbb{R}) = \{A \in \text{Mat}_n(\mathbb{R}) : AA^T = I\},$$

$$SO_n(\mathbb{R}) = \{A \in \text{Mat}_n(\mathbb{R}) : AA^T = I \text{ and } \det A = 1\}.$$

Prove that each one of these sets is a group under matrix multiplication. [Hint: It is helpful to remember that $\det(AB) = \det(A)\det(B)$, $\det(A^T) = \det(A)$ and $(AB)^T = B^T A^T$ for all matrices $A, B \in \text{Mat}_n(\mathbb{R})$.]

(a) **General Linear Group.** Recall that matrix multiplication is associative because it represents the composition of linear functions, and recall that the $n \times n$ identity matrix I satisfies $AI = IA = A$ for all $A \in \text{Mat}_n(\mathbb{R})$.

- Suppose that $A, B \in GL_n(\mathbb{R})$ so that $\det(A), \det(B) \neq 0$. Then we have $\det(AB) = \det(A)\det(B) \neq 0$, which implies that $AB \in GL_n(\mathbb{R})$.
- Since $\det(I) = 1 \neq 0$ we have $I \in GL_n(\mathbb{R})$.
- If $A \in GL_n(\mathbb{R})$ then since $\det(A) \neq 0$ we know that the two-sided inverse A^{-1} exists, and since $\det(A^{-1}) = 1/\det(A) \neq 0$ we see that $A^{-1} \in GL_n(\mathbb{R})$.

(b) **Special Linear Group.** Clearly $SL_n(\mathbb{R})$ is a subset of $GL_n(\mathbb{R})$. We will show that it is a subgroup:

- Consider any $A, B \in SL_n(\mathbb{R})$ so that $\det(A) = \det(B) = 1$. Then we have $\det(AB^{-1}) = \det(A)/\det(B) = 1$, which implies that $AB^{-1} \in SL_n(\mathbb{R})$.

(c) **Orthogonal Group.** If $AA^T = I$ then we have

$$\begin{aligned}\det(AA^T) &= \det(I) \\ \det(A)\det(A^T) &= 1 \\ \det(A)^2 &= 1,\end{aligned}$$

which implies that $\det(A) = \pm 1$. In particular, we see that $O_n(\mathbb{R})$ is a subset of $GL_n(\mathbb{R})$. We will prove that it is a subgroup:

- Consider any $A, B \in O_n(\mathbb{R})$ so that $AA^T = BB^T = I$, or in other words we have $A^{-1} = A^T$ and $B^{-1} = B^T$. But then we have

$$(AB^{-1})(AB^{-1})^T = AB^{-1}(B^{-1})^T A^T = A(B^{-1}B)A^{-1} = AA^{-1} = I,$$

which implies that $AB^{-1} \in O_n(\mathbb{R})$.

(d) **Special Orthogonal Group.** It is easy to show that the intersection of subgroups is a subgroup. Since $SL_n(\mathbb{R})$ and $O_n(\mathbb{R})$ are both subgroups of $GL_n(\mathbb{R})$, and since

$$SO_n(\mathbb{R}) = SL_n(\mathbb{R}) \cap O_n(\mathbb{R}),$$

we conclude that $SO_n(\mathbb{R})$ is a subgroup of $GL_n(\mathbb{R})$.

5. Order of an Element. Let $(G, *, \varepsilon)$ be a group and let $g \in G$ be any element. Then for all integers $n \in \mathbb{Z}$ we define the exponential notation

$$g^n := \begin{cases} \overbrace{g * g * \cdots * g}^{n \text{ times}} & \text{if } n > 0, \\ \varepsilon & \text{if } n = 0, \\ \underbrace{g^{-1} * g^{-1} * \cdots * g^{-1}}_{-n \text{ times}} & \text{if } n < 0. \end{cases}$$

- (a) Check that $g^m * g^n = g^{m+n}$ for all $m, n \in \mathbb{Z}$.
- (b) Use this to prove that $\langle g \rangle := \{g^n : n \in \mathbb{Z}\}$ is a subgroup of G .
- (c) If $\langle g \rangle$ is a finite set, prove that there exists some $n \geq 1$ such that $g^n = \varepsilon$.
- (d) If $\langle g \rangle$ is finite, and if $m \geq 1$ is the smallest number such that $g^m = \varepsilon$, prove that

$$\#\langle g \rangle = m.$$

This m is called the *order* of the element $g \in G$. If the set $\langle g \rangle$ is infinite we will say that g has *infinite order*.

(a) This is a boring case-by-case check, depending on whether m and n are negative, zero or positive. It's okay to do it in your head.

(b) *Proof.* To see that $\langle g \rangle$ is closed under $*$, consider any two elements $g^m, g^n \in \langle g \rangle$. Then from part (a) we have $g^m * g^n = g^{m+n} \in \langle g \rangle$. To see that $\langle g \rangle$ contains the identity, note from part (a) that $\varepsilon = g^0 \in \langle g \rangle$. Finally, to see that $\langle g \rangle$ is closed under inversion, consider any element $g^n \in \langle g \rangle$. Then from part (a) we have $g^n * g^{-n} = g^{-n} * g^n = g^0 = \varepsilon$, and it follows that $(g^n)^{-1} = g^{-n} \in \langle g \rangle$ as desired. [Incidentally, since $m + n = n + m$ for all $m, n \in \mathbb{Z}$, part (a) also implies that the group $\langle g \rangle$ is **abelian**.]

(c) *Proof.* If $\langle g \rangle$ is finite then there must exist integers $k < \ell$ such that $g^k = g^\ell$. Now define $n := \ell - k \geq 1$ and observe that

$$\begin{aligned} g^\ell &= g^k \\ g^\ell * g^{-k} &= g^k * g^{-k} \\ g^{\ell-k} &= g^0 \\ g^n &= \varepsilon. \end{aligned}$$

(d) Let $m \geq 1$ be the smallest positive integer such that $g^m = \varepsilon$. Then I claim that

$$\langle g \rangle = \{\varepsilon, g, g^2, g^3, \dots, g^{m-1}\}.$$

To see that every element of $\langle g \rangle$ has this form, consider the element g^n for any integer $n \in \mathbb{Z}$. Now divide n by m to obtain $n = qm + r$, where the remainder satisfies $0 \leq r < m$, and observe that

$$g^n = g^{qm+r} = (g^m)^q * g^r = \varepsilon^q * g^r = g^r.$$

Finally, to see that the m elements $\varepsilon, g, \dots, g^{m-1}$ are distinct, let us suppose for contradiction that we have $g^k = g^\ell$ for some $0 \leq k < \ell \leq m-1$. But then as in (c) we have $g^{\ell-k} = \varepsilon$ which together with $1 \leq \ell - k \leq m-1$ contradicts the minimality of m . [Essentially, we have just proved that the remainder of $n \bmod m$ is unique.]

6. Matrices of Finite and Infinite Order. Consider the matrices

$$J = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad \text{for any } \theta \in \mathbb{R}.$$

- (a) Show that J is invertible and has infinite order.
- (b) Show that $R_\theta R_{-\theta} = I$, hence R_θ is invertible.
- (c) More generally, show that $R_\alpha R_\beta = R_{\alpha+\beta}$ for all angles $\alpha, \beta \in \mathbb{R}$.
- (d) Conclude that for each integer $n \geq 1$ the matrix $R_{2\pi/n}$ has order n .
- (e) For which angles θ does the matrix R_θ have infinite order?

(a) *Proof.* To show this, we will prove by induction that

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \quad \text{for all } n \in \mathbb{Z}.$$

Indeed, the statement is true when $n = 0$ and $n = 1$. And if the statement is true for n then it's also true for $n + 1$ because

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{n+1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^n \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & n+1 \\ 0 & 1 \end{pmatrix}.$$

We have shown that the statement is true for all $n \geq 0$. Finally, we observe for all $n \geq 1$ that

$$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -n \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which implies that

$$\begin{pmatrix} 1 & -n \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}^{-1} = \left[\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^n \right]^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{-n}.$$

(b) First observe that

$$R_{-\theta} = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}.$$

Then we have

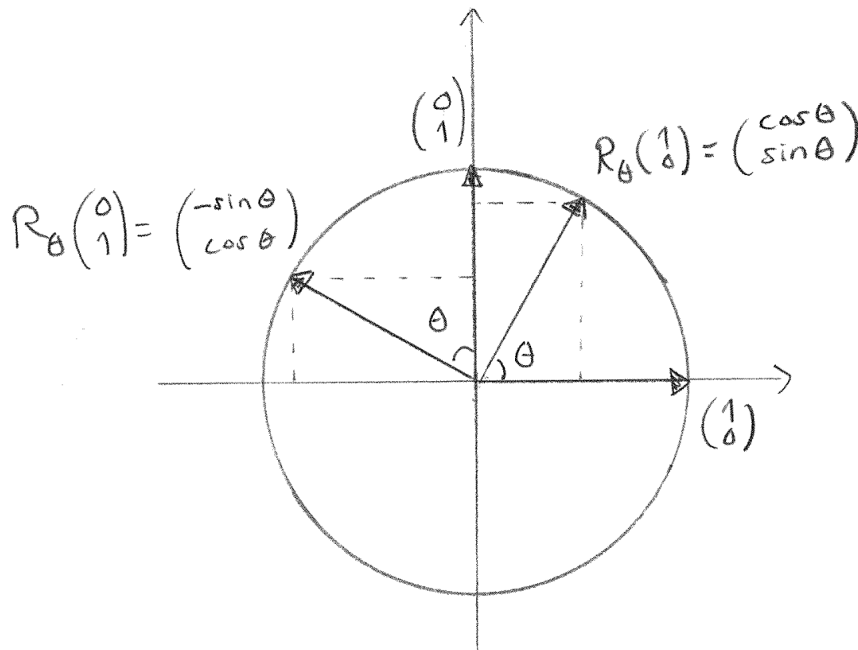
$$\begin{aligned} R_\theta R_{-\theta} &= \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \\ &= \begin{pmatrix} \cos^2 \theta + \sin^2 \theta & \cos \theta \sin \theta - \sin \theta \cos \theta \\ \sin \theta \cos \theta - \cos \theta \sin \theta & \sin^2 \theta + \cos^2 \theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

(c) The bad way to prove this is to expand out the product $R_\alpha R_\beta$ and use trig identities. The good way is to argue that the matrix R_θ represents the (linear) function $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ that rotates every vector counterclockwise by angle θ . Then since matrix multiplication is composition of linear functions we will conclude that

$$\begin{aligned} R_\alpha R_\beta &= R_\alpha \circ R_\beta \\ &= (\text{rotate ccw by } \alpha) \circ (\text{rotate ccw by } \beta) \\ &= (\text{first rotate ccw by } \beta \text{ then rotate ccw by } \alpha) \\ &= (\text{rotate ccw by } \alpha + \beta) \\ &= R_{\alpha+\beta}. \end{aligned}$$

So here's the proof that R_θ is a rotation.

Proof. Consider the standard basis vectors $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$. Here is a picture showing that R_θ rotates each of these vectors counterclockwise by angle θ :



Now let $f_\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the function that rotates every vector by counterclockwise around the origin by angle θ . We already know that $f_\theta(\mathbf{e}_1) = R_\theta \mathbf{e}_1$ and $f_\theta(\mathbf{e}_2) = R_\theta \mathbf{e}_2$, and we want to show that $f_\theta(\mathbf{x}) = R_\theta \mathbf{x}$ for all vectors $\mathbf{x} \in \mathbb{R}^2$. So consider any vector

$$\mathbf{x} = (\alpha, \beta) = (\alpha, 0) + (0, \beta) = \alpha(1, 0) + \beta(0, 1) = \alpha \mathbf{e}_1 + \beta \mathbf{e}_2.$$

Since the rotation function f_θ is linear (i.e., since it preserves parallelograms), we conclude that

$$f_\theta(\mathbf{x}) = f_\theta(\alpha \mathbf{e}_1 + \beta \mathbf{e}_2)$$

$$\begin{aligned}
&= \alpha f_\theta(\mathbf{e}_1) + \beta f_\theta(\mathbf{e}_2) \\
&= \alpha R_\theta \mathbf{e}_1 + \beta R_\theta \mathbf{e}_2 \\
&= R_\theta(\alpha \mathbf{e}_1 + \beta \mathbf{e}_2) \\
&= R_\theta \mathbf{x},
\end{aligned}$$

as desired.

(d) It follows from part (c) that $(R_\theta)^n = R_{n\theta}$ for any angle $\theta \in \mathbb{R}$ and for any integer $n \in \mathbb{Z}$. Thus we have

$$(R_{2\pi/n})^n = R_{2\pi} = R_0 = I.$$

Furthermore, since R_θ is rotation by θ we see that $R_\alpha = R_\beta$ if and only if $\alpha - \beta = 2\pi k$ for some integer $k \in \mathbb{Z}$. It follows from this that $(R_{2\pi/n})^k \neq I$ for all $1 \leq k \leq n - 1$.

(e) Good question. I guess that R_θ has infinite order if and only if $\theta = \alpha\pi$ for some **irrational** number $\alpha \in \mathbb{R}$. Otherwise, if $\theta = a\pi/b$ for some $a, b \in \mathbb{Z}$ then the order of R_θ is $\text{lcm}(a, 2b)/a$, where $\text{lcm}(a, 2b)$ is the least common multiple of a and $2b$.

Week 3

We have seen the definition of abstract groups and we have played with the main examples. It turns out that the three group axioms lead to an extraordinarily rich theory. The topic of “cyclic groups” will be our first glimpse of this theory.

Intersection of Subgroups is a Subgroup. Let $(G, *, \varepsilon)$ be a group and let $H_i \subseteq G$ be any family of subgroups (possibly infinite or even uncountable). Then the intersection

$$\bigcap_i H_i \subseteq G$$

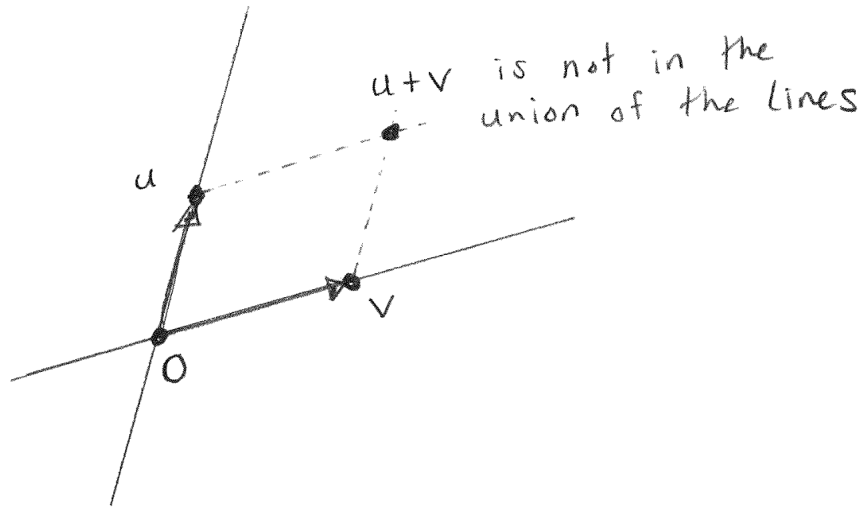
is also a subgroup.

Proof. Consider any elements a, b in the intersection. By definition this means that we have $a, b \in H_i$ for each index i . But then since H_i is a subgroup we must have $a * b^{-1} \in H_i$. Finally, since $a * b^{-1}$ is contained inside each subgroup H_i , it follows by definition that $a * b^{-1}$ is contained in the intersection. \square

However, the union of subgroups is not necessarily a subgroup. For example, consider the additive group $(\mathbb{R}^2, +, \mathbf{0})$ of vectors in n -dimensional Cartesian space and let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ be any two non-zero vectors satisfying $\mathbf{u} \neq \mathbf{v}$. Then each of the “lines”

$$\mathbb{R}\mathbf{u} := \{\alpha\mathbf{u} : \alpha \in \mathbb{R}\} \quad \text{and} \quad \mathbb{R}\mathbf{v} := \{\alpha\mathbf{v} : \alpha \in \mathbb{R}\}$$

is a subgroup of \mathbb{R}^n , but the union $\mathbb{R}\mathbf{u} \cup \mathbb{R}\mathbf{v}$ is not a subgroup because, for example, it does not contain the point $\mathbf{u} + \mathbf{v}$:



In linear algebra we fix this problem by defining the “linear span” of the vectors:

$$\mathbb{R}\mathbf{u} \cup \mathbb{R}\mathbf{v} \subsetneq \mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v} := \{\alpha\mathbf{u} + \beta\mathbf{v} : \alpha, \beta \in \mathbb{R}\},$$

and we call this the “plane” generated by \mathbf{u} and \mathbf{v} . This is a special case of a very general construction.

Subgroup Generated by a Subset. Let $(G, *, \varepsilon)$ be a group and let $S \subseteq G$ be any subset. Let $X = \{H : S \subseteq H\}$ be the set of all subgroups of G that contain the set S and consider the intersection

$$\langle S \rangle := \bigcap_{H \in X} H.$$

I claim that this $\langle S \rangle \subseteq G$ is the smallest subgroup of G that contains the set S . We call it the *subgroup of G generated by S* .

Proof. Since $S \subseteq H$ for all $H \in X$ we have by definition that S is contained in the intersection $\langle S \rangle$. Furthermore, we know from above that this intersection is a subgroup of G . Thus $\langle S \rangle$ is a subgroup of G that contains S . To see that this is the **smallest** such subgroup, let $K \subseteq G$ be **any** subgroup that contains S . By definition this means that $K \in X$, and hence

$$\langle S \rangle = \bigcap_{H \in X} H = K \cap \bigcap_{\substack{H \in X \\ H \neq K}} H \subseteq K,$$

as desired. □

Let’s examine the previous example in light of this definition.

Example: The Join of Two Subgroups. Let G any group and let $H, K \subseteq G$ be any two subgroups. We define their *join* as the smallest subgroup containing the union:

$$H \vee K := \langle H \cup K \rangle.$$

In terms of the group $(\mathbb{R}^n, +, \mathbf{0})$, I claim that the plane $\mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v}$ coincides with the join of the two lines $\mathbb{R}\mathbf{u}$ and $\mathbb{R}\mathbf{v}$.

Proof. The set inclusions $\mathbb{R}\mathbf{u} \subseteq \mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v}$ and $\mathbb{R}\mathbf{v} \subseteq \mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v}$ imply that the points $\alpha\mathbf{u}$ and $\beta\mathbf{v}$ are contained in $\mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v}$ for all $\alpha, \beta \in \mathbb{R}$. Then since the group $\mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v}$ is closed under addition we have $\alpha\mathbf{u} + \beta\mathbf{v} \in \mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v}$ and hence

$$\mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v} = \{\alpha\mathbf{u} + \beta\mathbf{v} : \alpha, \beta \in \mathbb{R}\} \subseteq \mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v}.$$

Conversely, one can check that the plane $\mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v}$ is a subgroup of \mathbb{R}^n . Then since the plane contains the union of the lines it must contain the subgroup **generated** by this union:

$$\mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v} = \langle \mathbb{R}\mathbf{u} \cup \mathbb{R}\mathbf{v} \rangle \subseteq \mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v}.$$

□

The key to this proof was the fact that $\mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v} \subseteq \mathbb{R}^n$ is a subgroup. On the homework you will show that a similar construction works for all abelian groups, but fails for non-abelian groups. Now let's examine the simplest possible case of a subgroup generated by a subset.

Definition of Cyclic Groups. Let $g \in G$ be an element in a group $(G, *, \varepsilon)$ and let $S = \{g\} \subseteq G$ be the subset containing just this element. Then we use the following notation for the group generated by S :

$$\langle g \rangle := \langle \{g\} \rangle = \langle S \rangle.$$

We call this the *cyclic subgroup generated by g* . In the special case that $G = \langle g \rangle$ for some element $g \in G$ we will say that G is a *cyclic group*.

Here's another point of view: On the homework you showed that for any group element $g \in G$ and for any integer $n \in \mathbb{Z}$ we can define the exponential notation $g^n \in G$ in such a way that

- $g^0 = \varepsilon$,
- $g^1 = g$,
- $(g^n)^{-1} = (g^{-1})^n = g^{-n}$ for all $n \in \mathbb{Z}$,
- and, more generally, $g^m * g^n = g^{m+n}$ for all $m, n \in \mathbb{Z}$.

It follows from this that the set of powers $\{g^n : n \in \mathbb{Z}\}$ is a subgroup of G . I claim that this group coincides with the cyclic group $\langle g \rangle$.

Proof. Since $g = g^1$ we see that the subgroup $\{g^n : n \in \mathbb{Z}\}$ contains the element g . Thus it must contain the smallest subgroup that contains g :

$$\langle g \rangle \subseteq \{g^n : n \in \mathbb{Z}\}.$$

On the other hand, we will prove by induction that every element g^n is contained in $\langle g \rangle$. Indeed, we have $g^1 = g \in \langle g \rangle$ by definition and we have $g^0 = \varepsilon \in \langle g \rangle$ since the subgroup $\langle g \rangle$ necessarily contains the identity. Furthermore, if $g^n \in \langle g \rangle$ for some $n \geq 1$ then since $\langle g \rangle$ is closed under $*$ we must have

$$g^{n+1} = g^n * g^1 \in \langle g \rangle.$$

Finally, since $\langle g \rangle$ is closed under inversion we conclude that

$$g^{-n} = (g^n)^{-1} \in \langle g \rangle \text{ for all } n \geq 1.$$

In summary, we conclude that $\{g^n : n \in \mathbb{Z}\} \subseteq \langle g \rangle$ as desired. \square

Last time I defined cyclic groups. Now some examples.

Example: \mathbb{Z}^+ is Cyclic. In an additive group $(G, +, 0)$ we prefer to write the inverse of $g \in G$ as $-g$ and we prefer to write the element g^n as $n \cdot g$, using the analogy that “multiplication is repeated addition.” To be precise, for each element $g \in G$ and each integer $n \in \mathbb{Z}$ we define

$$n \cdot g := \begin{cases} \underbrace{g + g + \cdots + g}_{n \text{ times}} & \text{if } n > 0, \\ 0 & \text{if } n = 0, \\ \underbrace{-g - g - \cdots - g}_{-n \text{ times}} & \text{if } n < 0. \end{cases}$$

When the group is $\mathbb{Z}^+ = (\mathbb{Z}, +, 0)$ this notation becomes completely literal:

$$\text{for all } k \in \mathbb{Z}^+ \text{ and } n \in \mathbb{Z} \text{ we have } n \cdot k = nk \in \mathbb{Z}^+.$$

It follows that the cyclic subgroup of \mathbb{Z}^+ generated by the element $k \in \mathbb{Z}^+$ is just the set of multiples of k . We have a special notation for this:

$$k\mathbb{Z} := \langle k \rangle = \{n \cdot k : n \in \mathbb{Z}\} = \{kn : n \in \mathbb{Z}\}.$$

Since every integer is a multiple of 1 (or -1) we conclude that the group \mathbb{Z}^+ is cyclic:

$$\mathbb{Z}^+ = \langle 1 \rangle = \langle -1 \rangle.$$

It will turn out later that \mathbb{Z}^+ is, in some sense, the only infinite cyclic group.

Example: Roots of Unity. Recall the “absolute value” of complex numbers,

$$| - | : \mathbb{C} \rightarrow \mathbb{R}_{\geq 0},$$

which is defined by $|a + ib| := a^2 + b^2$. Through some miracle it turns out that the absolute value respects multiplication. [Exercise: Check this.] It follows from this that the complex numbers of unit length form a subgroup of the multiplicative group \mathbb{C}^\times :

$$U(1) = \{\alpha \in \mathbb{C} : |\alpha| = 1\} \subseteq \mathbb{C}^\times = \{\alpha \in \mathbb{C} : \alpha \neq 0\}.$$

Since these numbers form a circle in the complex plane, we call $U(1)$ the *circle group*. Here's an interesting question:

Is the circle group a cyclic group?

Strictly speaking the answer is **no**. Indeed, the cyclic subgroup $\langle \omega \rangle \subseteq U(1)$ generated by any element $\omega \in U(1)$ is *countable*, but the number of points of the circle is *uncountable*. Let's examine these cyclic subgroups.

Recall that every unit length complex number has the form

$$\cos \theta + i \sin \theta = e^{i\theta} \quad \text{for some angle } 0 \leq \theta < 2\pi.$$

If $\omega = e^{2\pi i/n}$ for some integer $n \geq 1$ then we obtain a cyclic group of size n :

$$\langle \omega \rangle = \{1, \omega, \omega^2, \dots, \omega^{n-1}\} = \{1, e^{2\pi i/n}, e^{4\pi i/n}, \dots, e^{2\pi i(n-1)/n}\}.$$

This is the *group of n -th roots of unity*. [Exercise: Show that these are indeed all of the solutions of $x^n = 1$. The hard part is to show that there are no **other** solutions.] It is important to note that this group does not have a unique generator. On the homework you will show (indirectly) that the group of 12-th roots of unity has four possible generators

$$e^{2\pi i/12}, \quad e^{10\pi i/12}, \quad e^{14\pi i/12}, \quad e^{22\pi i/12}$$

which are called the *primitive 12-th roots of unity*.

If $\omega = e^{i\alpha\pi}$ for some **irrational** number $\alpha \in \mathbb{R}$ then one can show that the element ω has infinite order. This infinite set of powers $\langle \omega \rangle = \{\omega^n : n \in \mathbb{Z}\}$ does not coincide with the circle but it turns out that this set is *dense* in the circle. In other words, the circle group $U(1)$ is equal to the topological closure of this subgroup:

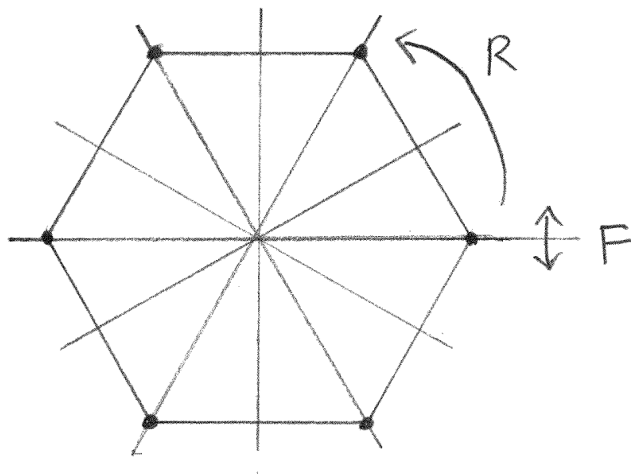
$$\overline{\langle \omega \rangle} = U(1).$$

So we might say the following:

The circle group is “almost cyclic.”

///

Example: Symmetries of a Regular Polygon. Consider a regular hexagon. In the following discussion we will show that this shape has exactly 12 symmetries, consisting of 6 rotation symmetries and 6 reflection symmetries:



6 rotations
&
6 reflections

We can think of a symmetry as a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ that leaves the hexagon looking the same. Thus, the symmetries can be combined by composition and they form a group called the *dihedral group* of size 12 (“dihedral” because the hexagon has two sides). If we let R denote any (primitive) rotation and let F denote any reflection then it turns out that the group can be generated by these two elements:

$$D_{12} = \langle R, F \rangle.$$

The dihedral group is **not** cyclic because neither of the generators can be expressed as a nontrivial power of the other.

It will take some time to prove these assertions but the proof will be very interesting. Here are the main steps:

- A “symmetry” of a regular polygon should preserve the distance between any two points and send the center of the polygon to itself.
- Any function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ that preserves distance and sends the origin to itself is a linear function.
- Any linear function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ has the form $f(\mathbf{x}) = A\mathbf{x}$ for some matrix $A \in \text{Mat}_2(\mathbb{R})$.
- If the linear function preserves distance then the matrix satisfies $A^T A = I$.
- Finally, any such matrix represents a rotation or a reflection.

I will prove some of this next week and you will prove the rest on the homework.

Week 4

What does “symmetry” mean in a geometric context?

Definition of Euclidean Space. Let \mathbb{R}^n the set of ordered n -tuples of real numbers, which we think of as *column vectors*:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n.$$

For any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we define the *standard inner product* as follows:

$$\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^T \mathbf{y} = (x_1 \ x_2 \ \cdots \ x_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

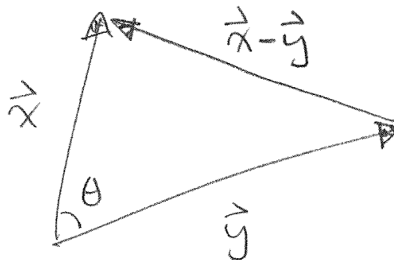
The *length* of a vector \mathbf{x} (i.e., the distance between the points \mathbf{x} and $\mathbf{0}$) is given by the extended Pythagorean theorem:

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

It follows from this that we can compute the *distance* between any two points:

$$(\text{distance between } \mathbf{x} \text{ and } \mathbf{y})^2 = \|\mathbf{x} - \mathbf{y}\|^2 = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle.$$

More surprisingly, we can use the inner product to compute the *angle* between any two vectors. To see this note that the vectors \mathbf{x} , \mathbf{y} and $\mathbf{x} - \mathbf{y}$ form three sides of a triangle:



By expanding the the expression $\|\mathbf{x} - \mathbf{y}\|^2 = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle$ in terms of algebra we get

$$\begin{aligned} \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle &= \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle - 2\langle \mathbf{x}, \mathbf{y} \rangle \\ \|\mathbf{x} - \mathbf{y}\|^2 &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\langle \mathbf{x}, \mathbf{y} \rangle. \end{aligned}$$

On the other hand the classical law of cosines tells us that

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta,$$

where θ is the angle between vectors \mathbf{x} and \mathbf{y} . Finally, combining the two equations gives

$$\begin{aligned} -2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta &= -2\langle\mathbf{x}, \mathbf{y}\rangle \\ \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta &= \langle\mathbf{x}, \mathbf{y}\rangle \\ \cos\theta &= \frac{\langle\mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{\langle\mathbf{x}, \mathbf{y}\rangle}{\sqrt{\langle\mathbf{x}, \mathbf{x}\rangle}\sqrt{\langle\mathbf{y}, \mathbf{y}\rangle}}. \end{aligned}$$

This implies that **any** geometric concept can be expressed in terms of the standard inner product on \mathbb{R}^n . To emphasize this situation we refer to the pair $(\mathbb{R}^n, \langle -, - \rangle)$ as *n-dimensional Euclidean space*. ///

Since the geometric structure of space is defined by the inner product $\langle -, - \rangle$, any function that preserves the inner product will preserve all geometric structure. Here is an infinite family of examples.

Orthogonal Matrices Preserve Geometry. On the previous homework we considered the “orthogonal group” of $n \times n$ orthogonal matrices:

$$O_n(\mathbb{R}) = \{A \in \text{Mat}_n(\mathbb{R}) : A^T A = I\}.$$

Recall that we can think of any $n \times n$ matrix A as a function \mathbb{R}^n to \mathbb{R}^n by multiplying column vectors on the left:

$$\mathbf{x} \mapsto A\mathbf{x}.$$

If the matrix satisfies $A^T A = I$ then for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we have

$$\langle A\mathbf{x}, A\mathbf{y} \rangle = (A\mathbf{x})^T (A\mathbf{y}) = (\mathbf{x}^T A^T)(A\mathbf{y}) = \mathbf{x}^T (A^T A)\mathbf{y} = \mathbf{x}^T I\mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle.$$

It follows that orthogonal matrices preserve all distances and angles. ///

The following surprising theorem shows that the converse is also true.

The Isometry Theorem. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be any function satisfying two conditions:

- f fixes the origin:

$$f(\mathbf{0}) = \mathbf{0}.$$

- f preserves distance (i.e., f is an isometry):

$$\text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \text{ we have } \|f(\mathbf{x}) - f(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|.$$

Then we have $f(\mathbf{x}) = A\mathbf{x}$ for some orthogonal matrix $A \in O_n(\mathbb{R})$.

Proof. See the homework. □

Example: Symmetries of a Regular Polygon (Continued). Now let's return to our discussion of a regular n -sided polygon in the Euclidean plane \mathbb{R}^2 . By a *symmetry* of the polygon I mean any function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ that

- sends points of the polygon to points of the polygon,
- preserves the distance between any two points, and
- sends the center of the polygon to itself.

For convenience, let's assume that the polygon is centered at the origin $\mathbf{0} \in \mathbb{R}^2$. Then the previous theorem implies that any symmetry has the form $f(\mathbf{x}) = A\mathbf{x}$ where A is a real 2×2 matrix satisfying $A^T A = I$. What are the possibilities for this matrix? Suppose that

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{for some } a, b, c, d \in \mathbb{R}.$$

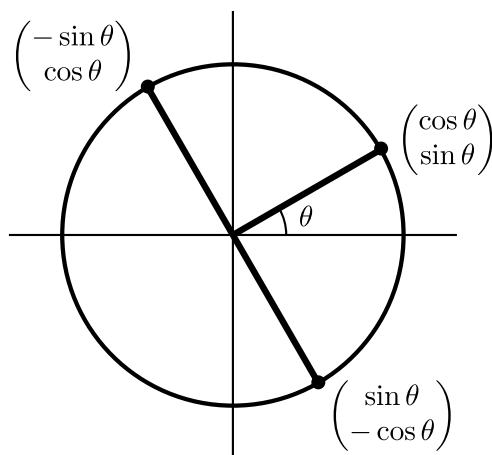
Then the equation $A^T A = I$ tells us that

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} (a \ c) \begin{pmatrix} a \\ c \end{pmatrix} & (a \ c) \begin{pmatrix} b \\ d \end{pmatrix} \\ (b \ d) \begin{pmatrix} a \\ c \end{pmatrix} & (b \ d) \begin{pmatrix} b \\ d \end{pmatrix} \end{pmatrix} = \begin{pmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{pmatrix}.$$

In other words, the two column vectors of A are **perpendicular unit vectors**. Since (a, c) is a unit vector we must have

$$(a, c) = (\cos \theta, \sin \theta) \quad \text{for some unique angle } 0 \leq \theta < 2\pi.$$

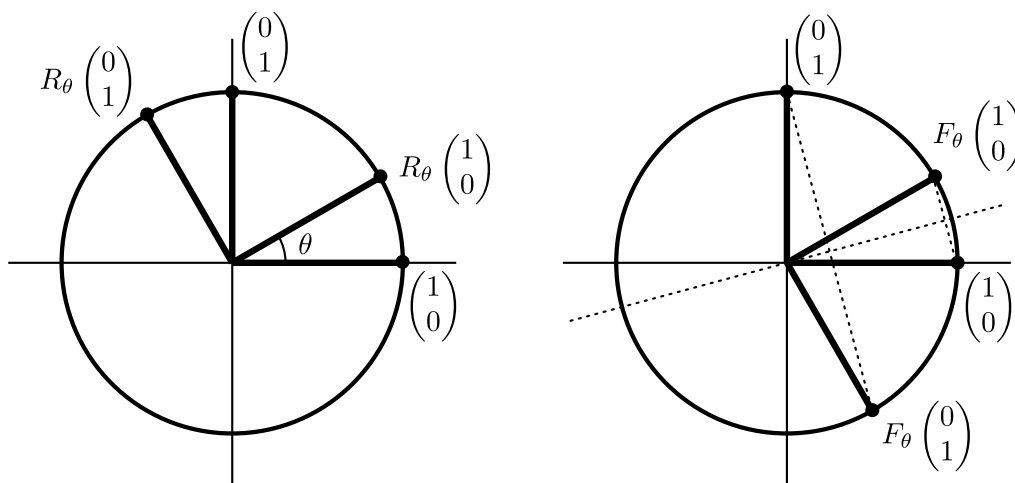
And then since (b, d) is a unit vector perpendicular to (a, c) , there are only two possibilities. Namely, we must have $(b, d) = (-\sin \theta, \cos \theta)$ or $(b, d) = (\sin \theta, -\cos \theta)$. Here is a picture:



In summary, we have shown that every 2×2 orthogonal matrix $A \in O_2(\mathbb{R})$ has one of the following two forms:

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad \text{or} \quad F_\theta = \begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix}.$$

We already know that $\mathbf{x} \mapsto R_\theta \mathbf{x}$ is the function that rotates the plane counterclockwise by angle θ . The following picture demonstrates that $\mathbf{x} \mapsto F_\theta \mathbf{x}$ is the function that **reflects** perpendicularly across the line that makes an angle of $\theta/2$ with the x -axis:



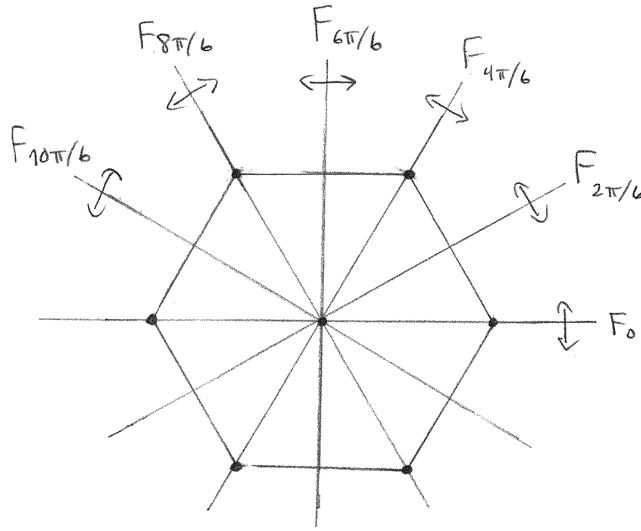
[Remark: R is for Rotation and F is for reFlection (or Flip).] At this point we know that the only possible symmetries of our regular n -gon are rotations and reflections. It turns out that there are n rotation symmetries and n reflection symmetries. The rotation symmetries form a cyclic group generated by $R := R_{2\pi/n}$:

$$\{I, R, R^2, \dots, R^{n-1}\} = \{R_0, R_{2\pi/n}, R_{4\pi/n}, \dots, R_{2\pi(n-1)/n}\}.$$

To determine the reflection symmetries we need to know the exact position of the n -gon. Let's assume for convenience that one of the vertices lies on the positive x -axis, so that F_0 (i.e., reflection across the x -axis) is a symmetry. Then the complete list of reflection symmetries is

$$\{F_0, F_{2\pi/n}, F_{4\pi/n}, \dots, F_{2\pi(n-1)/n}\}.$$

Here is the picture when $n = 6$:



The complete group of symmetries is called the *dihedral group* of size $2n$:

$$D_{2n} = \{R_0, R_{2\pi/n}, R_{4\pi/n}, \dots, R_{2\pi(n-1)/n}, F_0, F_{2\pi/n}, F_{4\pi/n}, \dots, F_{2\pi(n-1)/n}\}.$$

On the homework you will find a more efficient way to work with this group. Namely, if we define $R := R_{2\pi/n}$ and $F := F_0$ then you will show that the dihedral group is generated by these two elements as follows:

$$D_{2n} = \langle R, F \rangle = \{R^a F^b : a \in \{0, 1, \dots, n-1\} \text{ and } b \in \{0, 1\}\}.$$

///

You might have noticed that the n -th roots of unity and the rotation symmetries of a regular n -gon are really just “the same group” in two different disguises. Let’s formalize this idea.

Definition of Group Isomorphism. Let $(G, *)$ and (H, \bullet) be abstract groups. We will say that G and H are *isomorphic as groups*, and we will write

$$G \cong H,$$

if there exists a function $\varphi : G \rightarrow H$ satisfying the following properties:

- the function $\varphi : G \rightarrow H$ is invertible with inverse $\varphi^{-1} : H \rightarrow G$,
- for all $a, b \in G$ we have $\varphi(a * b) = \varphi(a) \bullet \varphi(b)$,

- for all $a, b \in H$ we have $\varphi^{-1}(a \bullet b) = \varphi^{-1}(a) * \varphi^{-1}(b)$.

Actually, you will show on the homework that the third condition follows automatically from the first two. When the three conditions are satisfied we say that the pair of functions φ, φ^{-1} defines an *isomorphism* between the groups $(G, *)$ and (H, \bullet) .

Example: Cyclic Groups. Let $(G, *, \varepsilon) = \langle g \rangle$ be a cyclic group. Recall from the first homework that if $\#G < \infty$ then there exists a **smallest positive integer** m such that $g^m = \varepsilon$, and it follows from this that

$$G = \{\varepsilon, g, g^2, \dots, g^{m-1}\}.$$

In this case I claim that

$$g^k = g^\ell \iff k - \ell \in m\mathbb{Z}.$$

Proof. If $k - \ell \in m\mathbb{Z}$ then by definition we have $k = \ell + mx$ for some $x \in \mathbb{Z}$ and hence

$$g^k = g^{\ell+mx} = g^\ell * g^{mx} = g^\ell * (g^m)^x = g^\ell * (\varepsilon)^x = g^\ell.$$

Conversely, let us suppose that $g^k = g^\ell$, and hence $g^{k-\ell} = \varepsilon$ for some $k, \ell \in \mathbb{Z}$. By computing the remainder of $k - \ell \bmod m$ we obtain

$$\begin{cases} k - \ell = qm + r, \\ 0 \leq r < m. \end{cases}$$

If $r \neq 0$ then we find that

$$g^r = g^{k-\ell-qm} = g^{k-\ell} * (g^m)^{-q} = \varepsilon * (\varepsilon)^{-q} = \varepsilon,$$

contradicting the minimality of m . Hence $k - \ell = qm + 0 \in m\mathbb{Z}$. ///

And if G is an **infinite** cyclic group then I claim that

$$g^k = g^\ell \iff k = \ell.$$

Proof. Clearly $k = \ell$ implies $g^k = g^\ell$. Conversely, suppose that we have $g^k = g^\ell$ for some $k \neq \ell$. Without loss let us assume that $k < \ell$. Then we have

$$\begin{aligned} g^\ell &= g^k \\ g^\ell * g^{-k} &= g^k * g^{-k} \\ g^{\ell-k} &= \varepsilon, \end{aligned}$$

for the positive integer $\ell - k$. If m is the **smallest** positive integer such that $g^m = \varepsilon$ then we again have $\#G = m$, which contradicts the fact that G is infinite. ///

With these facts in hand I can prove an important theorem about cyclic groups.

Theorem. Let $G = \langle g \rangle$ and $H = \langle h \rangle$ be cyclic groups. Then we have

$$G \cong H \iff \#G = \#H.$$

In other words, a cyclic group is determined up to isomorphism by its size.

Proof. Clearly $G \cong H$ implies $\#G = \#H$. Conversely, let us suppose that $\#G = \#H$. There are two cases. **(Case 1)** If the groups are finite then we have $\#G = \#H = m$ for some $m \geq 1$. Our goal is to define an isomorphism $\varphi : G \rightarrow H$ and there is an obvious candidate:

$$\text{let } \varphi(g^k) := h^k \text{ for all } k \in \mathbb{Z}.$$

There are four things to check:

- **Well-Defined.** Since the representation g^k is not unique we need to make sure that $g^k = g^\ell$ implies $\varphi(g^k) = \varphi(g^\ell)$. Indeed, from the above lemma we have

$$g^k = g^\ell \implies k - \ell \in m\mathbb{Z} \implies h^k = h^\ell \implies \varphi(g^k) = \varphi(g^\ell).$$

- **Surjective.** Every element of H has the form h^k for some $k \in \mathbb{Z}$ and hence has the form $\varphi(g^k)$ for some $g^k \in G$.

- **Injective.** We need to show that $\varphi(g^k) = \varphi(g^\ell)$ implies $g^k = g^\ell$. And, indeed,

$$\varphi(g^k) = \varphi(g^\ell) \implies h^k = h^\ell \implies k - \ell \in m\mathbb{Z} \implies g^k = g^\ell.$$

- **Homomorphism.** For all $k, \ell \in \mathbb{Z}$ we have

$$\varphi(g^k * g^\ell) = \varphi(g^{k+\ell}) = h^{k+\ell} = h^k \bullet h^\ell = \varphi(g^k) \bullet \varphi(g^\ell).$$

(Case 2) If $\#G = \#H = \infty$ then the proof is even easier because we don't need to check well-definedness. \square

It follows from this theorem that every infinite cyclic group is isomorphic to \mathbb{Z}^+ . We will see later that every cyclic group of size n is isomorphic to the quotient group $\mathbb{Z}/n\mathbb{Z}$, but before then I need to define quotient groups.

So much for cyclic groups. Now let's talk about the circle group.

Example: Euler's Isomorphism. This is as good a time as any for me to introduce *unitary matrices*. If $A \in \text{Mat}_n(\mathbb{R})$ is a real $n \times n$ matrix, recall that A^T denotes the *transpose* matrix. If $\langle -, - \rangle$ is the standard inner product on \mathbb{R}^n then the transpose matrix is defined by

$$\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A^T\mathbf{y} \rangle \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ in **complex** space we prefer to work with the "Hermitian" inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^* \mathbf{y} = \sum_i \bar{x}_i y_i,$$

where \mathbf{x}^* is the *conjugate transpose* row vector. More generally if $A \in \text{Mat}_n(\mathbb{C})$ is an $n \times n$ complex matrix then we let A^* denote the conjugate transpose of A . It is defined by the condition

$$\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A^*\mathbf{y} \rangle \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{C}^n.$$

Based on this, we define the (*special*) *orthogonal* and (*special*) *unitary* groups as follows:

$$\begin{aligned} O(n) &= \{A \in \text{Mat}_n(\mathbb{R}) : A^T A = I\}, \\ SO(n) &= \{A \in \text{Mat}_n(\mathbb{R}) : A^T A = I \text{ and } \det A = 1\}, \\ U(n) &= \{A \in \text{Mat}_n(\mathbb{C}) : A^* A = I\}, \\ SU(n) &= \{A \in \text{Mat}_n(\mathbb{C}) : A^* A = I \text{ and } \det A = 1\}, \end{aligned}$$

We have seen above that $O(n)$ and $SO(n)$ can be viewed as groups of symmetries of Euclidean space. The geometric meaning of $U(n)$ and $SU(n)$ is not so obvious but these groups are extremely important in physics. In general all four of these groups are distinct but for small values of n there can be “accidental isomorphisms.”

Theorem (Euler’s Isomorphism). We have $U(1) \cong SO(2)$.

[Remark: It is an amusing consequence of this theorem that $SO(2)$ is an **abelian** group, which is not obvious from the definition.]

Proof. Let $(\alpha) \in \text{Mat}_1(\mathbb{C})$ be a 1×1 complex matrix. Then the unitary condition says

$$\begin{aligned} (\alpha)^*(\alpha) &= (1) \\ (\alpha^*\alpha) &= (1) \\ (|\alpha|^2) &= (1), \end{aligned}$$

which implies that $|\alpha| = 1$. Euler showed that all such complex numbers have the form $e^{i\theta} = \cos \theta + i \sin \theta$. In other words, $U(1)$ is the familiar circle group:

$$U(1) = \{e^{i\theta} : \theta \in \mathbb{R}\}.$$

On the other hand, we proved above that any 2×2 real orthogonal matrix has the form R_θ (a rotation) or F_θ (a reflection). Since $\det R_\theta = 1$ and $\det F_\theta = -1$ for all $\theta \in \mathbb{R}$ we find that

$$SO(2) = \left\{ R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} : \theta \in \mathbb{R} \right\}.$$

I claim that we can define a group isomorphism $f : U(1) \rightarrow SO(2)$ by

$$\varphi(e^{i\theta}) := R_\theta \quad \text{for all } \theta \in \mathbb{R}.$$

There are four things to check:

- **Well-Defined.** For all $\eta, \theta \in \mathbb{R}$ we have

$$e^{i\eta} = e^{i\theta} \implies \eta - \theta \in 2\pi\mathbb{Z} \implies R_\eta = R_\theta \implies \varphi(e^{i\eta}) = \varphi(e^{i\theta}).$$

- **Surjective.** Every element of $SO(2)$ has the form R_θ for some $\theta \in \mathbb{R}$ and hence has the form $\varphi(e^{i\theta})$ for some $e^{i\theta} \in U(1)$.
- **Injective.** For all $\eta, \theta \in \mathbb{R}$ we have

$$\varphi(e^{i\eta}) = \varphi(e^{i\theta}) \implies R_\eta = R_\theta \implies \eta - \theta \in 2\pi\mathbb{Z} \implies e^{i\eta} = e^{i\theta}.$$

- **Homomorphism.** For all $\eta, \theta \in \mathbb{R}$ we have

$$\varphi(e^{i\eta}e^{i\theta}) = \varphi(e^{i(\eta+\theta)}) = R_{\eta+\theta} = R_\eta R_\theta = \varphi(e^{i\eta})\varphi(e^{i\theta}).$$

The third equality was proved on the previous homework.

□

[Remark: The name “Euler’s Isomorphism” is facetious.]

Note that this isomorphism restricts to an isomorphism between the n -th roots of unity under multiplication and the rotational symmetries of a regular n -gon under composition.

Problem Set 2

1. **Powers of a Cycle.** Consider the standard 12-cycle in cycle notation:

$$c := (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12) \in S_{12}.$$

Compute the first twelve powers $c, c^2, c^3, \dots, c^{12}$ and express each of them in cycle notation. Try to guess what the k -th power of an n -cycle looks like.

Solution. We have

$$\begin{aligned} c &= (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12) \\ c^2 &= (1, 3, 5, 7, 9, 11)(2, 4, 6, 8, 10, 12) \\ c^3 &= (1, 4, 7, 10)(2, 5, 8, 11)(3, 6, 9, 12) \\ c^4 &= (1, 5, 9)(2, 6, 10)(3, 7, 11)(4, 8, 12) \\ c^5 &= (1, 6, 11, 4, 9, 2, 7, 12, 5, 10, 3, 8) \\ c^6 &= (1, 7)(2, 8)(3, 9)(4, 10)(5, 11)(6, 12) \\ c^7 &= (1, 8, 3, 10, 5, 12, 7, 2, 9, 4, 11, 6) \\ c^8 &= (1, 9, 5)(2, 10, 6)(3, 11, 7)(4, 12, 8) \\ c^9 &= (1, 10, 7, 4)(2, 11, 8, 5)(3, 12, 9, 6) \end{aligned}$$

$$\begin{aligned}
c^{10} &= (1, 11, 9, 7, 5, 3)(2, 12, 10, 8, 6, 4) \\
c^{11} &= (1, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2) \\
c^{12} &= (1)(2)(3)(4)(5)(6)(7)(8)(9)(10)(11)(12) = \varepsilon.
\end{aligned}$$

Now let c be a general n -cycle. I guess that the k -th power c^k consists of d cycles each of length n/d , where $d = \gcd(n, k)$ is the greatest common divisor of n and k . I'll prove this later but not today.

2. Homomorphism and Isomorphism. Let $(G, *, \delta)$ and $(H, \bullet, \varepsilon)$ be abstract groups and let $\varphi : G \rightarrow H$ be a function. We say that φ is a (*group*) *homomorphism* if it satisfies the following condition:

$$\text{for all } a, b \in G \text{ we have } \varphi(a * b) = \varphi(a) \bullet \varphi(b).$$

- (a) If $\varphi : G \rightarrow H$ is a homomorphism, prove that $\varphi(\delta) = \varepsilon$.
- (b) If $\varphi : G \rightarrow H$ is a homomorphism, prove that $\varphi(a^{-1}) = \varphi(a)^{-1}$ for all $a \in G$.
- (c) Suppose that $\varphi : G \rightarrow H$ is homomorphism and that the inverse function exists. Prove that the function $\varphi^{-1} : H \rightarrow G$ is also a homomorphism. It follows that invertible homomorphisms are the same as isomorphisms.

[Remark: Instead of f I tend to use the letter φ for group homomorphisms.]

(a) *Proof.* Since $\delta * \delta = \delta$ we have

$$\begin{aligned}
\varphi(\delta * \delta) &= \varphi(\delta) \\
\varphi(\delta) \bullet \varphi(\delta) &= \varphi(\delta) \\
\varphi(\delta) \bullet \varphi(\delta) \bullet \varphi(\delta)^{-1} &= \varphi(\delta) \bullet \varphi(\delta)^{-1} \\
\varphi(\delta) &= \varepsilon.
\end{aligned}$$

(b) *Proof.* For all $a \in G$ we have

$$\begin{aligned}
a * a^{-1} &= \delta \\
\varphi(a * a^{-1}) &= \varphi(\delta) \\
\varphi(a) \bullet \varphi(a^{-1}) &= \varepsilon \\
\varphi(a)^{-1} \bullet \varphi(a) \bullet \varphi(a^{-1}) &= \varphi(a)^{-1} \bullet \varepsilon \\
\varphi(a^{-1}) &= \varphi(a)^{-1}.
\end{aligned}$$

(c) *Proof.* Assume that $\varphi : G \rightarrow H$ is an invertible homomorphism. Then for all $a, b \in H$ we have

$$\varphi(\varphi^{-1}(a) * \varphi^{-1}(b)) = \varphi(\varphi^{-1}(a)) \bullet \varphi(\varphi^{-1}(b)) = a \bullet b,$$

and applying φ^{-1} to both sides gives

$$\varphi^{-1}(a) * \varphi^{-1}(b) = \varphi^{-1}(a \bullet b)$$

as desired. □

3. Isometries = Orthogonal Matrices. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ be column vectors and let \mathbf{x}^T denote the row vector corresponding to \mathbf{x} . We define the standard inner product as follows:

$$\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^T \mathbf{y} = \sum_i x_i y_i.$$

Recall the the distance between two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined by $\|\mathbf{x} - \mathbf{y}\|^2 = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle$ and recall that the following properties are satisfied:

- We have $\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
- For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we have $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$.
- For all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$ we have $\langle \mathbf{x}, \alpha \mathbf{y} + \beta \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle + \beta \langle \mathbf{x}, \mathbf{z} \rangle$.

The goal of this problem is to show the following: If $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is any function that preserves distance and sends the origin to itself then it preserves the inner product. Hence the function is linear. Hence we have $f(\mathbf{x}) = A\mathbf{x}$ for some $n \times n$ matrix A , which must satisfy $A^T A = I$.

- (a) Assume that the function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ preserves the distance between any two points (i.e., $\|f(\mathbf{x}) - f(\mathbf{y})\|^2 = \|\mathbf{x} - \mathbf{y}\|^2$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$) and sends the origin to itself (i.e., $f(\mathbf{0}) = \mathbf{0}$). Prove that

$$\langle f(\mathbf{x}), f(\mathbf{y}) \rangle = \langle \mathbf{x}, \mathbf{y} \rangle \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

- (b) Continuing from part (a), prove that this f is a linear function. [Hint: For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ show that

$$\|f(\mathbf{x} + \mathbf{y}) - (f(\mathbf{x}) + f(\mathbf{y}))\|^2 = 0 \quad \text{and} \quad \|f(\alpha \mathbf{x}) - \alpha f(\mathbf{x})\|^2 = 0.]$$

- (c) Continuing from (a) and (b), show that $f(\mathbf{x}) = A\mathbf{x}$ for some $n \times n$ matrix satisfying $A^T A = I$. [Hint: Let $\mathbf{e}_1, \dots, \mathbf{e}_n \in \mathbb{R}^n$ be the standard basis vectors. Then $f(\mathbf{e}_i)$ is the i -th column of A . To show that $A^T A = I$ use the fact that $\mathbf{e}_i^T B \mathbf{e}_j$ is equal to the i, j -entry of an arbitrary matrix B .]

(a) *Proof.* Since $f(\mathbf{0}) = \mathbf{0}$ we have for all $\mathbf{x} \in \mathbb{R}^n$ that

$$\begin{aligned}\langle f(\mathbf{x}), f(\mathbf{x}) \rangle &= \|f(\mathbf{x})\|^2 \\ &= \|f(\mathbf{x}) - \mathbf{0}\|^2 \\ &= \|f(\mathbf{x}) - f(\mathbf{0})\|^2 = \|\mathbf{x} - \mathbf{0}\|^2 = \|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle.\end{aligned}$$

Then for any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we have

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

(b) *Proof.* Now consider any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. To show that $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$ we will verify that $f(\mathbf{x} + \mathbf{y}) - f(\mathbf{x}) - f(\mathbf{y})$ has length zero, hence it must be the zero vector. Here is the verification:

$$\begin{aligned}\|f(\mathbf{x} + \mathbf{y}) - f(\mathbf{x}) - f(\mathbf{y})\|^2 &= \langle f(\mathbf{x} + \mathbf{y}) - f(\mathbf{x}) - f(\mathbf{y}), f(\mathbf{x} + \mathbf{y}) - f(\mathbf{x}) - f(\mathbf{y}) \rangle \\ &= \langle f(\mathbf{x} + \mathbf{y}), f(\mathbf{x} + \mathbf{y}) \rangle + \langle f(\mathbf{x}), f(\mathbf{x}) \rangle + \langle f(\mathbf{y}), f(\mathbf{y}) \rangle \\ &\quad - 2\langle f(\mathbf{x} + \mathbf{y}), f(\mathbf{x}) \rangle - 2\langle f(\mathbf{x} + \mathbf{y}), f(\mathbf{y}) \rangle + 2\langle f(\mathbf{x}), f(\mathbf{y}) \rangle \\ &= \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle + \cancel{\langle \mathbf{x}, \mathbf{x} \rangle} + \cancel{\langle \mathbf{y}, \mathbf{y} \rangle} - 2\langle \mathbf{x} + \mathbf{y}, \mathbf{x} \rangle - 2\langle \mathbf{x} + \mathbf{y}, \mathbf{y} \rangle + 2\langle \mathbf{x}, \mathbf{y} \rangle \\ &= \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle - \langle \mathbf{x}, \mathbf{x} \rangle - \langle \mathbf{y}, \mathbf{y} \rangle - 2\langle \mathbf{y}, \mathbf{x} \rangle - \cancel{2\langle \mathbf{x}, \mathbf{y} \rangle} + \cancel{2\langle \mathbf{x}, \mathbf{y} \rangle} \\ &= \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle - \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle \\ &= 0.\end{aligned}$$

Then for any $\alpha \in \mathbb{R}$ we will verify that $f(\alpha\mathbf{x}) - \alpha f(\mathbf{x})$ has length zero, hence $f(\alpha\mathbf{x}) = \alpha f(\mathbf{x})$:

$$\begin{aligned}\|f(\alpha\mathbf{x}) - \alpha f(\mathbf{x})\|^2 &= \langle f(\alpha\mathbf{x}) - \alpha f(\mathbf{x}), f(\alpha\mathbf{x}) - \alpha f(\mathbf{x}) \rangle \\ &= \langle f(\alpha\mathbf{x}), f(\alpha\mathbf{x}) \rangle - 2\langle f(\alpha\mathbf{x}), \alpha f(\mathbf{x}) \rangle + \langle \alpha f(\mathbf{x}), \alpha f(\mathbf{x}) \rangle \\ &= \langle \alpha\mathbf{x}, \alpha\mathbf{x} \rangle - 2\alpha\langle f(\alpha\mathbf{x}), f(\mathbf{x}) \rangle + \alpha^2\langle f(\mathbf{x}), f(\mathbf{x}) \rangle \\ &= \alpha^2\langle \mathbf{x}, \mathbf{x} \rangle - 2\alpha\langle \alpha\mathbf{x}, \mathbf{x} \rangle + \alpha^2\langle \mathbf{x}, \mathbf{x} \rangle \\ &= \alpha^2\langle \mathbf{x}, \mathbf{x} \rangle - 2\alpha^2\langle \mathbf{x}, \mathbf{x} \rangle + \alpha^2\langle \mathbf{x}, \mathbf{x} \rangle \\ &= 0.\end{aligned}$$

(c) *Proof.* From (b) we know that $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a linear function. To show that f is a matrix, let $\mathbf{e}_1, \dots, \mathbf{e}_n \in \mathbb{R}^n$ be the standard basis and let A be the matrix whose i -th column is

$f(\mathbf{e}_i)$. If $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n \in \mathbb{R}^n$ is any vector then by definition of matrix multiplication we have

$$A\mathbf{x} = x_1f(\mathbf{e}_1) + \cdots + x_nf(\mathbf{e}_n) = f(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) = f(\mathbf{x}).$$

Finally, to show that $A^T A = I$, recall from part (a) that $\langle f(\mathbf{x}), f(\mathbf{y}) \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$ for all vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. We find that the i, j -entry of the matrix $A^T A$ is

$$\mathbf{e}_i^T (A^T A) \mathbf{e}_j = (A\mathbf{e}_i)^T (A\mathbf{e}_j) = \langle A\mathbf{e}_i, A\mathbf{e}_j \rangle = \langle f(\mathbf{e}_i), f(\mathbf{e}_j) \rangle = \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

In other words, $A^T A$ is the identity matrix. □

[Remark: The hardest part by far was to show that an origin-fixing isometry is linear. That's a pretty surprising fact.]

4. Rotation and Reflection. In class I showed that every element of $O_2(\mathbb{R})$ has the form

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad \text{or} \quad F_\theta = \begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix}.$$

- (a) Verify that $R_\theta \in SO_2(\mathbb{R})$ and that $F_\theta \in O_2(\mathbb{R}) - SO_2(\mathbb{R})$.
- (b) We saw on the previous homework that $\mathbf{x} \mapsto R_\theta \mathbf{x}$ is a rotation. Use a similar argument to prove that $\mathbf{x} \mapsto F_\theta \mathbf{x}$ is a reflection.
- (c) For all $\alpha, \beta \in \mathbb{R}$ prove that
- $R_\alpha R_\beta = R_{\alpha+\beta}$,
 - $F_\alpha F_\beta = R_{\alpha-\beta}$,
 - $R_\alpha F_\beta = F_\beta (R_\alpha)^{-1} = F_{\alpha+\beta}$.
- (d) Fix a positive integer n and define the matrices $R := R_{2\pi/n}$ and $F := F_0$. The subgroup of $O_2(\mathbb{R})$ generated by the set $\{R, F\}$ has $2n$ elements. Use (c) to find them all.

(a) Since $R_\theta^T = R_{-\theta}$, we already know from the first homework that $R_\theta^T R_\theta = I$. Now we need to check that $\det(R_\theta) = 1$. Indeed, we have

$$\det \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} = \cos^2 \theta - (-\sin^2 \theta) = 1.$$

Next we need to check that $F_\theta^T F_\theta = I$ and $\det(F_\theta) \neq 1$. For the first statement we have

$$\begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix} = \begin{pmatrix} \cos^2 \theta + \sin^2 \theta & \sin \theta \cos \theta - \sin \theta \cos \theta \\ \sin \theta \cos \theta - \sin \theta \cos \theta & \sin^2 \theta + \cos^2 \theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and for the second statement we have

$$\det \begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix} = -\cos^2 \theta - \sin^2 \theta = -1 \neq 1.$$

(b) Since $F_\theta^T = F_\theta$, we saw in part (a) that $F_\theta F_\theta = I$. [Jargon: We say that F_θ is an *involution*.] Now we will give a geometric reason for this. For any real number $0 \leq \theta < 2\pi$, let $f_\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the function that reflects orthogonally across the line that makes an angle of $\theta/2$ with the x -axis. I claim that $f_\theta(\mathbf{x}) = F_\theta \mathbf{x}$ for all vectors $\mathbf{x} \in \mathbb{R}^2$.

Proof. Consider the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2 \in \mathbb{R}^2$. We saw in the notes that $f_\theta(\mathbf{e}_1) = F_\theta \mathbf{e}_1$ and $f_\theta(\mathbf{e}_2) = F_\theta \mathbf{e}_2$. Now consider any vector $\mathbf{x} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 \in \mathbb{R}^2$. Then since reflection preserves linear combinations (i.e., it preserves parallelograms) we conclude that

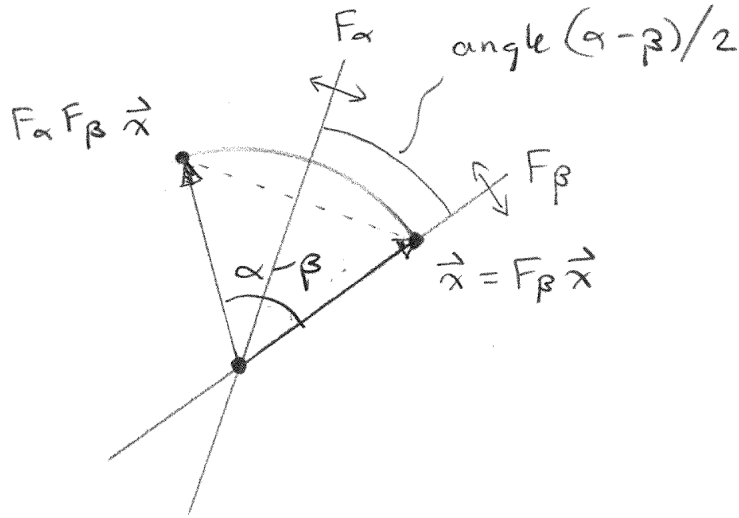
$$\begin{aligned} f_\theta(\mathbf{x}) &= f_\theta(x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2) \\ &= x_1 f_\theta(\mathbf{e}_1) + x_2 f_\theta(\mathbf{e}_2) \\ &= x_1 F_\theta \mathbf{e}_1 + x_2 F_\theta \mathbf{e}_2 \\ &= F_\theta(x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2) \\ &= F_\theta \mathbf{x}, \end{aligned}$$

as desired. □

(c) *Proof.* We saw on the first homework that $R_\alpha R_\beta = R_{\alpha+\beta}$. This follows from the fact that

$$(\text{rotate by } \beta \text{ and then rotate by } \alpha) = (\text{rotate by } \alpha + \beta).$$

One could prove $F_\alpha F_\beta = R_{\alpha-\beta}$ by quoting trig identities but I prefer to find a better way. Since $\det(F_\alpha) = \det(F_\beta) = -1$ we know that $\det(F_\alpha F_\beta) = \det(F_\alpha) \det(F_\beta) = 1$. Since we know that all 2×2 orthogonal matrices with determinant 1 have the form R_θ we conclude that $F_\alpha F_\beta = R_\theta$ for some θ . To compute the angle of rotation it is enough to apply $F_\alpha F_\beta$ to one specific vector and see what happens. Let's choose a vector on the reflecting line for F_β so that $F_\beta \mathbf{x} = \mathbf{x}$. Then we have the following picture:



Since the angle between the reflecting lines of F_α and F_β is $(\alpha - \beta)/2$, we conclude that the vector \mathbf{x} gets rotated by angle $\theta = \alpha - \beta$, i.e., **twice** the angle between the reflecting lines.

Finally, to prove that $R_\alpha F_\beta = F_{\alpha+\beta}$ we note that

$$\begin{aligned} F_{\alpha+\beta} F_\beta &= R_{\alpha+\beta-\beta} \\ F_{\alpha+\beta} F_\beta &= R_\alpha \\ F_{\alpha+\beta} \cancel{F_\beta} F_\beta &= R_\alpha F_\beta \\ F_{\alpha+\beta} &= R_\alpha F_\beta, \end{aligned}$$

and to prove that $F_\beta (R_\alpha)^{-1} = F_{\alpha+\beta}$ we note that

$$\begin{aligned} F_\beta F_{\alpha+\beta} &= R_{\beta-(\alpha+\beta)} \\ F_\beta F_{\alpha+\beta} &= R_{-\alpha} \\ \cancel{F_\beta} F_\beta F_{\alpha+\beta} &= F_\beta R_{-\alpha} \\ F_{\alpha+\beta} &= F_\beta (R_\alpha)^{-1}. \end{aligned}$$

□

(d) Fix $R := R_{2\pi/n}$ and $F := F_0$. Then I claim that

$$\langle R, F \rangle = \{R^a F^b : a \in \{0, 1, \dots, n-1\} \text{ and } b \in \{0, 1\}\}.$$

Proof. Since the subgroup $\langle R, F \rangle$ contains R and F and is closed under composition, we conclude that $\{R^a F^b\} \subseteq \langle R, F \rangle$. Conversely, we will show that $\{R^a F^b\}$ is a subgroup of $O_2(\mathbb{R})$ containing R and F , from which it will follow that $\langle R, F \rangle \subseteq \{R^a F^b\}$. Let's check:

- **Closure.** Since $FR = R^{-1}F$, one can prove by induction that $FR^c = R^{-c}F$ for all integers $c \in \mathbb{Z}$. Now consider any two elements R^aF^b and R^cF^d . There are two cases: **(Case 1)** If $b = 0$ then we have

$$(R^aF^b)(R^dF^d) = (R^a)(R^cF^d) = R^{a+c}F^d = R^kF^d,$$

where k is the remainder of $a + c \bmod n$. **(Case 2)** If $b = 1$ then we have

$$(R^aF^b)(R^cF^d) = R^a(FR^c)F^d = R^a(R^{-c}F)F^d = R^{a-c}F^{d+1} = R^kF^\ell,$$

where k is the remainder of $a - c \bmod n$ and ℓ is the remainder of $d + 1 \bmod 2$.

- **Identity.** The identity is $R^0F^0 = I$.
- **Inverses.** Consider any element R^aF^b . There are two cases: **(Case 1)** If $b = 0$ then $R^aF^b = R^a$ has inverse $R^{-a} = R^{n-a}F^0$, which has the correct form. **(Case 2)** If $b = 1$ then the element $R^aF^b = R^aF$ is equal to its own inverse:

$$(R^aF)(R^aF) = R^a(FR^a)F = R^a(R^{-a}F)F = (R^aR^{-a})(FF) = I.$$

Finally, let's prove that these $2n$ elements are distinct. If $R^aF^b = R^cF^d$ then multiplying on the left by R^{-c} and on the right by F^{-b} gives $R^{c-a} = F^{d-b}$. Since R^{c-a} always has determinant 1 this implies that $F^{b-d} = I$, and hence $F^b = F^d$. But then we must also have $R^{c-a} = I$ and hence $R^a = R^c$. \square

[Remark: It is also true that $\langle R, F \rangle = \{F^aR^b : a, b \in \mathbb{Z}\}$.]

5. The Fermat-Euler-Lagrange Theorem. Let $(G, *, \varepsilon)$ be a group and let $g \in G$ be any element. Define the function $f_g : G \rightarrow G$ by $f_g(a) := g * a$.

- Prove that $f_g : G \rightarrow G$ is a bijection.
- If G is a **finite abelian** group, prove that $g^{\#G} = \varepsilon$. [Hint: Suppose that $G = \{a_1, a_2, \dots, a_n\}$. Explain why $\prod_i a_i = \prod_i f_g(a_i)$. Rearrange and then cancel.]

[Remark: This theorem is also true for **finite non-abelian** groups but we don't have the technology to prove it yet. The technology we need is called "Lagrange's Theorem."]

(a) Fix an element $g \in G$. I claim that the function $f_g(a) = g * a$ is a bijection. *Proof.*

- **Injective.** Suppose that we have $f_g(a) = f_g(b)$ for some $a, b \in G$. Then by applying g^{-1} on the left we obtain

$$\begin{aligned} f_g(a) &= f_g(b) \\ g * a &= g * b \\ g^{-1} * g * a &= g^{-1} * g * b \\ a &= b. \end{aligned}$$

- **Surjective.** Consider any $b \in G$. Now define $a := g^{-1} * b$ and check that

$$f_g(a) = g * a = g * g^{-1} * b = b.$$

□

- (b) Now suppose that G is a **finite abelian** group with n elements, say

$$G = \{a_1, a_2, \dots, a_n\}.$$

For any element $g \in G$ I claim that $g^n = \varepsilon$. *Proof.* Let $b \in G$ denote the product of all n elements in some order:

$$b := a_1 * a_2 * \dots * a_n.$$

Since G is abelian the order of the product doesn't matter. On the other hand, we know from part (a) that $f_g : G \rightarrow G$ is a permutation of the elements of G . It follows that

$$\begin{aligned} b &= f_g(a_1) * f_g(a_2) * \dots * f_g(a_n) \\ &= (g * a_1) * (g * a_2) * \dots * (g * a_n) \\ &= (g * g * \dots * g) * (a_1 * a_2 * \dots * a_n) \\ &= g^n * (a_1 * a_2 * \dots * a_n) \\ &= g^n * b. \end{aligned}$$

Now multiply both sides by b^{-1} to obtain $g^n = \varepsilon$. □

[Remark: We will see later that this is a generalization of Fermat's Little Theorem and Euler's Totient Theorem.]

6. Join of Two Subgroups. Let G be a group and let $H, K \subseteq G$ be subgroups. Recall that the subgroup generated by the union $H \cup K$ is called the *join*:

$$H \vee K := \langle H \cup K \rangle = \text{the intersection of all subgroups that contain } H \cup K.$$

- (a) If $(G, +, 0)$ is abelian, we define the *sum* of H and K as follows:

$$H + K := \{h + k : h \in H, k \in K\}.$$

Prove that this is a subgroup.

- (b) If $(G, +, 0)$ is abelian, use part (b) to prove that $H \vee K = H + K$.
 (c) If $(G, *, \varepsilon)$ is non-abelian, show that the following set is **not** necessarily a subgroup, and hence it does not coincide with the join:

$$H * K := \{h * k : h \in H, k \in K\}.$$

[Hint: The smallest non-abelian group is S_3 .]

(a) Let $(G, +, 0)$ be an abelian group and let $H, K \subseteq G$ be any two subgroups. We will show that the sum

$$H + K = \{h + k : h \in H, k \in K\}$$

is a subgroup of G . *Proof.*

- **Closure.** Consider any two elements $h_1 + k_1$ and $h_2 + k_2$ from the set $H + K$. Since $H \subseteq G$ is a subgroup we have $h_1 + h_2 \in H$ and since $K \subseteq G$ is a subgroup we have $k_1 + k_2 \in K$. Then since G is abelian we see that the sum is in $H + K$:

$$\begin{aligned} (h_1 + k_1) + (h_2 + k_2) &= h_1 + (k_1 + h_2) + k_2 \\ &= h_1 + (h_2 + k_1) + k_2 \\ &= (h_1 + h_2) + (k_1 + k_2) \in H + K. \end{aligned}$$

- **Identity.** Since $0 \in H$ and $0 \in K$ we have $0 = 0 + 0 \in H + K$.
- **Inverses.** Consider any element $h + k \in H + K$. I claim that the inverse $-(h + k)$ is given by $(-h) + (-k)$. To see this we again use the fact that G is abelian:

$$(h + k) + (-h) + (-k) = (h - h) + (k - k) = 0 + 0 = 0.$$

But since H, K are subgroups we know that $-h \in H$ and $-k \in K$, which implies that

$$-(h + k) = (-h) + (-k) \in H + K,$$

as desired. □

[Shorter Proof: If $h_1 + k_1$ and $h_2 + k_2$ are in $H + K$ then

$$(h_1 + k_1) - (h_2 + k_2) = (h_1 - h_2) + (k_1 - k_2) \in H + K.]$$

(b) Following from (a), I claim that $H + K = H \vee K$. *Proof.* Consider any $h \in H$ and $k \in K$. Since the set $H \vee K$ contains $H \cup K$ it must contain the elements h and k , and since $H \vee K \subseteq G$ is a subgroup it must contain the sum $h + k$. Since this is true for all $h \in H$ and $k \in K$ we conclude that

$$H + K \subseteq H \vee K.$$

Conversely, note that for all $h \in H$ we have $h = h + 0 \in H + K$ and for all $k \in K$ we have $k = 0 + k \in H + K$. This implies that $H + K$ contains the set $H \cup K$. But from part (a) we know that $H + K \subseteq G$ is a subgroup. Finally, since $H \vee K \subseteq G$ is the smallest subgroup that contains $H \cup K$ we must have

$$H \vee K \subseteq H + K. \quad \square$$

(c) Consider the non-abelian group (S_3, \circ, id) and the (cyclic) subgroups

$$H = \{\text{id}, (12)\} \quad \text{and} \quad K = \{\text{id}, (23)\}.$$

Then by definition we have

$$\begin{aligned} H \circ K &= \{\text{id} \circ \text{id}, \text{id} \circ (23), (12) \circ \text{id}, (12) \circ (23)\} \\ &= \{\text{id}, (23), (12), (123)\}. \end{aligned}$$

This is **not** a subgroup of S_3 because the element $(132) = (123)^{-1} = (23) \circ (12)$ is not contained in the set.

[Remark: If I had instead chosen $G = S_4$ with subgroups $H = \{\text{id}, (12)\}$ and $K = \{\text{id}, (34)\}$ then I would have accidentally found that $H \circ K$ **is** a subgroup. We'll talk about that later.]

Week 5

This week we will dive into the structure of the infinite cyclic group $\mathbb{Z}^+ = (\mathbb{Z}, +, 0)$. In the process we will meet the concepts of “poset” (partially-ordered set) set and “lattice.”

Definition of Posets and Lattices. Let P be a set equipped with an abstract relation “ \leq .” We say that the pair (P, \leq) is a *poset* if the following three axioms are satisfied:

(P1) The relation \leq is *reflexive*: for all $a \in P$ we have

$$a \leq a.$$

(P2) The relation \leq is *anti-symmetric*: for all $a, b \in P$,

$$\text{if } a \leq b \text{ and } b \leq a \text{ then we have } a = b.$$

(P3) The relation \leq is *transitive*: for all $a, b, c \in P$,

$$\text{if } a \leq b \text{ and } b \leq c \text{ then we have } a \leq c.$$

Moreover, we say that the poset (P, \leq) is a *lattice* if it satisfies the following additional axiom:

(L) Every subset of P has a greatest lower bound and a least upper bound.

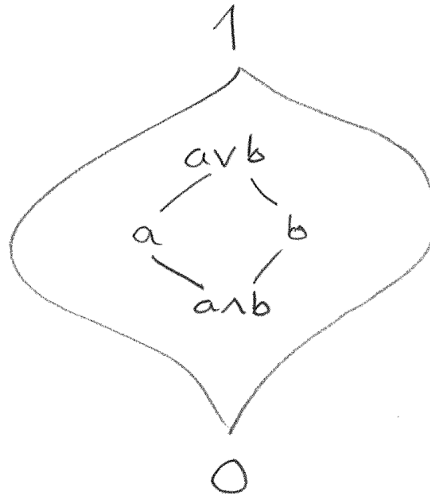
In the special case of two elements $a, b \in P$ we say that $g \in P$ is a *greatest lower bound* of a, b if it satisfies the following two properties:

- $g \leq a$ and $g \leq b$ (i.e., g is a lower bound of a and b),
- for all $c \in P$, if $c \leq a$ and $c \leq b$ then $c \leq g$ (i.e., every lower bound is less than g).

[Exercise: Check that this g is unique.] In this case we will write $g = a \wedge b$ and we will call g the *meet* of a and b . Dually, we say $\ell \in P$ is a (the) *least upper bound* if it satisfies:

- $a \leq \ell$ and $b \leq \ell$ (i.e., ℓ is an upper bound of a and b),
- for all $c \in P$, if $a \leq c$ and $b \leq c$ then $\ell \leq c$ (i.e., every lower bound is greater than ℓ).

We will write $\ell = a \vee b$ and call this the *join* of a and b . On the other extreme, we will use the symbols 0 and 1 for the greatest lower bound and the least upper bound of **all** elements in P , which satisfy $0 \leq a$ and $a \leq 1$ for all $a \in P$. Here is how I visualize a lattice:



///

Example: The Lattice of Subsets. Let U be any set and let 2^U be the set of all subsets of U . I claim that 2^U is a lattice with the following structure:

partial order \leq	set containment \subseteq
meet \wedge	intersection \cap
join \vee	union \cup
bottom 0	empty set \emptyset
top 1	the universe U

Example: The Lattice of Subgroups. Now let $(G, *, \varepsilon)$ be a group and let $\mathcal{L}(G)$ be the set of all subgroups of G . I claim that $\mathcal{L}(G)$ is a lattice with the following structure:

partial order \leq	set containment \subseteq
meet \wedge	intersection \cap
join \vee	join \vee
bottom 0	trivial group $\{\varepsilon\}$
top 1	full group G

In particular, note that the join operation (least upper bound) coincides with the join of subgroups that we defined earlier:

$$H \vee K = \langle H \cup K \rangle = \text{the intersection of all subgroups that contain } H \cup K.$$

That was good planning on my part.

///

It turns out that the lattice of subgroups of \mathbb{Z}^+ has a particularly nice structure.

Theorem (Subgroups of \mathbb{Z}^+). The lattice of subgroups of \mathbb{Z}^+ under containment is isomorphic to the lattice of non-negative integers under “reverse divisibility:”

$$(\mathcal{L}(\mathbb{Z}^+), \subseteq) \cong (\mathbb{N}, \text{reverse divisibility}).$$

The proof has three steps.

Step 1. We already know that the **cyclic** subgroups of \mathbb{Z}^+ have the form

$$m\mathbb{Z} = \{mk : k \in \mathbb{Z}\}$$

for some $m \in \mathbb{N}$. I claim that **every** subgroup has this form.

Proof. Let $H \subseteq \mathbb{Z}^+$ be any subgroup. If $H = \{0\}$ is the trivial group then we have $H = 0\mathbb{Z}$ as desired. Otherwise, suppose that $H \neq \{0\}$ and let m be the smallest positive element of H . In this case I claim that $H = m\mathbb{Z}$. Indeed, since $m\mathbb{Z} = \langle m \rangle$ is the smallest subgroup containing m we must have $m\mathbb{Z} \subseteq H$. On the other hand, let $n \in H$ be any element of H and divide it by m to obtain

$$\begin{cases} n = qm + r, \\ 0 \leq r < m. \end{cases}$$

We will show that $r = 0$. To see this, first observe that since n and m are in H we also have $r = n - qm \in H$. But if $r \neq 0$ then $0 < r < m$ contradicts the minimality of m . We conclude that $r = 0$ and hence $n = qm \in m\mathbb{Z}$. Finally, since $n \in H$ was arbitrary we conclude that $H \subseteq m\mathbb{Z}$ as desired. \square

Step 2. For all integers $a, b \in \mathbb{Z}$ recall that we define divisibility as follows:

$$"a|b" = "a \text{ divides } b" = "\exists k \in \mathbb{Z}, ak = b."$$

Then for all $a, b \in \mathbb{Z}$ I claim that

$$a\mathbb{Z} \subseteq b\mathbb{Z} \iff b|a.$$

Proof. If $a\mathbb{Z} \subseteq b\mathbb{Z}$ then since $a \in a\mathbb{Z} \subseteq b\mathbb{Z}$ we must have $a = bk$ for some $k \in \mathbb{Z}$, hence $b|a$. Conversely, suppose that $b|a$ so there exists $k \in \mathbb{Z}$ with $a = bk$. Then for any $al \in a\mathbb{Z}$ we have

$$al = (bk)l = b(kl) \in b\mathbb{Z}$$

and it follows that $a\mathbb{Z} \subseteq b\mathbb{Z}$ as desired. \square

Step 3. Now consider the function $f : \mathbb{N} \rightarrow \mathcal{L}(\mathbb{Z}^+)$ defined by $f(m) := m\mathbb{Z}$. I claim that this is an isomorphism of posets.

Proof. We saw in Step 1 that this function is surjective. If we can show that the function is **injective** (hence invertible) then it follows from Step 2 that the function f and its inverse f^{-1} both preserve order. So let us assume that $a\mathbb{Z} = b\mathbb{Z}$ for some non-negative integers $a, b \in \mathbb{N}$. From Step 2 we know that $a|b$ and $b|a$, hence there exist integers $k, \ell \in \mathbb{Z}$ with $ak = b$ and $b\ell = a$. If either a or b is zero this implies that $a = b = 0$ as desired. Otherwise, both a and b are positive and we have

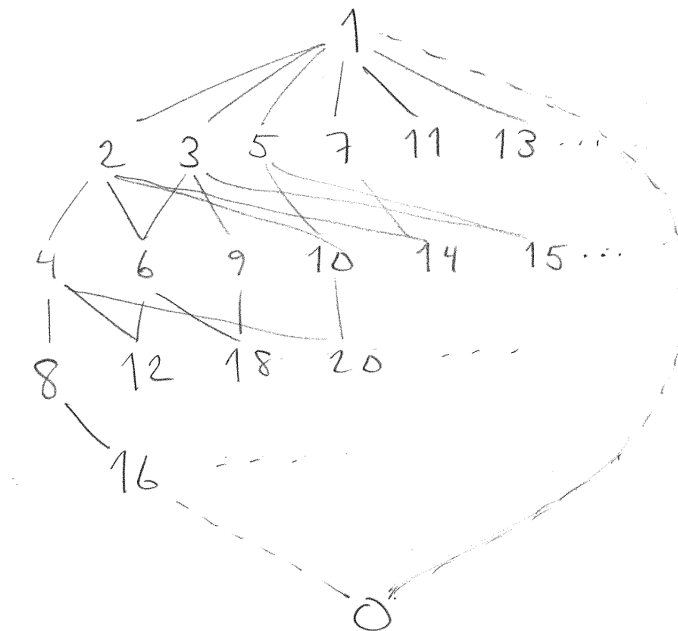
$$\begin{aligned} a &= b\ell \\ a &= ak\ell \\ a(1 - k\ell) &= 0 \\ (1 - k\ell) &= 0 \\ 1 &= k\ell. \end{aligned}$$

The only solutions are $k = \ell = \pm 1$ which implies that $a = \pm b$. Finally, since a, b are both positive we conclude that $a = b$ as desired. \square

Remarks:

- Since f preserves order we call it a *poset homomorphism*. Since f^{-1} exists and also preserves order we call the pair (f, f^{-1}) a poset isomorphism.
- Unlike in the case of groups, an invertible poset homomorphism is not necessarily an isomorphism. [Exercise: Find a small example.]

This completes the proof that $\mathcal{L}(\mathbb{Z}^+)$ is isomorphic to \mathbb{N} under reverse divisibility. Here is a picture of this lattice:



Notice that 0 is divisible by everything and everything is divisible by 1. The elements just below 1 are the prime numbers. We know that this poset is a lattice because it is isomorphic to the lattice $\mathcal{L}(\mathbb{Z}^+)$. What are the meet and join operations?

For any group G recall that the set $\mathcal{L}(G)$ of subgroups is a lattice under the containment partial order. For any subgroups $H, K \in \mathcal{L}(G)$ the meet $H \wedge K$ equals the intersection $H \cap K$ and the join $H \vee K$ can be defined as the intersection of all subgroups that contain the union. Moreover, if $(G, +, 0)$ is abelian then the join equals the sum: $H \vee K = H + K$.

Now let's consider the case $G = (\mathbb{Z}, +, 0)$. We saw above that every subgroup $H \subseteq G$ has the form $H = m\mathbb{Z}$ for a unique non-negative integer $m \in \mathbb{N}$. Hence for any two subgroups $a\mathbb{Z}$ and $b\mathbb{Z}$ **there exist unique non-negative integers** $m, d \in \mathbb{N}$ such that

$$\begin{aligned} a\mathbb{Z} \wedge b\mathbb{Z} &= a\mathbb{Z} \cap b\mathbb{Z} = m\mathbb{Z}, \\ a\mathbb{Z} \vee b\mathbb{Z} &= a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}. \end{aligned}$$

What are these numbers m and d ?

Theorem. We have $m = \text{lcm}(a, b)$ and $d = \text{gcd}(a, b)$.

Proof. First we will prove that $m = \text{lcm}(a, b)$. Since $m \in m\mathbb{Z} = a\mathbb{Z} \cap b\mathbb{Z}$ we must have $m \in a\mathbb{Z}$ and $m \in b\mathbb{Z}$, which implies that $a|m$ and $b|m$. In other words, m is a common multiple of a and b . Now let c be **any** common multiple of a and b , so that $c \in a\mathbb{Z}$ and $c \in b\mathbb{Z}$. It follows that $c \in a\mathbb{Z} \cap b\mathbb{Z} = m\mathbb{Z}$, which implies that $m|c$. In other words, m is the **least** common multiple.

Now we will prove that $d = \text{gcd}(a, b)$. Note that $a = a1+b0 \in a\mathbb{Z}+b\mathbb{Z}$ and $b = a0+b1 \in a\mathbb{Z}+b\mathbb{Z}$. Since $a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}$ this implies that $a \in d\mathbb{Z}$ and $b \in d\mathbb{Z}$, hence $d|a$ and $d|b$. In other words, d is a common divisor of a and b . Now let c be **any** common divisor, so that $a \in c\mathbb{Z}$ and $b \in c\mathbb{Z}$. It follows [Exercise: How?] that $d\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z} \subseteq c\mathbb{Z}$, which implies that $c|d$. In other words, d is the **greatest** common divisor. \square

Remarks:

- For any $a \in \mathbb{N}$ we have $a\mathbb{Z} \cap 0\mathbb{Z} = 0\mathbb{Z}$ and $a\mathbb{Z} + 0\mathbb{Z} = a\mathbb{Z}$, which implies that

$$\text{lcm}(a, 0) = 0 \quad \text{and} \quad \text{gcd}(a, 0) = a.$$

- For any $a \in \mathbb{N}$ we have $a\mathbb{Z} \cap 1\mathbb{Z} = a\mathbb{Z}$ and $a\mathbb{Z} + 1\mathbb{Z} = 1\mathbb{Z}$, which implies that

$$\text{lcm}(a, 1) = a \quad \text{and} \quad \text{gcd}(a, 1) = 1.$$

- Since $0\mathbb{Z} \cap 0\mathbb{Z} = 0\mathbb{Z}$ and $0\mathbb{Z} + 0\mathbb{Z} = 0\mathbb{Z}$, the theorem also says

$$\text{lcm}(0, 0) = 0 \quad \text{and} \quad \text{gcd}(0, 0) = 0,$$

but you may object to this. It's not important.

It follows that the poset $(\mathbb{N}, \text{reverse divisibility})$ is a lattice with $\text{meet}=\text{lcm}$ and $\text{join}=\text{gcd}$. [Equivalently: The poset $(\mathbb{N}, \text{divisibility})$ is a lattice with $\text{meet}=\text{gcd}$ and $\text{join}=\text{lcm}$.] To end this section I want to bring your attention to the following corollary which is very important in the study of prime numbers.

Important Corollary (Bézout's Identity). Let $a, b \in \mathbb{Z}$ be any two integers and let $d = \text{gcd}(a, b)$. Then there exist some (non-unique) integers $x, y \in \mathbb{Z}$ such that

$$d = ax + by.$$

Proof. From the above theorem we know that $d\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z}$. Then since $d = d1 \in d\mathbb{Z}$ we must have $d \in a\mathbb{Z} + b\mathbb{Z} = \{ax + by : x, y \in \mathbb{Z}\}$. \square

If $\langle g \rangle$ is an infinite cyclic group we have seen that the lattice of subgroups $\mathcal{L}\langle g \rangle$ is isomorphic to the lattice of natural numbers \mathbb{N} under reverse-divisibility. Specifically, the isomorphism $f : \mathbb{N} \rightarrow \mathcal{L}\langle g \rangle$ is defined by $f(m) = \langle g^m \rangle$. But what if the cyclic group $\langle g \rangle$ is **finite**?

I'll just tell you the answer and then we'll discuss how to prove it.

Fundamental Theorem of Cyclic Groups. For any $n \in \mathbb{N}$ let $\text{Div}(n) \subseteq \mathbb{N}$ be the set of non-negative divisors of n . Thus $\text{Div}(0) = \mathbb{N}$ and for $n \geq 1$ the set $\text{Div}(n)$ is finite.

If $\langle g \rangle$ is an infinite cyclic group, we have already seen that the function $f : \text{Div}(0) \rightarrow \mathcal{L}\langle g \rangle$ defined by $f(k) := \langle g^k \rangle$ defines a poset isomorphism:

$$(\mathcal{L}\langle g \rangle, \subseteq) \cong (\text{Div}(0), \text{reverse divisibility}).$$

Now if $\langle g \rangle$ is a **finite** cyclic group of size $n \geq 1$ then I claim that the same function f restricted to $\text{Div}(n) \subseteq \text{Div}(0)$ defines a poset isomorphism:

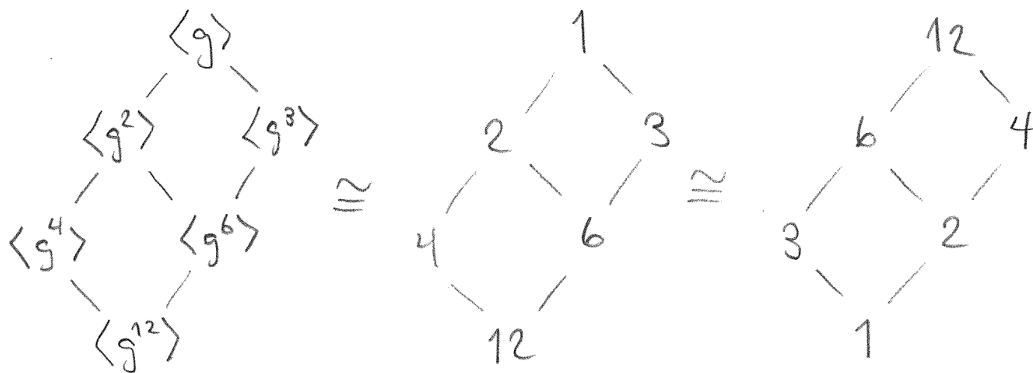
$$(\mathcal{L}\langle g \rangle, \subseteq) \cong (\text{Div}(n), \text{reverse divisibility}).$$

Finally, if $n \geq 1$ then the permutation $\text{Div}(n) \rightarrow \text{Div}(n)$ defined by $d \mapsto n/d$ switches the relations of reverse divisibility and divisibility, and hence

$$(\mathcal{L}\langle g \rangle, \subseteq) \cong (\text{Div}(n), \text{reverse divisibility}) \cong (\text{Div}(n), \text{divisibility}).$$

[Remark: This last isomorphism doesn't work when $n = 0$.] ///

For example, here is a picture of the theorem when $n = 12$:



I could give a quick and dirty proof right now but I prefer to develop a slower and more abstract proof that illustrates what's really going on. The fundamental idea is that of a "Galois connection" between posets.

Definition of Galois Connections.³ Let (P, \leq) and (Q, \leq) be posets and consider any functions $f : P \rightarrow Q$ and $g : Q \rightarrow P$. The pair f, g is a *Galois connection* if it satisfies

$$p \leq g(q) \iff f(p) \leq q \text{ for all } p \in P \text{ and } q \in Q.$$

///

For example, if $f : P \rightarrow Q$ is a poset isomorphism then the pair (f, f^{-1}) is a Galois connection. Indeed, in that case we have $f(p) \leq q \Rightarrow f^{-1}(f(p)) \leq f^{-1}(q) \Rightarrow p \leq f^{-1}(q)$ and vice versa. A general Galois connection f, g need **not** be an isomorphism, but it always **restricts** to an isomorphism between certain **subposets** $P' \subseteq P$ and $Q' \subseteq Q$.

Fundamental Theorem of Galois Connections. If $f : P \rightleftarrows Q : g$ is a Galois connection then we have the following properties:

- $f : P \rightarrow Q$ and $g : Q \rightarrow P$ are poset homomorphisms,
- $g \circ f : P \rightarrow P$ is increasing and $f \circ g : Q \rightarrow Q$ is decreasing,
- $f \circ g \circ f = f$ and $g \circ f \circ g = g$.

If we define the subposets

$$P' = g[Q] := \{g(q) : q \in Q\} \quad \text{and} \quad Q' = f[P] := \{f(p) : p \in P\},$$

then it follows from the above properties that f and g restrict to a **poset isomorphism**:

$$f : P' \xrightarrow{\sim} Q' : g.$$

Proof. See the homework. □

[Remark: Specific examples of this theorem are often called “correspondence theorems.” We will see one below.]

Galois connections are best understood with a picture. Since I can’t draw general posets, let me assume for convenience that $(P, \leq, \vee, \wedge, 0_P, 1_P)$ and $(Q, \leq, \vee, \wedge, 0_Q, 1_Q)$ are lattices. In this case I claim that $f(0_P) = 0_Q$ and $g(1_Q) = 1_P$.

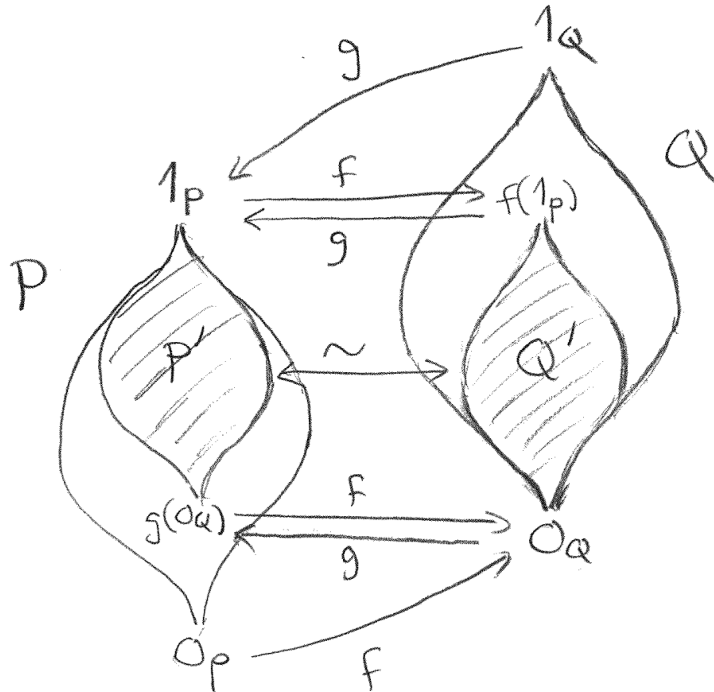
Proof. By definition we have $0_P \leq p$ for all $p \in P$ and $0_Q \leq q$ for all $q \in Q$. In particular, setting $q = f(0_P)$ in the second inequality gives $0_Q \leq f(0_P)$. On the other hand, the definition of Galois connections says that

$$0_P \leq g(q) \iff f(0_P) \leq q \text{ for all } q \in Q.$$

³You won’t find this concept in any of the standard algebra textbooks. I think that’s a shame.

In particular, since $0_P \leq g(0_Q)$ we conclude that $f(0_P) \leq 0_Q$, and then it follows from antisymmetry that $f(0_P) = 0_Q$. The proof of $g(1_Q) = 1_P$ is similar. \square

Then here is the picture:



Let me emphasize that the images P' and Q' are isomorphic as posets, but the original P and Q need not be. And what does all of this have to do with Évariste Galois? I'll tell you later. For now, an example.

Example: Image and Preimage. Let $(G, *, \delta)$ and $(H, \bullet, \varepsilon)$ be groups and let $\varphi : G \rightarrow H$ be any function. Then for all subsets $S \subseteq G$ and $T \subseteq H$ we define the *image set* $\varphi[S] \subseteq H$ and the *preimage set* $\varphi^{-1}[T] \subseteq G$ as follows:

$$\begin{aligned}\varphi[S] &:= \{\varphi(g) : g \in S\} \subseteq H, \\ \varphi^{-1}[T] &:= \{g \in G : \varphi(g) \in T\} \subseteq G.\end{aligned}$$

Remarks:

- I use square brackets to distinguish between the function $\varphi : G \rightarrow H$ that sends elements to elements and the function $\varphi : 2^G \rightarrow 2^H$ that sends subsets to subsets.

- The **preimage** function $\varphi^{-1} : 2^H \rightarrow 2^G$ always exists, but the **inverse** function $\varphi^{-1} : H \rightarrow G$ need not exist. The inverse function exists if and only if for all $h \in H$ the preimage $\varphi^{-1}[\{h\}] \subseteq G$ consists one element, which we may then call $\varphi^{-1}(h)$.

If we think of $(2^G, \subseteq)$ and $(2^H, \subseteq)$ as posets then I claim that the image and preimage functions are a Galois connection:

$$\varphi : 2^G \rightleftarrows 2^H : \varphi^{-1}.$$

Proof. For all subsets $S \subseteq G$ and $T \subseteq H$ we have

$$\begin{aligned} S \subseteq \varphi^{-1}[T] &\iff \forall s \in S, s \in \varphi^{-1}[T] \\ &\iff \forall s \in S, \varphi(s) \in T \\ &\iff \varphi[S] \subseteq T. \end{aligned}$$

□

So far these remarks apply to any sets G, H and any function $\varphi : G \rightarrow H$. Now let us assume that φ is a **group homomorphism**. In this case you will verify on the homework that

- if $S \subseteq G$ is a subgroup then $\varphi[S] \subseteq H$ is a subgroup,
- if $T \subseteq H$ is a subgroup then $\varphi^{-1}[T] \subseteq G$ is a subgroup,

and hence we obtain a Galois connection $\varphi : \mathcal{L}(G) \rightleftarrows \mathcal{L}(H) : \varphi^{-1}$ between the lattices of subgroups. It follows from the Fundamental Theorem of Galois Connections that image and preimage restrict to an **isomorphism** between certain posets of subgroups:

$$\varphi : \mathcal{L}(G)' \xrightarrow{\sim} \mathcal{L}(H)' : \varphi^{-1}.$$

In other words, we obtain an order-preserving bijection between certain kinds of subgroups of G and certain kinds of subgroups of H .

What kinds of subgroups? Let me spoil the surprise right now: It will turn out that $\mathcal{L}(G)'$ consists of subgroups that **contain** the kernel “ $\ker \varphi$,” and $\mathcal{L}(H)'$ consists of subgroups of H that are **contained in** the image “ $\text{im } \varphi$.” Next time I will define the notions “ $\ker \varphi$ ” and “ $\text{im } \varphi$ ” and I will prove these assertions.

We have seen that any group homomorphism $\varphi : G \rightarrow H$ induces the image and preimage functions $\varphi : \mathcal{L}(G) \rightleftarrows \mathcal{L}(H) : \varphi^{-1}$ which form an abstract Galois connection. Among the images and preimages there are two important special cases.

Definition of the Kernel and Image. Let $\varphi : (G, *, \delta) \rightarrow (H, \bullet, \varepsilon)$ be a group homomorphism. We define the *kernel of φ* as the preimage of the trivial subgroup $\{\varepsilon\} \subseteq H$:

$$\ker \varphi := \varphi^{-1}[\{\varepsilon\}] = \{g \in G : \varphi(g) = \varepsilon\}.$$

And we define the *image of φ* as the image of the full group G :

$$\text{im } \varphi := \varphi[G] = \{\varphi(g) : g \in G\}.$$

From general properties (proved on the homework) we know that $\ker \varphi \subseteq G$ and $\text{im } \varphi \subseteq H$ are subgroups. ///

We now have the ingredients necessary to state and prove an important general theorem. Afterwards, we will obtain the Fundamental Theorem of Cyclic Groups as an easy corollary.

The Correspondence Theorem for Groups. Let $\varphi : (G, *, \delta) \rightarrow (H, \bullet, \varepsilon)$ be any group homomorphism and define the following sets, partially ordered by containment:

$$\begin{aligned} \mathcal{L}(G, \ker \varphi) &:= \{\text{subgroups } \ker \varphi \subseteq K \subseteq G\} \\ \mathcal{L}(\text{im } \varphi) &:= \{\text{subgroups } L \subseteq \text{im } \varphi\}. \end{aligned}$$

I claim that the image and preimage $\varphi : \mathcal{L}(G) \rightleftarrows \mathcal{L}(H) : \varphi^{-1}$ restrict to a poset isomorphism:

$$\varphi : \mathcal{L}(G, \ker \varphi) \xrightarrow{\sim} \mathcal{L}(\text{im } \varphi) : \varphi^{-1}.$$

///

Proof. Since $\varphi : \mathcal{L}(G) \rightleftarrows \mathcal{L}(H) : \varphi^{-1}$ is a Galois connection we automatically obtain a poset isomorphism $\varphi : \mathcal{L}(G)' \longleftrightarrow \mathcal{L}(H)' : \varphi^{-1}$ between certain subposets $\mathcal{L}(G)' \subseteq \mathcal{L}(G)$ and $\mathcal{L}(H)' = \mathcal{L}(H)$. On the homework you will show that these subposets are

$$\begin{aligned} \mathcal{L}(G)' &= \{K \subseteq G : K = \varphi^{-1}[\varphi[K]]\}, \\ \mathcal{L}(H)' &= \{L \subseteq H : L = \varphi[\varphi^{-1}[L]]\}. \end{aligned}$$

Assuming this, I will prove that

$$\mathcal{L}(G)' = \mathcal{L}(G, \ker \varphi) \quad \text{and} \quad \mathcal{L}(H)' = \mathcal{L}(\text{im } \varphi).$$

There are two steps in the proof.

Step 1. For all subgroups $K \subseteq G$ and $L \subseteq H$ I claim that

$$\begin{aligned} \varphi[\varphi^{-1}[L]] &= L \wedge \text{im } \varphi, \\ \varphi^{-1}[\varphi[K]] &= K \vee \ker \varphi. \end{aligned}$$

For the first equality, note that $\varphi^{-1}[L] \subseteq G$ implies $\varphi[\varphi^{-1}[L]] \subseteq \varphi[G] = \text{im } \varphi$ because $\varphi[-]$ preserves order, and that $\varphi[\varphi^{-1}[L]] \subseteq L$ because $\varphi \circ \varphi^{-1}[-]$ is decreasing. Therefore we have $\varphi[\varphi^{-1}[L]] \subseteq L \cap \text{im } \varphi = L \wedge \text{im } \varphi$. For the converse, consider any element $h \in L \wedge \text{im } \varphi = L \cap \text{im } \varphi$. Since $h \in \text{im } \varphi$ we must have $h = \varphi(g)$ for some $g \in G$ and since $h \in L$ we must have $g \in \varphi^{-1}[L]$. Now it follows that $h = \varphi(g) \in \varphi[\varphi^{-1}[L]]$ and hence $L \wedge \text{im } \varphi \subseteq \varphi[\varphi^{-1}[L]]$.

For the second equality, note that $\{\varepsilon\} \subseteq \varphi[K]$ implies $\ker \varphi = \varphi^{-1}[\{\varepsilon\}] \subseteq \varphi^{-1}[\varphi[K]]$ because $\varphi^{-1}[-]$ is order-preserving, and that $K \subseteq \varphi^{-1}[\varphi[K]]$ because $\varphi^{-1} \circ \varphi[-]$ is increasing. Then since $\varphi^{-1}[\varphi[K]]$ is a subgroup of G containing $K \cup \ker \varphi$ we must have $K \vee \ker \varphi \subseteq \varphi^{-1}[\varphi[K]]$. For the converse, consider any element $g \in \varphi^{-1}[\varphi[K]]$. By definition this means that $\varphi(g) = \varphi(k)$ for some $k \in K$. Then by general properties of homomorphisms we have

$$\begin{aligned}\varphi(g) &= \varphi(k) \\ \varphi(k)^{-1} \bullet \varphi(g) &= \varepsilon \\ \varphi(k^{-1} * g) &= \varepsilon,\end{aligned}$$

and hence $k^{-1} * g \in \ker \varphi$. Finally, consider the product set $K \ker \varphi := \{k * \ell : k \in K, \ell \in \ker \varphi\}$. One can show that this set is a group, and hence that $K \ker \varphi = K \vee \ker \varphi$. [Remark: Later we will show the more general fact that $KN = K \vee N$ for any subgroups $K, N \subseteq G$ such that $N \subseteq G$ is “normal.”] It follows from this that

$$g = k * (k^{-1} * g) \in K \ker \varphi = K \vee \ker \varphi,$$

and hence $\varphi^{-1}[\varphi[K]] \subseteq K \vee \ker \varphi$, as desired. □

[Remark: The first three inclusions were purely formal. Only the last inclusion $\varphi^{-1}[\varphi[K]] \subseteq K \vee \ker \varphi$ used any non-trivial details about groups.]

Step 2. In Step 1 we proved that

$$\begin{aligned}\mathcal{L}(G)' &= \{K \subseteq G : K = K \vee \ker \varphi\}, \\ \mathcal{L}(H)' &= \{L \subseteq H : L = L \wedge \text{im } \varphi\}.\end{aligned}$$

Now it only remains to show that

$$\begin{aligned}K = K \vee \ker \varphi &\iff \ker \varphi \subseteq K, \\ L = L \wedge \text{im } \varphi &\iff L \subseteq \text{im } \varphi.\end{aligned}$$

This has nothing to do with groups so I will prove it for general lattices. Let $(L, \leq, \vee, \wedge, 0, 1)$ be a lattice and consider any elements $a, b \in L$. Then I claim that

$$\begin{aligned}a = a \vee b &\iff b \leq a, \\ a = a \wedge b &\iff a \leq b.\end{aligned}$$

For the first statement, if $a = a \vee b$ then by definition we have $b \leq a \vee b = a$. Conversely, suppose that $b \leq a$. Then we have $a \leq a \vee b$ by definition and we have $a \vee b \leq a$ because a is an upper bound of a and b . Hence $a = a \vee b$.

For the second statement, if $a = a \wedge b$ then by definition we have $a = a \wedge b \leq b$. Conversely, suppose that $a \leq b$. Then we have $a \wedge b \leq a$ by definition and we have $a \leq a \wedge b$ because a is a lower bound of a and b . Hence $a = a \wedge b$. \square

This completes the proof of the Correspondence Theorem for Groups. Finally, we obtain a free proof of the Fundamental Theorem of Cyclic Groups.

Corollary (Fundamental Theorem of Cyclic Groups). Let $\langle g \rangle$ be a cyclic group and consider the group homomorphism $\varphi : \mathbb{Z}^+ \rightarrow \langle g \rangle$ defined by $\varphi(k) := g^k$. Note that we have $\text{im } \varphi = \langle g \rangle$ by definition, and since the kernel is a subgroup of \mathbb{Z}^+ we must have $\ker \varphi = n\mathbb{Z}$ for some unique $n \in \mathbb{N}$. If $n = 0$ then $\langle g \rangle$ is infinite and otherwise we have $\# \langle g \rangle = n$.

Now we conclude from the Correspondence Theorem that

$$\mathcal{L}\langle g \rangle = \mathcal{L}(\text{im } \varphi) \cong \mathcal{L}(\mathbb{Z}^+, \ker \varphi) = \mathcal{L}(1\mathbb{Z}, n\mathbb{Z}).$$

But recall that the subgroups of \mathbb{Z}^+ between $1\mathbb{Z}$ and $n\mathbb{Z}$ have the form $d\mathbb{Z}$ where d is a divisor of n , and that these groups are ordered by “reverse divisibility.” It follows that

$$\mathcal{L}\langle g \rangle \cong \mathcal{L}(1\mathbb{Z}, n\mathbb{Z}) \cong (\text{Div}(n), \text{reverse divisibility}),$$

and the explicit isomorphism $\text{Div}(n) \rightarrow \mathcal{L}\langle g \rangle$ is given by the image function $\varphi[-]$:

$$d \mapsto d\mathbb{Z} \mapsto \varphi[d\mathbb{Z}] = \{g^{dk} : k \in \mathbb{Z}\} = \langle g^d \rangle.$$

\square

Week 6

Last week we discussed the abstract properties of the symbol “ \leq .” This week we’ll discuss the abstract properties of the symbol “ $=$.”

Definition of Equivalence and Classes. Let S be a set and let \sim be a relation on S . Technically: This means that \sim is a subset of $S \times S$. We will write “ $a \sim b$ ” to mean that “ $(a, b) \in \sim$.” We say that \sim is an *equivalence relation* if the following three axioms are satisfied:

(E1) The relation \sim is *reflexive*: for all $a \in S$ we have

$$a \sim a.$$

(E2) The relation \sim is *symmetric*: for all $a, b \in P$,

$$\text{if } a \sim b \text{ then } b \sim a.$$

(E3) The relation \sim is *transitive*: for all $a, b, c \in P$,

$$\text{if } a \sim b \text{ and } b \sim c \text{ then we have } a \sim c.$$

[Remark: The symbol “ \sim ” always denotes our favorite equivalence relation on a given set.]
For each element $a \in S$ we define its *equivalence class* as follows:

$$[a]_{\sim} := \{b \in S : a \sim b\}.$$

We will use the notation

$$S/\sim = “S \text{ mod } \sim” = \text{the set of equivalence classes.}$$

One should check that for all $a, b \in S$ the following conditions are equivalent:

- $a \sim b$,
- $[a]_{\sim} = [b]_{\sim}$,
- $[a]_{\sim} \cap [b]_{\sim} \neq \emptyset$.

It follows that the equivalence classes are a *partition* of the set S . In other words, we can express S as a disjoint union:

$$S = \coprod_{X \in S/\sim} X.$$

///

I assume that you are familiar with the following example.

Example: Equivalence Modulo an Integer. Fix an integer $n \in \mathbb{Z}$. Then for all integers $a, b \in \mathbb{Z}$ we define

$$a \sim_n b \iff b - a \in n\mathbb{Z} \iff n|(b - a).$$

We call this relation “equivalence mod n .” I won’t bother to prove that this is an equivalence relation because it will follow from a more general proof in the next example.

In the case of equivalence mod n we have a special notation for equivalence classes:

$$\begin{aligned} [a]_{\sim_n} &= \{b \in \mathbb{Z} : a \sim_n b\} \\ &= \{b \in \mathbb{Z} : b - a \in n\mathbb{Z}\} \\ &= \{b \in \mathbb{Z} : b - a = nk \text{ for some } k \in \mathbb{Z}\} \\ &= \{b \in \mathbb{Z} : b = a + nk \text{ for some } k \in \mathbb{Z}\} \end{aligned}$$

$$\begin{aligned}
&= \{a + nk : k \in \mathbb{Z}\} \\
&=: a + n\mathbb{Z}.
\end{aligned}$$

The equivalence class $[a]_{\sim_n} = a + n\mathbb{Z}$ is called a *coset* of the subgroup $n\mathbb{Z} \subseteq \mathbb{Z}$. If $n = 0$ then each coset has a single element:

$$a + 0\mathbb{Z} = \{a + 0k : k \in \mathbb{Z}\} = \{a\}.$$

Thus we see that “equivalent mod 0” is just a fancy way to say “equals:”

$$a \sim_0 b \iff a + 0\mathbb{Z} = b + 0\mathbb{Z} \iff \{a\} = \{b\} \iff a = b.$$

If $n \neq 0$ then each coset is in one-to-one correspondence with \mathbb{Z} :

$$a + n\mathbb{Z} = \{\dots, a - 2n, a - n, a, a + n, a + 2n, \dots\}.$$

In this case, the partition of \mathbb{Z} into equivalence classes is called “division with remainder.” By convention we say that a “remainder mod n ” must satisfy $0 \leq r < |n|$. Therefore we have the following disjoint union:

$$\mathbb{Z} = \coprod_{r=0}^{|n|-1} \{\text{integers with remainder } r \text{ mod } n\} = \coprod_{r=0}^{|n|-1} (r + n\mathbb{Z}).$$

And instead of \mathbb{Z}/\sim_n we use the following notation for the set of cosets:

$$\mathbb{Z}/n\mathbb{Z} = \{n\mathbb{Z}, 1 + n\mathbb{Z}, 2 + n\mathbb{Z}, \dots, (|n| - 1) + n\mathbb{Z}\}.$$

///

That was just an example. Here is the general concept.

Definition of Equivalence Modulo a Subgroup. Let $(G, *, \varepsilon)$ be a group and let $H \subseteq G$ be any subgroup. Then for all $a, b \in G$ we define

$$\begin{aligned}
a \sim_H b &\iff a^{-1} * b \in H \iff b^{-1} * a \in H, \\
a \underset{H}{\sim} b &\iff a * b^{-1} \in H \iff b * a^{-1} \in H.
\end{aligned}$$

I claim that each of \sim_H and $\underset{H}{\sim}$ is an equivalence relation on G . We call these relations left and right *equivalence mod H* .

Proof. Let $H \subseteq G$ be a subgroup. We will prove that left equivalence \sim_H is an equivalence relation and leave the proof of right equivalence $\underset{H}{\sim}$ to the reader.

(E1) Consider any $a \in G$. Since the subgroup H contains the identity ε we have $a^{-1} * a = \varepsilon \in H$, and hence $a \sim_H a$.

(E2) Consider any $a, b \in G$ such that $a \sim_H b$. By definition this means that $a^{-1} * b \in H$. Then since the subgroup H is closed under inversion we have $b^{-1} * a = (a^{-1} * b)^{-1} \in H$, and hence $b \sim_H a$.

(E3) Consider any $a, b, c \in G$ such that $a \sim_H b$ and $b \sim_H c$. By definition this means that $a^{-1} * b \in H$ and $b^{-1} * c \in H$. Then since the subgroup H is closed under $*$ we have

$$a^{-1} * c = (a^{-1} * b) * (b^{-1} * c) \in H,$$

and hence $a \sim_H c$. □

[Remark: Note that the three axioms of equivalence correspond perfectly with the three axioms of a subgroup. This is excellent motivation for the definition.]

Now let's examine the equivalence classes. The classes of \sim_H are called *left cosets*:

$$\begin{aligned} [a]_{\sim_H} &= \{b \in G : a \sim_H b\} \\ &= \{b \in G : a^{-1} * b \in H\} \\ &= \{b \in G : a^{-1} * b = h \text{ for some } h \in H\} \\ &= \{b \in G : b = a * h \text{ for some } h \in H\} \\ &= \{a * h : h \in H\} \\ &=: aH. \end{aligned}$$

And the classes of ${}_H\sim$ are called *right cosets*:

$$[a]_{{}_H\sim} = \{h * a : h \in H\} =: Ha.$$

Instead of G/\sim_H and $G/{}_H\sim$, we will use the following notations for sets of cosets:

$$\begin{aligned} G/H &:= \text{the set of left cosets of } H, \\ H \setminus G &:= \text{the set of right cosets of } H. \end{aligned}$$

///

Now before we go any further let me explain the meaning of the notation " G/H ." The following theorem is definitely not due to Lagrange, but he did prove a special case in his 1770 paper on Lagrange resolvents.

Lagrange's Theorem. Let $(G, *, \varepsilon)$ be a group and let $H \subseteq G$ be any subgroup. Then there is a bijection between any two left (or right) cosets of H . If G is **finite** it follows that

$$\#(H \setminus G) = \#(G/H) = \#G/\#H,$$

hence the number of elements of H divides the number of elements of G . [Remark: I take this as excellent motivation for the fractional notation.] ///

Proof. For each element $a \in G$ note that the surjective function $H \rightarrow aH$ defined by $h \mapsto a * h$ is invertible with inverse $g \mapsto a^{-1} * g$. Similarly, the surjective function $H \rightarrow Ha$ defined by $h \mapsto h * a$ is invertible with inverse $g \mapsto g * a^{-1}$. We have shown that each left (or right) coset is in bijection with H , hence any two cosets are in bijection with one another.

Next let us assume that G is finite. Then for all $a \in G$ the above bijections prove that

$$\#(aH) = \#H = \#(Ha).$$

Finally, since the set G is a disjoint union of left (or right) cosets, each having size $\#H$, we conclude $\#G$ equals the number of left (or right) cosets times $\#H$:

$$\#G = \#(G/H) \cdot \#H = \#(H \backslash G) \cdot \#H.$$

□

On the second homework you proved the following theorem for finite abelian groups. Now we can use Lagrange's Theorem to prove it for all finite groups.

Corollary (The Fermat-Euler-Lagrange Theorem). Let $(G, *, \varepsilon)$ be a **finite** group and let $g \in G$ be any element. Then we have

$$g^{\#G} = \varepsilon.$$

Proof. Suppose that g has order d , so that $\#\langle g \rangle = d$. Then since $\langle g \rangle \subseteq G$ is a **subgroup**, Lagrange's Theorem tells us that $\#G = dk$ for some $k \in \mathbb{Z}$. Finally, we have

$$g^{\#G} = g^{dk} = (g^d)^k = \varepsilon^k = \varepsilon.$$

□

I still haven't told you what this has to do with Fermat and Euler. Be patient.

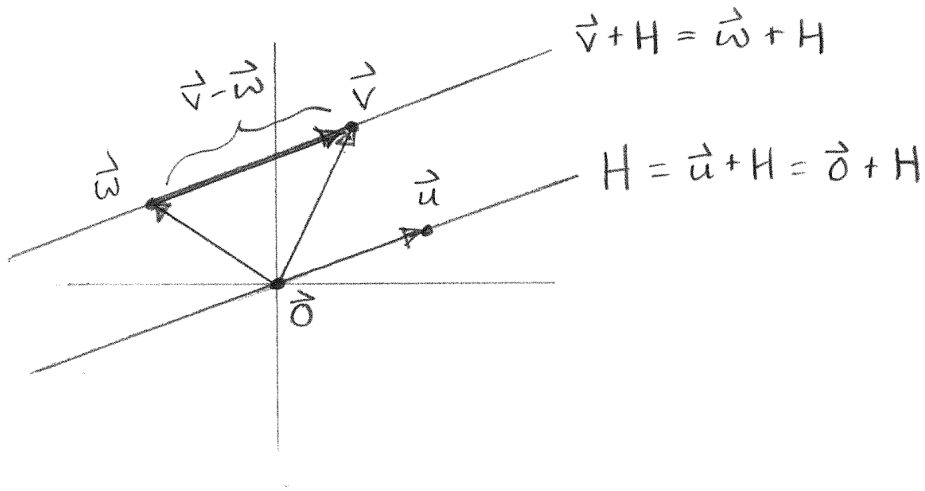
That was the theory. Now let's see some examples of cosets.

Example: Parallel Lines. Let $G = (\mathbb{R}^2, +, \mathbf{0})$ be the additive group of points in the plane, and for any nonzero vector $\mathbf{0} \neq \mathbf{u} \in \mathbb{R}^2$ let $H = \mathbb{R}\mathbf{u} := \{\alpha\mathbf{u} : \alpha \in \mathbb{R}\}$ be the line through the origin in the direction of \mathbf{u} . Note that $H \subseteq G$ is a subgroup.

Since G is abelian there is no difference between left and right cosets. We will emphasize this fact by writing the cosets additively. That is, for any vector $\mathbf{v} \in \mathbb{R}^2$ we will write

$$\mathbf{v} + H = H + \mathbf{v} = \{\mathbf{v} + h : h \in H\}.$$

The following picture shows that the cosets of H are precisely the **lines parallel to H** :



Indeed, for any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ we have by definition that

$$\mathbf{v} + H = \mathbf{w} + H \iff \mathbf{v} - \mathbf{w} \in H \iff \mathbf{v} - \mathbf{w} \text{ is parallel to } H.$$

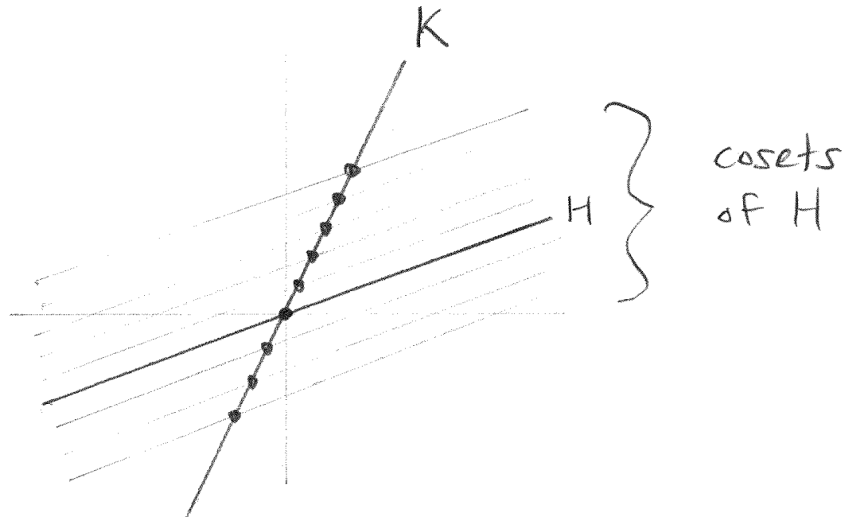
In this case the bijection $H \rightarrow \mathbf{v} + H$ defined by $\mathbf{x} \mapsto \mathbf{v} + \mathbf{x}$ is called *translation by \mathbf{v}* , and one can show that this function is in fact an **isometry**. [Exercise: Check this.] This explains why all of the cosets “look the same.” However, note that only one of the cosets (namely, H itself) is a subgroup of G because only one of the parallel lines contains the origin $\mathbf{0} \in \mathbb{R}^2$. In summary, we have

$$G/H = \mathbb{R}^2/\mathbb{R}\mathbf{u} = \text{the set of all lines parallel to } \mathbb{R}\mathbf{u}.$$

Is it possible to describe this set more efficiently? Sure. Let $K = \mathbb{R}\mathbf{v}$ be the line generated by any vector $\mathbf{v} \notin H$. Then each coset $\mathbf{w} + H$ intersects K in a unique point so we obtain a bijection $G/H \leftrightarrow K$ defined as follows:

$$\text{the line } \mathbf{w} + H \iff \text{the point of intersection } (\mathbf{w} + H) \cap K.$$

Following old Euclidean terminology we will call any such bijection a *transversal* of the cosets. Here is a picture:



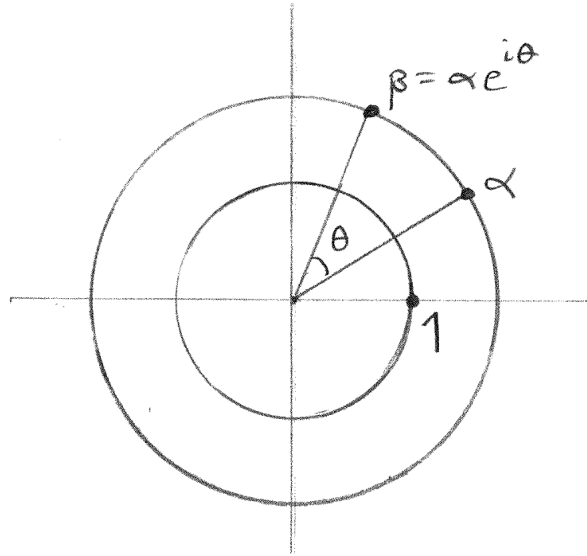
Of course the choice of the line K was arbitrary. In linear algebra it is common to let $K = H^\perp$ be the line (more generally, the complementary subspace) that is orthogonal to H . Then we obtain a bijection

$$G/H \longleftrightarrow H^\perp.$$

But note that $(H^\perp, +, 0)$ is a group. Does that mean that G/H is a group? ///

Example: Concentric Circles. Let $\mathbb{C}^\times = (\mathbb{C} - \{0\}, \times, 1)$ be the multiplicative group of nonzero complex numbers and let $U(1) = \{e^{i\theta} : \theta \in \mathbb{R}\}$ be the circle group. Note that $U(1) \subseteq \mathbb{C}^\times$ is a subgroup.

This time we will write the cosets multiplicatively, but there is still no difference between left and right cosets because \mathbb{C}^\times is abelian. The following picture shows that the cosets of $U(1)$ are precisely the **circles centered at $0 \in \mathbb{C}$** :



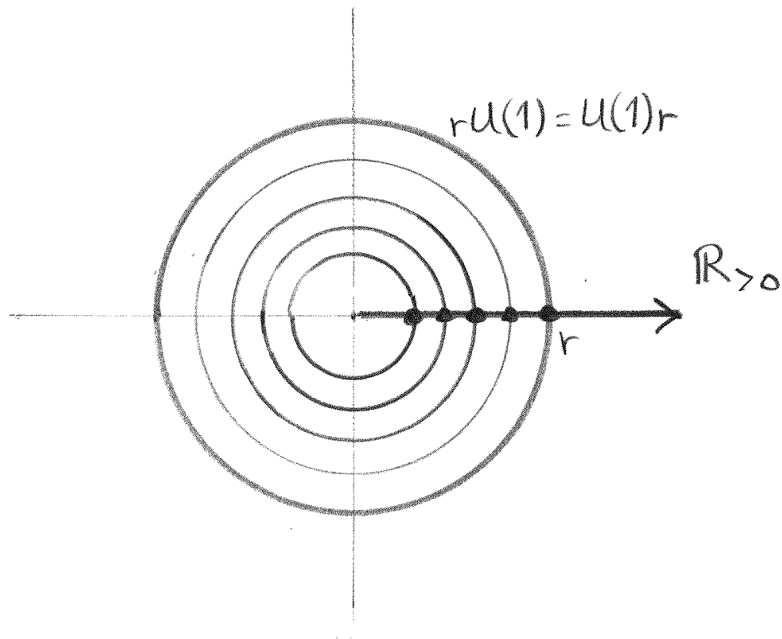
Indeed, we observe that any two nonzero numbers $\alpha, \beta \in \mathbb{C}^\times$ are in the same coset if and only if they differ by a rotation:

$$\alpha U(1) = \beta U(1) \iff \alpha^{-1}\beta \in U(1) \iff \beta = \alpha e^{i\theta} \text{ for some } \theta \in \mathbb{R}.$$

Note that any infinite ray gives rise to a transversal of the cosets. In particular, let $\mathbb{R}_{>0} = \{\alpha \in \mathbb{R} : \alpha > 0\}$ be the infinite ray of **positive real numbers**. Then we obtain a bijection between $\mathbb{C}^\times/U(1)$ and $\mathbb{R}_{>0}$ as follows:

$$\text{the circle } rU(1) = \{r e^{i\theta} : \theta \in \mathbb{R}\} \longleftrightarrow \text{the positive real number } r \in \mathbb{R}_{>0}.$$

Here is a picture:



In summary, we have a bijection

$$\mathbb{C}^\times / U(1) \longleftrightarrow \mathbb{R}_{>0}.$$

But note that $(\mathbb{R}_{>0}, \times, 1)$ is a group. Does that mean that $\mathbb{C}^\times / U(1)$ is a group? ///

Example: Modular Arithmetic. Let $\langle g \rangle$ be a cyclic group of order $n \geq 1$ and consider the group homomorphism $\varphi : \mathbb{Z}^+ \rightarrow \langle g \rangle$ defined by

$$\varphi(k) := g^k.$$

By convention the preimage of a single element is called a *fiber*. Note that the fibers of the function φ have the form

$$\begin{aligned} \varphi^{-1}[\{g^k\}] &= \{\ell \in \mathbb{Z} : \varphi(\ell) = g^k\} \\ &= \{\ell \in \mathbb{Z} : g^\ell = g^k\} \\ &= \{\ell \in \mathbb{Z} : \ell - k \in \mathbb{Z}\} \\ &= \{\ell \in \mathbb{Z} : \ell - k = nm \text{ for some } m \in \mathbb{Z}\} \\ &= \{\ell \in \mathbb{Z} : \ell = k + nm \text{ for some } m \in \mathbb{Z}\} \\ &= \{k + nm : m \in \mathbb{Z}\} \\ &= k + n\mathbb{Z}. \end{aligned}$$

In other words, we have a bijection between the cosets of the subgroup $n\mathbb{Z} \subseteq \mathbb{Z}$ and the elements of the cyclic group $\langle g \rangle$:

$$\mathbb{Z}/n\mathbb{Z} \longleftrightarrow \langle g \rangle.$$

But we know that $\langle g \rangle$ is a group. Does that mean that the set of cosets $\mathbb{Z}/n\mathbb{Z}$ is also a group?

Sure, why not? We can simply define a group structure on $\mathbb{Z}/n\mathbb{Z}$ by transferring it from $\langle g \rangle$ via the bijection. To be specific, since $g^k * g^\ell = g^{k+\ell}$ for all $k, \ell \in \mathbb{Z}$ we will define the “same operation” on the fibers:

$$(k + n\mathbb{Z}) * (\ell + n\mathbb{Z}) := (k + \ell) + n\mathbb{Z}.$$

But now the symbol “*” looks silly, so let’s replace it by “+.”

$$(k + n\mathbb{Z}) + (\ell + n\mathbb{Z}) := (k + \ell) + n\mathbb{Z}.$$

[Warning: The “+” symbol here is just an analogy. We are really “adding” two infinite sets of integers to obtain another infinite set of integers. It just happens that everything works out nicely.] In summary, the set of cosets $\mathbb{Z}/n\mathbb{Z}$ has a natural group structure which makes it isomorphic to the cyclic group $\langle g \rangle$. ///

Based on these three examples, it is not surprising that we can define a group structure on the set of cosets G/H whenever G is an abelian group. Here is the official statement.

Quotients of Abelian Groups. Let $(G, +, 0)$ be any abelian group and let $H \subseteq G$ be any subgroup. Since G is abelian the, left and right cosets of H are equal. That is, for all elements $a \in G$ we have

$$a + H = \{a + h : h \in H\} = \{h + a : h \in H\} = H + a.$$

Now we want to define “addition of cosets” so that the following equation makes sense:

$$(a + H) + (b + H) = (a + b) + H.$$

What needs to be checked? In the three examples above we knew ahead of time that everything would work out, but in the abstract setting we need to show that this operation is *well-defined*. In other words, we need to show that the definition does not depend on the particular choice of coset representatives a and b .

Proof. Suppose that $a + H = a' + H$ and $b + H = b' + H$, so that $a - a' \in H$ and $b - b' \in H$. Then since H is closed under addition we have

$$(a + b) - (a' + b') = (a - a') + (b - b') \in H,$$

and hence $(a + b) + H = (a' + b') + H$. [Remark: I used the fact that G is abelian when I switched $b - a'$ with $-a' + b$. This proof doesn’t work for non-abelian groups.] □

Having checked that “addition of cosets” is well-defined, I claim that $(G/H, +)$ is a group.

Proof. The identity element is $H = 0 + H$ since for all $a \in G$ we have

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

And the inverse of $(a + H)$ is $(-a + H)$ because

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

Finally, associativity is inherited from G because for all $a, b, c \in G$ we have

$$\begin{aligned} (a + H) + [(b + H) + (c + H)] &= (a + H) + ([b + c] + H) \\ &= (a + [b + c]) + H \\ &= ([a + b] + c) + H \\ &= ([a + b] + H) + (c + H) \\ &= [(a + H) + (b + H)] + (c + H). \end{aligned}$$

□

In summary, for every abelian group $(G, +, 0)$ and for every subgroup $H \subseteq G$ we have constructed the *quotient group* $(G/H, +, H)$. Next week we'll consider the more difficult case when G is non-abelian.

Problem Set 3

1. Order of a Power. Let G be a group and let $g \in G$ be an element of order n .

- (a) For all $k \in \mathbb{Z}$, prove that $\langle g^k \rangle = \langle g^d \rangle$ where $d = \gcd(n, k)$. [Hint: $n\mathbb{Z} + k\mathbb{Z} = d\mathbb{Z}$.]
- (b) For any positive divisor $d|n$ show that g^d has order n/d .
- (c) Combine (a) and (b) to prove that for any $k \in \mathbb{Z}$ the element g^k has order $n/\gcd(n, k)$.

(a) By definition $d = \gcd(n, k)$ is a divisor of k . In other words, we have $k = d\ell$ for some integer $\ell \in \mathbb{Z}$. This implies that

$$g^k = g^{d\ell} = (g^d)^\ell \in \langle g^d \rangle,$$

and hence $\langle g^k \rangle \subseteq \langle g^d \rangle$. For the other direction, recall that $n\mathbb{Z} + k\mathbb{Z} = d\mathbb{Z}$. Since $d \in d\mathbb{Z}$ this implies that $d \in n\mathbb{Z} + k\mathbb{Z}$ and hence there exist integers $x, y \in \mathbb{Z}$ such that $d = nx + ky$. [Recall: This is called Bézout's Identity.] Then we have

$$g^d = g^{nx+ky} = (g^n)^x * (g^k)^y = \varepsilon^x * (g^k)^y = (g^k)^y \in \langle g^k \rangle,$$

which implies that $\langle g^d \rangle \subseteq \langle g^k \rangle$.

(b) Let d be a positive divisor of n , so that $n = dm$ for some integer $m \geq 1$. It follows that

$$(g^d)^m = g^{md} = g^n = \varepsilon.$$

Our goal is to show that the m elements $\varepsilon, g^d, (g^d)^2, \dots, (g^d)^{m-1}$ are distinct. So let us assume for contradiction that there exist integers $0 \leq k < \ell \leq m - 1$ with $(g^d)^k = (g^d)^\ell$, and hence

$$\begin{aligned}(g^d)^\ell &= (g^d)^k \\ g^{d\ell} &= g^{dk} \\ g^{d(\ell-k)} &= \varepsilon.\end{aligned}$$

But then since $1 \leq \ell - k < m$ we must have $1 \leq d \leq d(\ell - k) < dm = n$, which contradicts the fact that g has order n .

(c) For any $k \in \mathbb{Z}$ we showed in part (a) that $\langle g^k \rangle = \langle g^d \rangle$, where $d = \gcd(n, k)$. Then since d is a positive divisor of n it follows from part (b) that

$$\#\langle g^k \rangle = \#\langle g^d \rangle = n/d = n/\gcd(n, k).$$

□

[Remark: This proves the conjecture that we made on the previous homework.]

2. Multiplication of Subgroups. Let $(G, *, \varepsilon)$ be a group and let $H, K \subseteq G$ be any two subgroups. Consider the Cartesian product of sets

$$H \times K := \{(h, k) : h \in H, k \in K\}$$

and the “multiplication function” $\mu : H \times K \rightarrow G$ defined by $\mu(h, k) := h * k$.

- (a) Prove that μ is injective if and only if $H \cap K = \{\varepsilon\}$.
- (b) We can think of the set $H \times K$ as an abstract group by defining

$$(h_1, k_1) * (h_2, k_2) := (h_1 * h_2, k_1 * k_2) \quad \text{for all } h_1, h_2 \in H \text{ and } k_1, k_2 \in K.$$

In this case we call $(H \times K, *)$ the *direct product* of H and K . Prove that μ is a group homomorphism if and only if we have $h * k = k * h$ for all $h \in H$ and $k \in K$.

- (c) The image of $\mu : H \times K \rightarrow G$ is the “internal product set”

$$HK := \{h * k : h \in H, k \in K\} \subseteq G.$$

Prove that $HK \subseteq G$ is a subgroup if and only if $HK = KH$.

- (a) Assume that μ is injective and let $g \in H \cap K$, hence also $g^{-1} \in H \cap K$. Now we have

$$\mu(g, g^{-1}) = g * g^{-1} = \varepsilon = \varepsilon * \varepsilon = \mu(\varepsilon, \varepsilon),$$

and it follows from injectivity that $(g, g^{-1}) = (\varepsilon, \varepsilon)$, hence $g = \varepsilon$. Conversely, let $H \cap K = \{\varepsilon\}$ and consider any pairs $(h_1, k_1), (h_2, k_2) \in H \times K$ such that $\mu(h_1, k_1) = \mu(h_2, k_2)$. Then

$$\begin{aligned}\mu(h_1, k_1) &= \mu(h_2, k_2) \\ h_1 * k_1 &= h_2 * k_2 \\ h_2^{-1} * h_1 &= k_2 * k_1^{-1}.\end{aligned}$$

Since $h_2^{-1} * h_1 \in H$ and $k_2 * k_1^{-1} \in K$ this implies that $h_2^{-1} * h_1 = k_2 * k_1^{-1} \in H \cap K$ and hence

$$h_2^{-1} * h_1 = k_2 * k_1^{-1} = \varepsilon.$$

Finally, $h_2^{-1} * h_1 = \varepsilon$ implies $h_1 = h_2$ and $k_2 * k_1^{-1} = \varepsilon$ implies $k_1 = k_2$. \square

(b) Assume that $h * k = k * h$ for all $h \in H$ and $k \in K$. Then multiplication defines a homomorphism $\mu : H \times K \rightarrow G$ because

$$\begin{aligned}\mu((h_1, k_1) * (h_2, k_2)) &= \mu(h_1 * h_2, k_1 * k_2) \\ &= (h_1 * h_2) * (k_1 * k_2) \\ &= (h_1 * k_1) * (h_2 * k_2) \\ &= \mu(h_1, k_1) * \mu(h_2, k_2).\end{aligned}$$

Conversely, assume that μ is a homomorphism. Then for all $h \in H$ and $k \in K$ we have

$$(h * k) * (h^{-1} * k^{-1}) = \mu(h, k) * \mu(h^{-1}, k^{-1}) = \mu(h * h^{-1}, k * k^{-1}) = \mu(\varepsilon, \varepsilon) = \varepsilon$$

and it follows that

$$\begin{aligned}h * k * h^{-1} * k^{-1} &= \varepsilon \\ h * k &= k * h.\end{aligned}$$

\square

(c) Assume that $HK = KH$ and consider any two elements $h_1 k_1, h_2 k_2 \in HK$. Then since $k_1 k_2^{-1} h_2^{-1} \in KH = HK$ we observe that $k_1 k_2^{-1} h_2^{-1} = h_3 k_3$ for some $h_3 \in H$ and $k_3 \in K$. It follows that

$$(h_1 k_1)(h_2 k_2)^{-1} = h_1 k_1 k_2^{-1} h_2^{-1} = h_1 h_3 k_3 \in HK,$$

and hence HK is a subgroup. Conversely, assume that HK is a subgroup. We will show that $KH \subseteq HK$ and $HK \subseteq KH$. For the first inclusion, we will show that $kh \in HK$ for all $k \in K$ and $h \in H$. Indeed, since $k = \varepsilon k \in HK$ and $h = h\varepsilon \in HK$ and since HK is a subgroup we must have $kh \in HK$. For the other inclusion, we will show that $hk \in KH$ for all $h \in H$ and $k \in K$. Indeed, since $hk \in HK$ and since HK is a subgroup we have

$$k^{-1} h^{-1} = (hk)^{-1} \in HK,$$

which implies that $k^{-1}h^{-1} = h'k'$ for some $h' \in H$ and $k' \in K$. Now take the inverse of both sides to get

$$hk = (k')^{-1}(h')^{-1} \in KH,$$

as desired. □

[Remark: In the third proof I used “juxtaposition” instead of $*$ to save space.]

3. Why Does $AB = I$ Imply $BA = I$? Given a field \mathbb{F} and a positive integer n we define

$$\mathbb{M} := \text{Mat}_n(\mathbb{F}) = \text{the set of } n \times n \text{ matrices with entries in } \mathbb{F}.$$

I claim that this set is a *vector space of dimension n^2* over the field \mathbb{F} . Now consider any two matrices $A, B \in \mathbb{M}$ such that $AB = I$.

- (a) Show that the set $B\mathbb{M} := \{BM : M \in \mathbb{M}\}$ is a *vector subspace* of \mathbb{M} . In other words, for all matrices $X, Y \in B\mathbb{M}$ and scalars $\alpha, \beta \in \mathbb{F}$, show that $\alpha X + \beta Y \in B\mathbb{M}$.
- (b) More generally, for each integer $k \geq 0$ define the set $B^k\mathbb{M} := \{B^k M : M \in \mathbb{M}\}$ and show that $B^{k+1}\mathbb{M}$ is a vector subspace of $B^k\mathbb{M}$.
- (c) I claim that a finite-dimensional vector space has no infinite descending chain of subspaces. Use this fact to prove that there exists an integer $k \geq 0$ and a matrix $C \in \mathbb{M}$ satisfying $B^k = B^{k+1}C$.
- (d) Let C be as in part (c). Prove that $BC = I$ and hence $C = A$. It follows that $BA = I$.

[Remark: Believe it or not, this is the shortest proof I know.]

(a) Consider any $X, Y \in B\mathbb{M}$, so that $X = BM$ and $Y = BN$ for some matrices $M, N \in \mathbb{M}$. Then for all scalars $\alpha, \beta \in \mathbb{F}$ we have

$$\alpha X + \beta Y = \alpha BM + \beta BN = B(\alpha M + \beta N) \in B\mathbb{M}.$$

(b) Let $k \geq 0$ and consider any element $X \in B^{k+1}\mathbb{M}$, so that $X = B^{k+1}M$ for some matrix $M \in \mathbb{M}$. It follows that $X = B^{k+1}M = B^k(BM) = B^k(BM) \in B^k\mathbb{M}$ and hence $B^{k+1}\mathbb{M} \subseteq B^k\mathbb{M}$. Then since we know from (a) that $B^{k+1}\mathbb{M}$ is a vector subspace of \mathbb{M} , it automatically follows that $B^{k+1}\mathbb{M}$ is a vector subspace of $B^k\mathbb{M}$.

(c) From part (b) we have an infinite chain of vector subspaces:

$$\mathbb{M} \supseteq B\mathbb{M} \supseteq B^2\mathbb{M} \supseteq B^3\mathbb{M} \supseteq \dots$$

If each of the inclusions were strict then we would obtain a contradiction to the finite dimensionality of \mathbb{M} . Therefore we must have $B^k\mathbb{M} = B^{k+1}\mathbb{M}$ for some $k \geq 0$. In particular, since $B^k \in B^k\mathbb{M} = B^{k+1}\mathbb{M}$, there must exist some matrix $C \in \mathbb{M}$ such that $B^k = B^{k+1}C$.

(d) Since $AB = I$ one can show by induction that $A^k B^k = I$ for all $k \geq 0$. Now multiply both sides of the equation $B^k = B^{k+1}C$ on the left by A^k to obtain

$$\begin{aligned} B^k &= B^{k+1}C \\ A^k B^k &= A^k B^k BC \\ I &= BC. \end{aligned}$$

In summary, we have shown that any matrix B with a left inverse A must also have a right inverse C . Finally, we observe that the left inverse and right inverse are equal:

$$A = AI = A(BC) = (AB)C = IC = C.$$

□

[Remark: The same proof shows that $ab = 1$ implies $ba = 1$ in any “Artinian ring.” The prototypical examples of Artinian rings are finite rings and finite dimensional algebras (for example, matrices) over a field.]

4. Conjugation is an Automorphism. Let $(G, *, \varepsilon)$ be a group and let $g \in G$ be any element. Define the function $\varphi_g : G \rightarrow G$ by $\varphi_g(a) := g * a * g^{-1}$.

- (a) Prove that $\varphi_g : G \rightarrow G$ is a bijection.
- (b) Prove that $\varphi_g : G \rightarrow G$ is a homomorphism, hence it is an *automorphism* of G .
- (c) Application: Consider any two elements $a, b \in G$. Prove that the cyclic groups $\langle a * b \rangle$ and $\langle b * a \rangle$ are isomorphic, hence the elements $a * b$ and $b * a$ have the same order.

(a) To prove that φ_g is a bijection I will show that it has an inverse. So consider the function $\psi_g : G \rightarrow G$ defined by $\psi_g(a) := g^{-1} * a * g$. Then for all $a \in G$ we have

$$\psi_g(\varphi_g(a)) = g^{-1} * (g * a * g^{-1}) * g = (g^{-1} * g) * a * (g * g^{-1}) = a$$

and

$$\varphi_g(\psi_g(a)) = g * (g^{-1} * a * g) * g^{-1} = (g * g^{-1}) * a * (g^{-1} * g) = a,$$

which by definition means that $\psi_g = \varphi_g^{-1}$.

(b) To see that $\varphi_g : G \rightarrow G$ is a homomorphism, note that for all $a, b \in G$ we have

$$\begin{aligned} \varphi_g(a) * \varphi_g(b) &= (g * a * g^{-1}) * (g * b * g^{-1}) \\ &= g * a * (g^{-1} * g) * b * g^{-1} \\ &= g * (a * b) * g^{-1} = \varphi_g(a * b). \end{aligned}$$

For posterity let me record that this implies $\varphi_g(a^n) = \varphi_g(a)^n$ for all $n \in \mathbb{Z}$. [Proof: Induction.]

(c) Now let $a, b \in G$ and consider the cyclic subgroups $\langle a * b \rangle$ and $\langle b * a \rangle$. I claim that the automorphism $\varphi_a : G \leftrightarrow G : \psi_a$ from parts (a) and (b) restricts to an isomorphism

$$\varphi_a : \langle b * a \rangle \xrightarrow{\sim} \langle a * b \rangle : \psi_a.$$

To see this, consider any integer $n \in \mathbb{Z}$ and observe from part (b) that

$$\varphi_a((b * a)^n) = \varphi_a(b * a)^n = (a * b * a * a^{-1})^n = (a * b)^n \in \langle a * b \rangle$$

and

$$\psi_a((a * b)^n) = \psi_a(a * b)^n = (a^{-1} * a * b * a)^n = (b * a)^n \in \langle b * a \rangle.$$

It follows that the two elements $a * b$ and $b * a$ have the same order. \square

5. Galois Connections. Let (P, \leq) and (Q, \leq) be posets and let $f : P \rightarrow Q$ and $g : Q \rightarrow P$ be any functions satisfying

$$p \leq g(q) \iff f(p) \leq q \quad \text{for all } p \in P \text{ and } q \in Q.$$

(a) For all $p \in P$ and $q \in Q$ prove that

$$p \leq g(f(p)) \quad \text{and} \quad f(g(q)) \leq q.$$

(b) For all $p_1, p_2 \in P$ and $q_1, q_2 \in Q$ prove that

$$p_1 \leq p_2 \Rightarrow f(p_1) \leq f(p_2) \quad \text{and} \quad q_1 \leq q_2 \Rightarrow g(q_1) \leq g(q_2).$$

(c) For all $p \in P$ and $q \in Q$ prove that

$$f(p) = f(g(f(p))) \quad \text{and} \quad g(q) = g(f(g(q))).$$

(d) Define the “images” $P' := g[Q] := \{g(q) : q \in Q\}$ and $Q' := f[P] := \{f(p) : p \in P\}$. Prove that these are the same as the sets of “closed elements”

$$P' = \{p \in P : p = g(f(p))\} \quad \text{and} \quad Q' = \{q \in Q : q = f(g(q))\}.$$

(e) Prove that the functions f, g restrict to an isomorphism of posets:

$$f : P' \longleftrightarrow Q' : g.$$

(a) For all $p \in P$ we have $f(p) \leq f(p)$, which by definition of Galois connection implies $p \leq g(f(p))$. The proof for $q \in Q$ is similar.

(b) Consider any $p_1, p_2 \in P$ such that $p_1 \leq p_2$. Then from part (a) we have $p_1 \leq p_2 \leq g(f(p_2))$ and transitivity implies $p_1 \leq g(f(p_2))$. Finally, the definition of Galois connection gives $f(p_1) \leq f(p_2)$. The proof for $q_1 \leq q_2$ is similar.

(c) Consider any $p \in P$. Then from (a) we have $p \leq g(f(p))$ and from (b) we have $f(p) \leq f(g(f(p)))$. Conversely, since $g(f(p)) \leq g(f(p))$ the definition of Galois connection gives $f(g(f(p))) \leq f(p)$, and hence $f(p) = f(g(f(p)))$. The proof for $q \in Q$ is similar.

(d) We will show that $f[P] = \{q \in Q : q = f(g(q))\}$. So consider any element $q = f(p) \in f[P]$. Then applying g and using part (c) gives $q = f(p) = f(g(f(p))) = f(g(q))$. Conversely, consider any $q \in Q$ such that $q = f(g(q))$. Then $q = f(p)$ for $p = g(q) \in P$ and hence $q \in f[P]$. The proof of $g[Q] = \{p \in P : p = g(f(p))\}$ is similar.

(e) We already know from part (a) that each of f, g is a poset homomorphism. Thus we only need to show that $f : P' \leftrightarrow Q' : g$ is a bijection. So consider any $p \in P' = g[Q]$. Then by definition we have $f(p) \in Q' = f[P]$ and from (d) we have $g(f(p)) = p$. Similarly we have for all $q \in Q'$ that $g(q) \in P'$ and $f(g(q)) = q$. This completes the proof. \square

6. Image and Preimage. Let $(G, *, \delta)$ and $(H, \bullet, \varepsilon)$ be groups and let $\varphi : G \rightarrow H$ be any group homomorphism. For every subset $S \subseteq G$ we define the *image set*

$$\varphi[S] := \{\varphi(g) : g \in S\} \subseteq H,$$

and for every subset $T \subseteq H$ we define the *preimage set*

$$\varphi^{-1}[T] := \{g \in G : \varphi(g) \in T\} \subseteq G.$$

- (a) Show that the function $\varphi^{-1} : H \rightarrow G$ exists if and only if $\#\varphi^{-1}[\{h\}] = 1$ for all $h \in H$.
- (b) If $S \subseteq G$ is a subgroup prove that the image $\varphi[S] \subseteq H$ is a subgroup.
- (c) If $T \subseteq H$ is a subgroup prove that the preimage $\varphi^{-1}[T] \subseteq G$ is a subgroup.
- (d) Now you have two functions $\varphi : \mathcal{L}(G) \leftrightarrow \mathcal{L}(H) : \varphi^{-1}$ between the subgroup lattices. Prove that this is a Galois connection.

(a) Consider a function $\varphi : G \rightarrow H$. We make two basic observations:

- The function is injective if and only if $\#\varphi^{-1}[\{h\}] \leq 1$ for all $h \in H$.
- The function is surjective if and only if $\#\varphi^{-1}[\{h\}] \geq 1$ for all $h \in H$.

Therefore the function is bijective if and only if $\#\varphi^{-1}[\{h\}] = 1$ for all $h \in H$.

(b) Let $S \subseteq G$ be a subgroup and consider any two elements $h_1, h_2 \in \varphi[S] \subseteq H$. By definition this means that $\varphi(s_1) = h_1$ and $\varphi(s_2) = h_2$ for some $s_1, s_2 \in S$ and since S is a subgroup we must have $s_1 * s_2^{-1} \in S$. But then

$$h_1 \bullet h_2^{-1} = \varphi(s_1) \bullet \varphi(s_2)^{-1} = \varphi(s_1 * s_2^{-1}) \in \varphi[S],$$

which implies that $\varphi[S] \subseteq H$ is a subgroup.

(c) Let $T \subseteq H$ be a subgroup and consider any two elements $g_1, g_2 \in \varphi^{-1}[T]$. By definition this means that $\varphi(g_1) \in T$ and $\varphi(g_2) \in T$. But then since T is a subgroup we have

$$\varphi(g_1 * g_2^{-1}) = \varphi(g_1) \bullet \varphi(g_2)^{-1} \in T,$$

which implies that $g_1 * g_2^{-1} \in \varphi^{-1}[T]$. We conclude that $\varphi^{-1}[T] \subseteq G$ is a subgroup.

(d) I proved this in the notes for the image and preimage of subsets. The same proof applies to subgroups.

Week 7

Let me recall the last proof we did using more generic language. If $(G, *, \varepsilon)$ is a group and if $H \subseteq G$ is any subgroup, then it seems natural to define the following operation on left cosets:

$$(aH) * (bH) := (a * b)H \quad \text{for all } a, b \in G.$$

However, we need to be careful because this definition is stated in terms of non-unique representatives of equivalence classes. To make sure there is no logical contradiction we must check that $a_1H = a_2H$ and $b_1H = b_2H$ imply $(a_1 * b_1)H = (a_2 * b_2)H$.

Check. Assume that $a_1H = a_2H$ and $b_1H = b_2H$, which by definition means that $a_1^{-1}a_2 \in H$ and $b_1^{-1}b_2 \in H$. In this case we want to show that $(a_1 * b_1)H = (a_2 * b_2)H$, which by definition means that $(a_1 * b_1)^{-1} * (a_2 * b_2) \in H$. If G is **abelian** then we have

$$\begin{aligned} (a_1 * b_1)^{-1} * (a_2 * b_2) &= b_1^{-1} * [(a_1^{-1} * a_2) * b_2] \\ &= b_1^{-1} * [b_2 * (a_1^{-1} * a_2)] \\ &= (b_1^{-1} * b_2) * (a_1^{-1} * a_2) \in H \end{aligned}$$

because H is closed under the operation “*.”

///

If G is **non-abelian** then we might have

$$(a_1 * a_2)^{-1} * b_2 \neq b_2 * (a_1^{-1} * a_2),$$

which seems to break the proof. But all is not lost. If there exists some $h \in H$ such that

$$(a_1^{-1} * a_2) * b_2 = b_2 * h$$

then the operation “*” on cosets is still well-defined because

$$\begin{aligned} (a_1 * b_1)^{-1} * (a_2 * b_2) &= b_1^{-1} * [(a_1^{-1} * a_2) * b_2] \\ &= b_1^{-1} * [b_2 * h] \\ &= (b_1^{-1} * b_2) * h \in H. \end{aligned}$$

///

To paraphrase: The proof still works if

for all $h \in H$ and $g \in G$ there exists some $h' \in H$ such that $h * g = g * h'$.

This strange kind of subgroup was defined by Galois all the way back in 1830, who apparently called them “invariant” subgroups. Today we call them “normal.”

Theorem (Definition of Normal Subgroups). Let $(G, *, \varepsilon)$ be a group and let $H \subseteq G$ be any subgroup. Then the following three statements are equivalent:

(N1) Left and right equivalence mod H are the same relation. In other words, the partitions of G into left and right cosets of H are the same:

$$G/H = H \backslash G.$$

(N2) For all $g \in G$ the left and right cosets containing g are equal:

$$gH = Hg.$$

(N3) For all $g \in G$ and $h \in H$ we have

$$g * h * g^{-1} \in H.$$

In other words, the subgroup H is closed under conjugation by elements of G .

Any subgroup $H \subseteq G$ satisfying one (and hence all) of these conditions is called *normal*. In this case we will use the notation

$$H \trianglelefteq G.$$

///

[Remark: Condition (N3) is the standard textbook definition of “normal,” and it is usually the easiest condition to check.]

Proof. We will show that (N1) \Rightarrow (N2) \Rightarrow (N3) \Rightarrow (N1).

(N1) \Rightarrow (N2): Assume that $G/H = H \backslash G$ and consider any element $g \in G$. Since $gH \in G/H$ we must also have $gH \in H \backslash G$, which means that $gH = Ha$ for some $a \in G$. Then since $g = g * \varepsilon \in gH$ we must have $g \in Ha$. In other words, the right cosets Hg and Ha both contain the element g . Finally, since non-equal cosets are disjoint this implies that $Hg = Ha$. We conclude that

$$gH = Ha = Hg.$$

(N2) \Rightarrow (N3): Assume that $gH = Hg$ for all $g \in G$. Then for all $g \in G$ and $h \in H$, since $g * h \in gG$ we must have $g * h \in Hg$. In other words, there exists some $h' \in H$ such that $g * h = h' * g$ and we conclude that

$$\begin{aligned} g * h &= h' * g \\ g * h * g^{-1} &= h' \in H. \end{aligned}$$

(N3) \Rightarrow (N1): Assume that for all $g \in G$ and $h \in H$ we have $g * h * g^{-1} \in H$. Our goal is to show that left and right equivalence mod H are the same relation on G . In other words, for all $a, b \in G$ we want to prove that

$$a^{-1} * b \in H \iff b * a^{-1} \in H.$$

For one direction, assume that $a^{-1} * b = h \in H$. Then conjugating by $a \in G$ gives

$$b * a^{-1} = a * (a^{-1} * b) * a^{-1} = a * h * a^{-1} \in H.$$

For the other direction, assume that $b * a^{-1} = h' \in H$. Then conjugating by $a^{-1} \in G$ gives

$$a^{-1} * b = a^{-1} * (b * a^{-1}) * (a^{-1})^{-1} = a^{-1} * h' * (a^{-1})^{-1} \in H.$$

□

Important Example: Every subgroup of an abelian group is normal.

Smallest Non-Example: Consider the smallest non-abelian group, which sometimes is called the symmetric group S_3 and at other times is called the dihedral group D_6 . Today we will call it D_6 . Recall that this group has a specific representation

$$D_6 = \{I, R, R^2, F, RF, R^2F\},$$

where $R = R_{2\pi/3}$ is rotation counterclockwise by $2\pi/3$ and $F = F_0$ is reflection across the x -axis. In other words:

$$R = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix} \quad \text{and} \quad F = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

One can check directly from the matrices that $RF = R^2F$ and $FR = R^2F$. To see this geometrically we should think of R^2, R as clockwise and counterclockwise rotation of an equilateral triangle and we should think of F as flipping the triangle over. Note that flipping the triangle and then rotating clockwise is the same as first rotating counterclockwise and then flipping the triangle. More generally, for any angle θ we have

$$\begin{aligned} (\text{rotate clockwise by } \theta) \circ (\text{flip}) &= (\text{flip}) \circ (\text{rotate counterclockwise by } \theta) \\ (R_\theta)^{-1}F &= FR_\theta. \end{aligned}$$

Now consider the cyclic subgroup $\langle R \rangle = \{I, R, R^2\} \subseteq D_6$. By Lagrange's Theorem we have

$$\#(\langle R \rangle \backslash D_6) = \#(D_6 / \langle R \rangle) = \#D_6 / \#\langle R \rangle = 2,$$

which tells us that there are two left cosets and two right cosets. Furthermore, since $\langle R \rangle$ itself is both a left **and** a right coset, it follows (somewhat accidentally) that

$$D_6 / \langle R \rangle = \{\{I, R, R^2\}, \{F, RF, R^2F\}\} = \langle R \rangle \backslash D_6.$$

In other words, $\langle R \rangle \trianglelefteq D_6$ is a normal subgroup. [Exercise: The same counting argument shows that $H \trianglelefteq G$ whenever $\#G/\#H = 2$.]

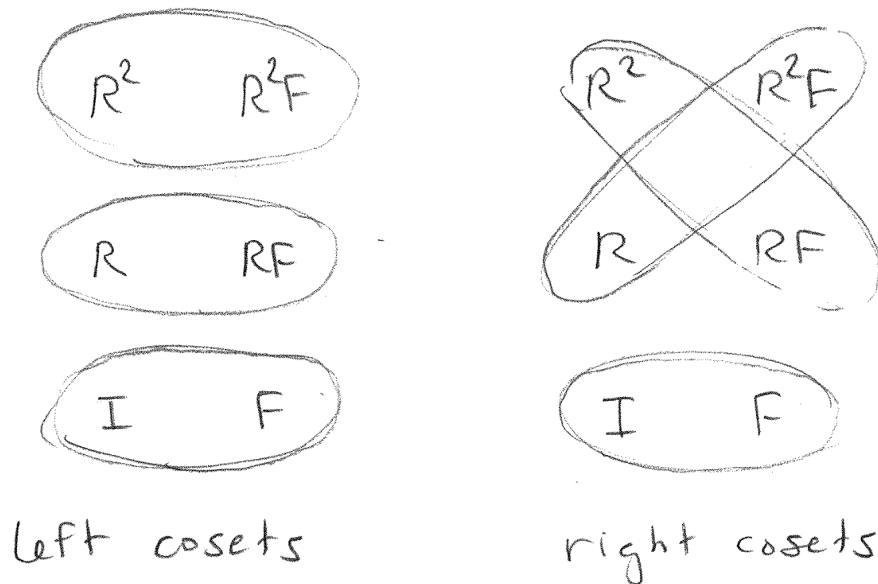
Now for the non-example. Consider the cyclic subgroup $\langle F \rangle = \{I, F\} \subseteq D_6$. Since $\#D_6/\#\langle F \rangle = 3$ it follows again from Lagrange's Theorem that there are three left cosets and three right cosets. With a little thought, the left cosets are

$$\begin{aligned} D_6 / \langle F \rangle &= \{\langle F \rangle, R\langle F \rangle, R^2\langle F \rangle\} \\ &= \{\{I, F\}, \{R, RF\}, \{R^2, R^2F\}\} \end{aligned}$$

and the right cosets are

$$\begin{aligned} \langle F \rangle \backslash D_6 &= \{\langle F \rangle, \langle F \rangle R, \langle F \rangle R^2\} \\ &= \{\{I, F\}, \{R, FR\}, \{R^2, FR^2\}\}. \end{aligned}$$

But recall that $RF = R^2F$ and $FR = R^2F$. It follows that the partitions into left and right cosets are not the same:



In other words, the subgroup $\langle F \rangle \subseteq D_6$ is **not normal**. This is the smallest possible example of a non-normal subgroup. [Exercise: Work out the details of this example in the language of the symmetric group S_3 . Hint: Let $R = (123)$ and $F = (12)$.]

The definition of normal subgroups above might have seemed a bit arbitrary and hard to remember. Today I'll show you that it's really a natural concept.

Theorem (Definition of Quotient Groups). Let $(G, *, \varepsilon)$ be a group and let $H \subseteq G$ be any subgroup. Then the following are equivalent:

- (N) $H \trianglelefteq G$ is normal,
- (N4) H is the kernel of a group homomorphism $\varphi : G \rightarrow G'$.

///

Proof. Let $(G', *, \varepsilon')$ be any group and let $\varphi : G \rightarrow G'$ be any group homomorphism. Recall that the kernel is defined as the preimage of the trivial subgroup $\{\varepsilon'\}$:

$$\ker \varphi = \varphi^{-1}[\{\varepsilon'\}] = \{g \in G : \varphi(g) = \varepsilon'\}.$$

To show that $\ker \varphi \subseteq G$ is normal, consider any elements $g \in G$ and $h \in \ker \varphi$. From general properties of homomorphisms we have

$$\begin{aligned} \varphi(g * h * g^{-1}) &= \varphi(g) *' \varphi(h) *' \varphi(g)^{-1} \\ &= \varphi(g) *' \varepsilon' *' \varphi(g)^{-1} \\ &= \varphi(g) *' \varphi(g)^{-1} \\ &= \varepsilon' \end{aligned}$$

and hence $g * h * g^{-1} \in \ker \varphi$. It follows from condition (N3) above that $\ker \varphi \trianglelefteq G$ is a normal subgroup.

Conversely, suppose that $H \trianglelefteq G$ is normal. In this case we want to define a group G' and a group homomorphism $\varphi : G \rightarrow G'$ such that $\ker \varphi = H$. This might be difficult if we were doing it from scratch, but luckily I set up all the ingredients in the previous lectures. The idea is to let G' be the set of left (or right) cosets of H :

$$G' = G/H \quad (= H \backslash G).$$

Since $H \trianglelefteq G$ is normal we have seen that the following operation on cosets is well-defined:

$$(aH) * (bH) := (a * b)H.$$

[Exercise: Check it again if you don't remember the proof.] Furthermore, we have seen that this operation makes G/H into a group with identity element $\varepsilon H = H$. To complete the proof we only need to find a group homomorphism $\varphi : G \rightarrow G/H$ such that $\ker \varphi = H$, and the choice is completely obvious: for all $g \in G$ we define

$$\varphi(g) := gH.$$

This function is a group homomorphism **by definition** and the kernel is

$$\ker \varphi = \{g \in G : gH = H\} = H.$$

□

Remarks:

- If I were teaching this course for graduate students I would probably take this as the **definition** of normal subgroups, and derive the properties (N1), (N2), (N3) as theorems.
- The homomorphism φ in the proof is called *canonical* because there is only one possible choice. It is important to remember that the quotient group is really a pair $(G/H, \varphi)$, where G/H is the group and $\varphi : G \rightarrow G/H$ is the *canonical surjection*.
- If $\varphi : G \rightarrow G'$ is **any** surjective homomorphism then we will see below that φ is secretly the canonical surjection onto the quotient group $G/\ker \varphi$.

Example: “Special” Matrix Groups. Every kind of matrix group has a “special” subgroup. I claim that these subgroups are normal. For example:

$$\begin{aligned}SL_n(\mathbb{F}) &\trianglelefteq GL_n(\mathbb{F}), \\SO(n) &\trianglelefteq O(n), \\SU(n) &\trianglelefteq U(n).\end{aligned}$$

Proof. Let G be a group of square matrices over a field \mathbb{F} and recall that the determinant satisfies $\det(AB) = \det(A)\det(B)$ for all $A, B \in G$. In other words, the determinant is a group homomorphism from G into the multiplicative group of nonzero elements of \mathbb{F} :

$$\det : G \rightarrow \mathbb{F}^\times = (\mathbb{F} - \{0\}, \times, 1).$$

It follows that the kernel of the determinant is a normal subgroup. By definition we call this the “special” subgroup:

$$SG := \ker(\det) = \{A \in G : \det(A) = 1\} \trianglelefteq G.$$

□

On the homework you will use the “same proof” to show that the group of alternating permutations is a normal subgroup of the symmetric group:

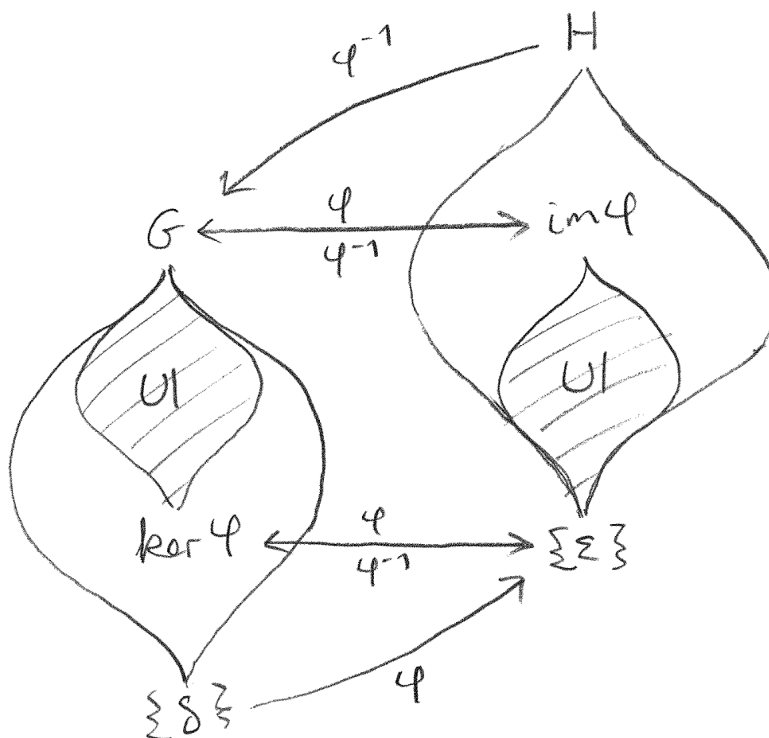
$$A_n \trianglelefteq S_n.$$

How should one visualize a group homomorphism? I have two pictures in my mind. We have already discussed one of them.

The Lattice Picture of a Group Homomorphism. If $\varphi : G \rightarrow H$ is a homomorphism of groups then we have seen that there is an isomorphism (called a “Galois Correspondence”)

$$\varphi : \mathcal{L}(G, \ker \varphi) \xrightarrow{\sim} \mathcal{L}(\text{im } \varphi) : \varphi^{-1}$$

between the the lattice $\mathcal{L}(\text{im } \varphi)$ of subgroups of the image and the lattice $\mathcal{L}(G, \ker \varphi)$ of subgroups of G that contain the kernel:



///

Today I will give you a second picture, related to the “fibers” of the homomorphism.

Definition of Fibers. Let G, H be sets and let $\varphi : G \rightarrow H$ be any function. Recall that for each subset $T \subset H$ we define the *preimage* as follows:

$$\varphi^{-1}[T] = \{g \in G : \varphi(g) \in T\} \subseteq G.$$

If the set $T = \{h\}$ contains just one element $h \in H$ then we prefer to call this the *fiber over h* :

$$\varphi^{-1}(h) := \varphi^{-1}[\{h\}] \subseteq G.$$

Warning: This notation does not imply that the inverse function exists. In fact the inverse function exists if and only if each fiber contains a single element:

$$\varphi^{-1} : H \rightarrow G \text{ exists} \iff \#\varphi^{-1}(h) = 1 \text{ for all } h \in H.$$

In this sense the “fiber function” $\varphi^{-1} : H \rightarrow 2^G$ from elements of H to subsets of G is a generalization of the concept of “inverse.” ///

For a general function the fibers can be strange but the fibers of a group homomorphism are particularly nice.

Lemma (Fibers of a Group Homomorphism). Let $(G, *, \delta)$ and $(H, \bullet, \varepsilon)$ be groups and let $\varphi : G \rightarrow H$ be a group homomorphism. Then for all $g \in G$ we have

$$\varphi^{-1}(\varphi(g)) = g(\ker \varphi).$$

In other words: The (non-empty) **fibers** of a group homomorphism are the (left or right) **cosets of the kernel**.⁴ For all $h \in H - \text{im } \varphi$ we have $\varphi^{-1}(h) = \emptyset$ by definition. ///

Proof. Fix an element $g \in G$. Then for all elements $a \in G$ we have

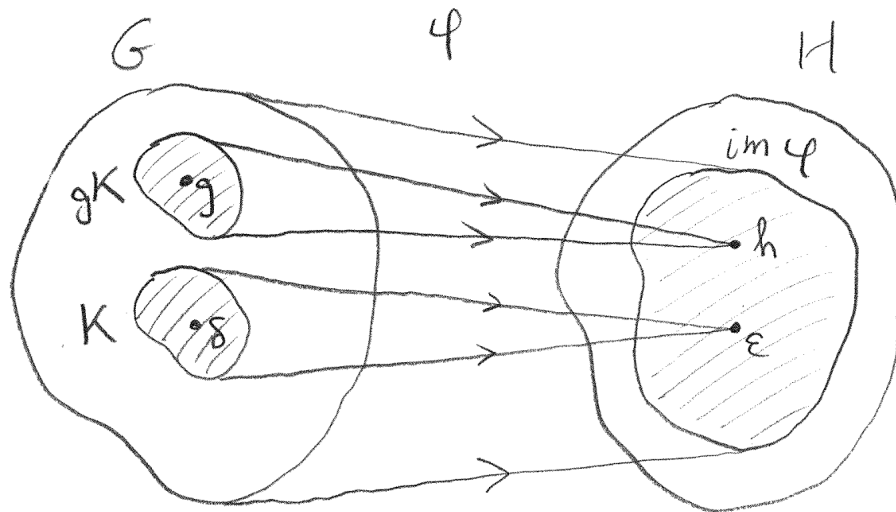
$$\begin{aligned} a \in \varphi^{-1}(\varphi(g)) &\iff \varphi(a) = \varphi(g) \\ &\iff \varphi(g)^{-1} \bullet \varphi(a) = \varepsilon \\ &\iff \varphi(g^{-1} * a) = \varepsilon \\ &\iff g^{-1} * a \in \ker \varphi \\ &\iff a \in g(\ker \varphi). \end{aligned}$$

□

Here is the picture.

The Fiber Picture of a Group Homomorphism. Let $\varphi : G \rightarrow H$ be a group homomorphism with kernel $K = \ker \varphi$. Instead of thinking of the lattice of subgroups, I will think of G and H as sets of points. For each $h = \varphi(g)$ in the image the fiber is the coset gK , and for each h not in the image the fiber is empty:

⁴Recall that kernels are normal subgroups, so there is no difference between left and right cosets.



///

But the image is a group and the set of cosets of the kernel is also a group (because the kernel is normal). Thus we obtain the following basic theorem.

Theorem (First Isomorphism Theorem). Let $\varphi : G \rightarrow H$ be a group homomorphism. Then the fiber function $\varphi^{-1} : H \rightarrow 2^G$ restricts to a group isomorphism $\text{im } \varphi \cong G / \ker \varphi$:

$$\varphi^{-1} : \text{im } \varphi \xrightarrow{\cong} G / \ker \varphi.$$

Proof. If $h \in \text{im } \varphi$ then we have $h = \varphi(g)$ for some $g \in G$ and it follows from the lemma that $\varphi^{-1}(h) = g(\ker \varphi)$ is a coset of the kernel. We need to show that this function is injective, surjective and a homomorphism.

- **Injective.** For all $h_1, h_2 \in \text{im } \varphi$ there exist $g_1, g_2 \in G$ such that $h_1 = \varphi(g_1)$ and $h_2 = \varphi(g_2)$. Then we have

$$\begin{aligned} \varphi^{-1}(h_1) = \varphi^{-1}(h_2) &\iff g_1(\ker \varphi) = g_2(\ker \varphi) \\ &\iff g_1^{-1} * g_2 \in \ker \varphi \\ &\iff \varphi(g_1^{-1} * g_2) = \epsilon \\ &\iff \varphi(g_1) = \varphi(g_2) \\ &\iff h_1 = h_2. \end{aligned}$$

- **Surjective.** This is true by definition.

- **Homomorphism.** For all $\varphi(a), \varphi(b) \in \text{im } \varphi$, the lemma says that

$$\begin{aligned} \varphi^{-1}(\varphi(a)) * \varphi^{-1}(\varphi(b)) &= a(\ker \varphi) * b(\ker \varphi) \\ &= (a * b) \ker \varphi \\ &= \varphi^{-1}(\varphi(a * b)) \\ &= \varphi^{-1}(\varphi(a) \bullet \varphi(b)). \end{aligned}$$

□

Now here is a summary of everything we know about group homomorphisms.

Summary. For any group homomorphism $\varphi : G \rightarrow H$ we have

- (1) an isomorphism of posets $\varphi : \mathcal{L}(G, \ker \varphi) \xrightarrow{\sim} \mathcal{L}(\text{im } \varphi)$,
- (2) an isomorphism of groups $\varphi^{-1} : \text{im } \varphi \xrightarrow{\sim} G / \ker \varphi$.

Furthermore, the isomorphism of groups (2) induces

- (3) an isomorphism of posets $\mathcal{L}(\text{im } \varphi) \xrightarrow{\sim} \mathcal{L}(G / \ker \varphi)$.

To be specific, we know from (1) that each subgroup of $\text{im } \varphi$ has the form $\varphi[K]$ for some unique subgroup $\ker \varphi \subseteq K \subseteq G$. The map (3) sends this to the following subgroup of $G / \ker \varphi$:

$$K / \ker \varphi = \{k(\ker \varphi) : k \in K\} \subseteq G / \ker \varphi.$$

Finally, by composing (1) and (3) we obtain an isomorphism from $\mathcal{L}(G, \ker \varphi)$ to $\mathcal{L}(G / \ker \varphi)$:

$$\begin{array}{ccccc} \mathcal{L}(G, \ker \varphi) & \xrightarrow{\sim} & \mathcal{L}(\text{im } \varphi) & \xrightarrow{\sim} & \mathcal{L}(G / \ker \varphi) \\ K & \mapsto & \varphi[K] & \mapsto & K / \ker \varphi. \end{array}$$

In other words, every subgroup of $G / \ker \varphi$ has the form $K / \ker \varphi$ for some unique subgroup $\ker \varphi \subseteq K \subseteq G$, and this correspondence preserves order. ///

Historical Remark: This is the post-1930 “modern” view of group theory. Apparently these “isomorphism theorems” emerged from the lectures of Emil Artin and Emmy Noether at Göttingen, which were then immortalized by Bartel van der Waerden in his textbook *Moderne Algebra* (1930). There are a couple more decorations we could add to this picture (namely, the Second and Third Isomorphism Theorems) but I will save those for future homework problems. For now, just a quick example.

Example: Cyclic Groups. Let G be a group and let $g \in G$ be any element. Then we have a group homomorphism from the additive integers:

$$\begin{array}{ccc} \varphi : \mathbb{Z} & \rightarrow & G \\ k & \mapsto & g^k. \end{array}$$

The image is (by definition) the cyclic subgroup $\langle g \rangle \subseteq G$ and the kernel, being a subgroup of \mathbb{Z} , has the form $n\mathbb{Z}$ some unique integer $n \geq 0$. Thus we obtain an isomorphism of groups

$$\langle g \rangle = \text{im } \varphi \cong \mathbb{Z} / \ker \varphi = \mathbb{Z} / n\mathbb{Z},$$

and two isomorphisms of lattices:

$$\begin{array}{ccccc} \mathcal{L}(\mathbb{Z}, n\mathbb{Z}) & \xrightarrow{\sim} & \mathcal{L}\langle g \rangle & \xrightarrow{\sim} & \mathcal{L}(\mathbb{Z}/n\mathbb{Z}) \\ d\mathbb{Z} & \mapsto & \langle g^d \rangle & \mapsto & d\mathbb{Z}/n\mathbb{Z}. \end{array}$$

The elements of the leftmost lattice (and hence all three lattices) are in bijection with the set of divisors $\text{Div}(n) = \{d \geq 0 : d|n\}$, which is finite for $n \geq 1$ and infinite for $n = 0$.

Remark: From the isomorphism $\langle g \rangle \cong \mathbb{Z}/n\mathbb{Z}$ we obtain bijections between the corresponding subgroups. If $n \geq 1$ then it follows for all $d \in \text{Div}(n)$ that

$$\#(d\mathbb{Z}/n\mathbb{Z}) = \#\langle g^d \rangle = n/d.$$

What happens if you try to use Lagrange's Theorem?

I don't want to end it like that. The First Isomorphism Theorem is so important that we should examine it from every angle. Today I'll show you the less sophisticated point of view. This is the kind of thing that appears on exams.

The Less Sophisticated Point of View. Let $\varphi : G \rightarrow H$ be a group homomorphism. It follows from general properties of image and preimage that $\ker \varphi \subseteq G$ and $\text{im } \varphi \subseteq H$ are subgroups, but let's prove it anyway.

Proof. For all $h_1, h_2 \in \text{im } \varphi$ there exist $g_1, g_2 \in G$ such that $\varphi(g_1) = h_1$ and $\varphi(g_2) = h_2$. Thus

$$h_1 h_2^{-1} = \varphi(g_1) \varphi(g_2)^{-1} = \varphi(g_1 g_2^{-1}) \in \text{im } \varphi.$$

For all $g_1, g_2 \in \ker \varphi$ we have $\varphi(g_1) = \varphi(g_2) = \varepsilon_H$. But then

$$\varphi(g_1 g_2^{-1}) = \varphi(g_1) \varphi(g_2)^{-1} = \varepsilon_H \varepsilon_H^{-1} = \varepsilon_H,$$

which implies that $g_1 g_2^{-1} \in \ker \varphi$. □

Now the homomorphism $\varphi : G \rightarrow H$ may not be bijective. To be specific, we observe that φ is surjective if and only if $\text{im } \varphi = H$ and φ is injective if and only if $\ker \varphi = \{\varepsilon_G\}$.

Proof. The first statement is obvious. For the second statement, assume that φ is injective. Then for all $g \in \ker \varphi$ we have $\varphi(g) = \varepsilon_H = \varphi(\varepsilon_G)$ and hence $g = \varepsilon_G$. Conversely, suppose that $\ker \varphi = \{\varepsilon_G\}$. Then for all $g_1, g_2 \in G$ we have

$$\varphi(g_1) = \varphi(g_2) \implies \varphi(g_1 g_2^{-1}) = \varepsilon_H$$

$$\begin{aligned} &\implies g_1 g_2^{-1} \in \ker \varphi \\ &\implies g_1 g_2^{-1} = \varepsilon_G \\ &\implies g_1 = g_2. \end{aligned}$$

□

Thus we can force $\varphi : G \rightarrow H$ to be **surjective** by restricting the codomain to the image:

$$\varphi : G \rightarrow \text{im } \varphi.$$

To make φ **injective** we need to “kill the kernel.” It turns out that the quotient group construction is the unique way to do this:

$$\begin{aligned} \varphi : G/\ker \varphi &\rightarrow \text{im } \varphi \\ g(\ker \varphi) &\mapsto \varphi(g). \end{aligned}$$

Assuming that the quotient group $G/\ker \varphi$ exists (i.e., that the operation on cosets is well-defined) this map is by definition a surjective group homomorphism. Furthermore, the map is well-defined and injective because

$$g_1(\ker \varphi) = g_2(\ker \varphi) \iff g_1^{-1}g_2 \in \ker \varphi \iff \varphi(g_1^{-1}g_2) = \varepsilon_H \iff \varphi(g_1) = \varphi(g_2).$$

Thus we obtain the First Isomorphism Theorem again:

$$G/\ker \varphi \cong \text{im } \varphi.$$

Week 8

Last week we saw some of the main theorems of “modern” group theory as it existed in the 1930s. The point of this theory is to prove theorems at the greatest possible level of generality as a way of compactifying our knowledge into a small conceptual space. Of course, there is no reason to do this unless we have a large stock of interesting examples.

Definition of Automorphism Groups. Let X be any “set with structure.” For example, X could be a topological space, a manifold, a vector space, a group/ring/field, or any kind of mathematical structure. By an *automorphism of X* we mean any invertible function $f : X \rightarrow X$ such that f and f^{-1} both “preserve the structure of X .” (You’ll see what this means in the examples below.) We denote the set of automorphisms by

$$\text{Aut}(X) = \{\text{invertible } f : X \rightarrow X \text{ such that } f, f^{-1} \text{ preserve the structure of } X\}.$$

It follows directly from the definition that $(\text{Aut}(X), \circ, \text{id})$ is a group under composition, with identity given by the identity function $\text{id} : X \rightarrow X$. ///

Example: Permutations. Let X be just a set (i.e., with no extra structure). Then the automorphisms of X are called *permutations*. In this case we use the notation

$$\text{Aut}(\text{set } X) = \text{Perm}(X) = S_X.$$

[The S is for *symmetric group*, which is another name for this group.] If the set X is finite with $\#X = n$ then we might as well say that $X = \{1, 2, \dots, n\}$, in which case we have

$$S_X = S_{\{1,2,\dots,n\}} = S_n.$$

Historical Remark: Prior to 1880s the word “group” was (almost) exclusively applied to groups of permutations. The first textbook on the subject was Camille Jordan’s *Traité des Substitutions* (1870). Here “substitution” means a permutation of the inputs of a multivariable function. The key fact (going back to Galois) is that the collection of substitutions that leave a given function invariant is a subgroup of S_n . The notion of an abstract group had been studied by Arthur Cayley in a series of three papers: *On the Theory of Groups, as depending on the Symbolic Equation $\theta^n = 1$* (1854 and 1859). However, these papers were ignored until Cayley republished them in 1878. ///

Example: Matrices. Let V be a vector space over a field \mathbb{F} . (See the homework for an axiomatic definition.) Homomorphisms of vector spaces are called *linear functions* and the group of automorphisms of V is called the *general linear group* of V :

$$\text{Aut}(\text{vector space } V) = GL(V).$$

Now suppose that V has dimension n . Given a basis $\mathcal{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subseteq V$ we can represent each vector $\mathbf{x} \in V$ as an $n \times 1$ column by defining

$$[\mathbf{x}]_{\mathcal{U}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n \iff \mathbf{x} = \sum_i x_i \mathbf{u}_i.$$

Then for each linear function $f : V \rightarrow V$ we define the $n \times n$ matrix $[f]_{\mathcal{U}} \in \text{Mat}_n(\mathbb{F})$ whose j -th column is $[f(\mathbf{u}_j)]_{\mathcal{U}}$ and it follows from linearity that

$$[f(\mathbf{x})]_{\mathcal{U}} = [f]_{\mathcal{U}} [\mathbf{x}]_{\mathcal{U}} \quad \text{for all } \mathbf{x} \in V.$$

In summary, the basis \mathcal{U} gives us an identification of the group $GL(V)$ of linear automorphisms with the group of $n \times n$ invertible matrices over \mathbb{F} :

$$\text{Aut}(\text{vector space } V \text{ with a fixed basis } \mathcal{U}) = GL_n(\mathbb{F})$$

However, there is no canonical choice of basis. If $\mathcal{U} \subseteq V$ and $\mathcal{V} \subseteq V$ are two bases for the vector space V and if $f : V \rightarrow V$ is a linear function then I claim that

$$C[f]_{\mathcal{U}} C^{-1} = [f]_{\mathcal{V}},$$

where $C \in \text{Mat}_n(\mathbb{F})$ is the (unique, invertible) matrix satisfying $C[\mathbf{x}]_{\mathcal{U}} = [\mathbf{x}]_{\mathcal{V}}$ for all $x \in V$.

Proof. For all $\mathbf{x} \in V$ we have

$$\begin{aligned} (C[f]_{\mathcal{U}}C^{-1})[\mathbf{x}]_{\mathcal{V}} &= C[f]_{\mathcal{U}}(C^{-1}[\mathbf{x}]_{\mathcal{V}}) \\ &= C[f]_{\mathcal{U}}[\mathbf{x}]_{\mathcal{U}} \\ &= C([f]_{\mathcal{U}}[\mathbf{x}]_{\mathcal{U}}) \\ &= C[f(\mathbf{x})]_{\mathcal{U}} \\ &= [f(\mathbf{x})]_{\mathcal{V}} \\ &= [f]_{\mathcal{V}}[\mathbf{x}]_{\mathcal{V}}. \end{aligned}$$

Then by substituting $\mathbf{x} = \mathbf{v}_j$ we see that the j -th columns of the matrices $C[f]_{\mathcal{U}}C^{-1}$ and $[f]_{\mathcal{V}}$ are equal for all j . \square

In other words, we have shown that conjugate elements of the group $GL_n(\mathbb{F})$ represent the same linear function with respect to with respect to different bases. On the homework you will show that a similar idea holds for the symmetric group. $///$

Example: Orthogonal (and Unitary) Matrices. Let V be an n -dimensional Euclidean vector space. In other words, let V be an n -dimensional vector space over \mathbb{R} , equipped with a symmetric and positive-definite bilinear form

$$\langle -, - \rangle : V \times V \rightarrow \mathbb{R}.$$

The group of automorphisms of this structure is called the *orthogonal group* of V :

$$\text{Aut}(\text{Euclidean space } V) = O(V).$$

If $\mathcal{U} \subseteq V$ is an *orthonormal basis* (consisting of orthogonal unit vectors) then for each automorphism $f : V \rightarrow V$ one can show that $[f]_{\mathcal{U}} \in \text{Mat}_n(\mathbb{R})$ is an orthogonal matrix. (In fact you already showed this on the homework.) Such a basis gives us an identification of the group $O(V)$ with the group $O(n)$ of $n \times n$ orthogonal matrices:

$$\text{Aut}(\text{Euclidean space } V \text{ with a fixed orthonormal basis } \mathcal{U}) = O(n).$$

However, there is no canonical choice of basis. In this case, conjugate elements of the group $O(n)$ represent the same linear function with respect to some orthogonal (i.e., distance preserving) change of coordinates.

More generally, if V is an n -dimensional ‘‘Hermitian space’’ over \mathbb{C} with positive-definite ‘‘sesquilinear form’’⁵ $\langle -, - \rangle : V \times V \rightarrow \mathbb{C}$ then all of the same remarks apply for the unitary groups $U(V)$ and $U(n)$. $///$

⁵A sesquilinear (one-and-a-half times linear) form satisfies $\langle \alpha \mathbf{x}, \mathbf{y} \rangle = \bar{\alpha} \langle \mathbf{x}, \mathbf{y} \rangle$ and $\langle \mathbf{x}, \alpha \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle$ for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{C}$, where $\bar{\alpha} \in \mathbb{C}$ is the complex conjugate of α .

These examples include all of the interesting kinds of (non-abelian) groups that we have studied in this course. Indeed, the subject of abstract group theory is meant to synthesize the study of concrete groups such as

$$S_n, \quad GL_n, \quad O(n), \quad U(n)$$

into one coherent theory. After studying groups from the abstract point of view, however, we might want to go back to concrete examples.

The heart's desire of an abstract group is to "act" on a nice structure.

Definition of Group Actions. Let $(G, *, \varepsilon)$ be an abstract group and let X be a set with structure. Let $G \times X \rightarrow X$ be some function written as $(g, x) \mapsto g(x)$. Equivalently, for each group element $g \in G$ we let $x \mapsto g(x)$ be an arbitrary function from X to itself. We call this function a *group action* if the following two axioms are satisfied:

(A1) The group operation acts like composition of functions:

$$(g * h)(x) = g(h(x)) \quad \text{for all } g, h \in G \text{ and } x \in X.$$

(A2) Each element of G acts like an automorphism of X :

$$\text{for all } g \in G \text{ the function } x \mapsto g(x) \text{ is in } \text{Aut}(X).$$

But there is a quicker way to say this. Equivalently, a group action is defined by a group homomorphism from G into the automorphisms of X :

$$\varphi : G \rightarrow \text{Aut}(X).$$

Proof. Given any function $(g, x) \mapsto g(x)$ satisfying (A1) and (A2) we will define $\varphi_g(x) := g(x)$. By axiom (A2) the function φ_g is in $\text{Aut}(X)$. Then by axiom (A1) we have

$$\varphi_{g*h}(x) = \varphi_g(\varphi_h(x)) = (\varphi_g \circ \varphi_h)(x) \quad \text{for all } g, h \in G \text{ and } x \in X.$$

It follows that

$$\varphi_{g*h} = \varphi_g \circ \varphi_h$$

and hence the function $\varphi : G \rightarrow \text{Aut}(X)$ sending $g \in G$ to $\varphi_g : X \rightarrow X$ is a group homomorphism. Conversely, suppose we have a group homomorphism $\varphi : G \rightarrow \text{Aut}(X)$ denoted by $\varphi \mapsto \varphi_g$. Now define a function $G \times X \rightarrow X$ by $(g, x) \mapsto \varphi_g(x)$ and observe that this function satisfies (A1) and (A2). \square

I like the homomorphism definition better because it emphasizes that a given group G can act on a given structure X in different ways, corresponding to different homomorphisms $\varphi : G \rightarrow \text{Aut}(X)$.

Remarks:

- The notation $\varphi_g(x)$ is a *simile* because it says that the group element g “acts like a function.” The notation $g(x)$ is a *metaphor* because it says that the group element g “is a function.” Of course that is not literally true.
- My definition of group action is slightly nonstandard. Most books only define the action of groups on sets, not on “sets with structure.” The standard definition says that (1) $\varepsilon(x) = x$ for all $x \in X$, and (2) $(g*h)(x) = g(h(x))$ for all $g, h \in G$ and $x \in X$. Exercise: Prove that this is equivalent to my definition when $\text{Aut}(X) = \text{Perm}(X)$, i.e., when the set X has no additional structure.
- Sometimes we use the notation $G \curvearrowright X$ to indicate that G acts on X . If we want to be specific about the homomorphism $\varphi : G \rightarrow \text{Aut}(X)$ then we can write

$$G \overset{\varphi}{\curvearrowright} X.$$

- Jargon: If V is a vector space, then an action $\varphi : G \rightarrow GL(V)$ is also called a *linear representation* of G , and the vector space V is called a *G -module*. More generally, the study of group actions is called *representation theory* by mathematicians. Physicists just call it *group theory*.

You won't appreciate the definition of group action until you understand some examples.

Example: A Group Acts on Itself in Two Ways. Let $(G, *, \varepsilon)$ be a group.

- **Translation.** For all $g \in G$ let $\tau_g : G \rightarrow G$ be the function defined by

$$\tau_g(a) := g * a \quad \text{for all } a \in G.$$

You proved on a previous homework that this function is invertible with $\tau_g^{-1} = \tau_{g^{-1}}$. Here's the proof again: For all $a \in G$ we have

$$\tau_g(\tau_{g^{-1}}(a)) = g * (g^{-1} * a) = (g * g^{-1}) * a = \varepsilon * a = a$$

and

$$\tau_{g^{-1}}(\tau_g(a)) = g^{-1} * (g * a) = (g^{-1} * g) * a = \varepsilon * a = a.$$

□

Thus we obtain a function $\tau : G \rightarrow \text{Perm}(G)$. Moreover, I claim that τ is a group homomorphism.

Proof. For all $a, g, h \in G$ we have

$$\tau_{g*h}(a) = (g * h) * a = g * (h * a) = (\tau_g \circ \tau_h)(a),$$

and hence $\tau_{g*h} = \tau_g \circ \tau_h$. □

To summarize: we say that G acts on itself (as a set) by *translation*.

- **Conjugation.** For all $g \in G$ let $\kappa_g : G \rightarrow G$ be the function defined by

$$\kappa_g(a) := g * a * g^{-1} \quad \text{for all } a \in G.$$

I claim that this function is invertible with $\kappa_g^{-1} = \kappa_{g^{-1}}$.

Proof. For all $a, g \in G$ we have

$$\kappa_g(\kappa_{g^{-1}}(a)) = g * (g^{-1} * a * g) * g^{-1} = \varepsilon * a * \varepsilon = a$$

and

$$\kappa_{g^{-1}}(\kappa_g(a)) = g^{-1} * (g * a * g^{-1}) * g = \varepsilon * a * \varepsilon = a.$$

□

Thus we obtain a function $\kappa : G \rightarrow \text{Perm}(G)$. Moreover, I claim that κ is a group homomorphism.

Proof. For all $a, g, h \in G$ we have

$$\kappa_{g*h}(a) = (g * h) * a * (g * h)^{-1} = g * (h * a * h^{-1}) * g^{-1} = (\kappa_g \circ \kappa_h)(a),$$

and hence $\kappa_{g*h} = \kappa_g \circ \kappa_h$. □

But even more is true. I claim that the image of κ is contained in the subgroup $\text{Aut}(G) \subseteq \text{Perm}(G)$ of automorphisms, i.e., the subgroup of permutations that preserve the group structure.

Proof. For all $g, a, b \in G$ we have

$$\begin{aligned} \kappa_g(a) * \kappa_g(b) &= (g * a * g^{-1}) * (g * b * g^{-1}) \\ &= g * a * (g * g^{-1}) * b * g^{-1} \\ &= g * a * \varepsilon * b * g^{-1} \\ &= g * (a * b) * g^{-1} \\ &= \kappa_g(a * b), \end{aligned}$$

and hence $\kappa_g \in \text{Aut}(G)$. □

Thus we obtain a group homomorphism $\kappa : G \rightarrow \text{Aut}(G)$, and we say that G acts on itself (as a group) by *conjugation*.

///

Remarks:

- The action of G on itself by translation does **not** preserve the group structure of G . In other words, the image of the homomorphism $\tau : G \rightarrow \text{Perm}(G)$ is **not** contained in the subgroup $\text{Aut}(G) \subseteq \text{Perm}(G)$.
- The actions τ and κ defined here are sometimes called “left translation” and “left conjugation,” and the notion of action defined above is sometimes called a “left action.” There is an associated notion of “right action,” which is defined by an **anti-homomorphism**

$$\varphi : G \rightarrow \text{Aut}(X).$$

In other words, a right action must satisfy $\varphi_{g*h} = \varphi_h \circ \varphi_g$ for all $g, h \in G$. Exercise: Define the notions of “right translation” and “right conjugation,” and prove that these are “right actions.”

Application: Cayley’s Theorem. What happens when we apply the First Isomorphism Theorem to the translation homomorphism $\tau : G \rightarrow \text{Perm}(G)$? First of all, I claim that τ is injective.

Proof. It is enough to show that $\ker \tau = \{\varepsilon\}$. So consider any $g \in \ker \tau$. By definition this means that $\tau_g : G \rightarrow G$ is the identity function:

$$\tau_g(a) = a \quad \text{for all } a \in G.$$

In particular, we have $\varepsilon = \tau_g(\varepsilon) = g * \varepsilon = g$. □

It follows that G is isomorphic to its image, which is a subgroup of $\text{Perm}(G)$:

$$G = G / \ker \tau \cong \text{im } \tau \subseteq \text{Perm}(G).$$

So what? In the 1850s the word “group” meant a “group of permutations.” When Arthur Cayley promoted an axiomatic definition of groups in 1854 he had to overcome this bias. *Cayley’s Theorem* says that every abstract group G is isomorphic to a group of permutations of some set (namely, itself). This shows that the concept of abstract group is **not more general** than the concept of permutation group. [Remark: The subgroup $\text{im } \tau \subseteq \text{Perm}(G)$ is certainly not equal to the full group, because

$$\#\text{im } \tau = \#G < (\#G)! = \#\text{Perm}(G).]$$

Application: Definition of the Center and Inner Automorphisms. If we apply the First Isomorphism Theorem to the conjugation homomorphism $\kappa : G \rightarrow \text{Aut}(G)$ then we obtain

$$G / \ker \kappa \cong \text{im } \kappa \subseteq \text{Aut}(G).$$

We have a special name for the kernel. We call it the *center* of G [German: *Zentrum*]:

$$\begin{aligned} Z(G) &:= \ker \kappa \\ &= \{g \in G : \kappa_g = \text{id}\} \\ &= \{g \in G : \kappa_g(a) = a \text{ for all } a \in G\} \\ &= \{g \in G : g * a * g^{-1} = a \text{ for all } a \in G\} \\ &= \{g \in G : g * a = a * g \text{ for all } a \in G\}. \end{aligned}$$

This is the set of elements of G that commute with everything. Being a kernel, it is necessarily a normal subgroup:

$$Z(G) \trianglelefteq G.$$

And what about the image? An automorphism of a group that arises from conjugation is called an *inner automorphism*, and we use the notation

$$\text{Inn}(G) := \text{im } \kappa \subseteq \text{Aut}(G).$$

It follows from the First Isomorphism Theorem that

$$\text{Inn}(G) \cong G/Z(G).$$

I have nothing interesting to say about this right now.

Problem Set 4

1. Permutation Matrices. Let S_n be the group of permutations of the set $\{1, 2, \dots, n\}$, and for each permutation $f \in S_n$ let $[f] \in \text{Mat}_n(\mathbb{R})$ be the matrix whose i, j -entry is 1 if $f(j) = i$ and 0 if $f(j) \neq i$.

- If $\mathbf{e}_1, \dots, \mathbf{e}_n \in \mathbb{R}^n$ is the standard basis, prove that $[f]\mathbf{e}_j = \mathbf{e}_{f(j)}$ for all $j \in \{1, \dots, n\}$.
- Use (a) to prove that the function $f \mapsto [f]$ is a group homomorphism $S_n \rightarrow O(n)$.
- Let $\det : O(n) \rightarrow \{\pm 1\}$ be the determinant. Use (b) to prove that $\varphi(f) := \det[f]$ is a group homomorphism $\varphi : S_n \rightarrow \{\pm 1\}$.
- Show that the kernel of φ is the alternating subgroup $A_n \subseteq S_n$ which was defined on the first homework. [Hint: If $t \in S_n$ is a transposition then $\varphi(t) = -1$.]
- Use the First Isomorphism Theorem and Lagrange's Theorem to conclude that

$$\#A_n = n! / 2.$$

(a) By definition, the j -th column of the matrix $[f]$ has a 1 in the $f(j)$ -th position and 0s elsewhere. In other words, $[f]\mathbf{e}_j = \mathbf{e}_{f(j)}$.

(b) For all $f, g \in S_n$ and $j \in \{1, \dots, n\}$ we will show that the j -th column of $[f \circ g]$ equals the j -th column of the matrix product $[f][g]$. This follows from repeated application of part (a) and the associativity of matrix multiplication:

$$[f \circ g]\mathbf{e}_j = \mathbf{e}_{(f \circ g)(j)} = \mathbf{e}_{f(g(j))} = [f]\mathbf{e}_{g(j)} = [f]([g]\mathbf{e}_j) = ([f][g])\mathbf{e}_j.$$

We have shown that the function $f \mapsto [f]$ is a group homomorphism from S_n to $GL_n(\mathbb{R})$, which implies that $[f^{-1}] = [f]^{-1}$ for all $f \in S_n$. It only remains to show that each matrix $[f]$ is orthogonal. In other words, we need to show that $[f^{-1}] = [f]^T$. This follows directly from the definition:

$$[f^{-1}]_{ij} = \begin{cases} 1 & \text{if } f^{-1}(j) = i \\ 0 & \text{if } f^{-1}(j) \neq i \end{cases} = \begin{cases} 1 & \text{if } j = f(i) \\ 0 & \text{if } j \neq f(i) \end{cases} = [f]_{ji} = [f]_{ij}^T.$$

(c) We assume that the determinant preserves multiplication.⁶ This implies that for all orthogonal matrices $A \in O(n)$ we have

$$1 = \det(I) = \det(A^T A) = \det(A^T) \det(A) = \det(A)^2,$$

and hence $\det(A) = \pm 1$. In other words, we have a group homomorphism $\det : O(n) \rightarrow \{\pm 1\}$, and by composing this with part (a) we obtain a homomorphism $\varphi : S_n \rightarrow \{\pm 1\}$. [Exercise: The composition of homomorphisms is a homomorphism.]

(d) We showed on a previous homework that every permutation $f \in S_n$ can be expressed (non-uniquely) as a composition of transpositions:

$$f = t_1 \circ t_2 \circ \dots \circ t_k.$$

Since each transposition has $\varphi(t) = \det[t] = -1$, this implies that

$$\varphi(f) = \varphi(t_1)\varphi(t_2)\dots\varphi(t_k) = (-1)^k.$$

We conclude that $f \in \ker \varphi$ if and only if f can be expressed as a composition of an **even** number of transpositions, i.e., if and only if $f \in A_n$.

(e) The homomorphism $\varphi : S_n \rightarrow \{\pm 1\}$ is surjective with kernel A_n . It follows from the First Isomorphism Theorem that

$$\{\pm 1\} = \text{im } \varphi \cong S_n / \ker \varphi = S_n / A_n.$$

Finally, we conclude from Lagrange's Theorem that

$$\#\{\pm 1\} = \#S_n / \#A_n$$

⁶Sorry, I'm not going to prove this.

$$\begin{aligned}\#A_n &= \#S_n / \#\{\pm 1\} \\ &= n!/2.\end{aligned}$$

□

2. Dimension of a Vector Space. Let $(\mathbb{F}, +, \times, 0, 1)$ be a field (of “scalars”) and let $(V, +, \mathbf{0})$ be an abelian group (of “vectors”). We say that V is a *vector space over* \mathbb{F} if there exists a function $\mathbb{F} \times V \rightarrow V$ denoted by $(a, \mathbf{u}) \mapsto a\mathbf{u}$ that satisfies four axioms:

- For all $\mathbf{u} \in V$ we have $1\mathbf{u} = \mathbf{u}$.
- For all $a, b \in \mathbb{F}$ and $\mathbf{u} \in V$ we have $(ab)\mathbf{u} = a(b\mathbf{u})$.
- For all $a, b \in \mathbb{F}$ and $\mathbf{u} \in V$ we have $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$.
- For all $a \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in V$ we have $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.

(a) In this case prove that $0\mathbf{u} = \mathbf{0}$ for all $\mathbf{u} \in V$ and $a\mathbf{0} = \mathbf{0}$ for all $a \in \mathbb{F}$.

(b) **Steinitz Exchange.** For all vectors $\mathbf{u}_1, \dots, \mathbf{u}_m \in V$ we define their *span* as the set

$$\mathbb{F}(\mathbf{u}_1, \dots, \mathbf{u}_m) := \{a_1\mathbf{u}_1 + \dots + a_m\mathbf{u}_m : a_1, \dots, a_m \in \mathbb{F}\} \subseteq V$$

and we say that $\mathbf{u}_1, \dots, \mathbf{u}_m$ is a *spanning set* when $\mathbb{F}(\mathbf{u}_1, \dots, \mathbf{u}_m) = V$. We say that $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ is an *independent set* if for all $b_1, \dots, b_n \in \mathbb{F}$ we have

$$(b_1\mathbf{v}_1 + \dots + b_n\mathbf{v}_n = \mathbf{0}) \Rightarrow (b_1 = \dots = b_n = 0).$$

If $\mathbf{u}_1, \dots, \mathbf{u}_m$ are spanning and $\mathbf{v}_1, \dots, \mathbf{v}_n$ are independent, prove that $n \leq m$. [Hint: Assume for contradiction that $m < n$. Since the \mathbf{u}_i are spanning we have $\mathbf{v}_1 = \sum_i a_i\mathbf{u}_i$ and since the \mathbf{v}_j are independent, not all of the coefficients are zero. Without loss suppose that $a_1 \neq 0$ and use this to show that $\mathbf{v}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$ is spanning. Now show by induction that $\mathbf{v}_1, \dots, \mathbf{v}_m$ is a spanning set and use this to obtain a contradiction.]

(c) An independent spanning set is called a *basis* of V . If V has a finite spanning set, prove that V has a finite basis.

(d) Continuing from (b) and (c), prove that any two finite bases have the same size. This size is called the *dimension* of the vector space V .

(a) For all $\mathbf{u} \in V$ we have

$$\begin{aligned}0 + 0 &= 0 \\ (0 + 0)\mathbf{u} &= 0\mathbf{u} \\ 0\mathbf{u} + 0\mathbf{u} &= 0\mathbf{u} \\ 0\mathbf{u} + 0\mathbf{u} - 0\mathbf{u} &= 0\mathbf{u} - 0\mathbf{u} \\ 0\mathbf{u} &= \mathbf{0},\end{aligned}$$

and for all $a \in \mathbb{F}$ we have

$$\begin{aligned} \mathbf{0} + \mathbf{0} &= \mathbf{0} \\ a(\mathbf{0} + \mathbf{0}) &= a\mathbf{0} \\ a\mathbf{0} + a\mathbf{0} &= a\mathbf{0} \\ a\mathbf{0} + a\mathbf{0} - a\mathbf{0} &= a\mathbf{0} - a\mathbf{0} \\ a\mathbf{0} &= \mathbf{0}. \end{aligned}$$

(b) **Steinitz Exchange.** Let $\mathbf{u}_1, \dots, \mathbf{u}_m \in V$ be a spanning set, let $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ be an independent set, and assume for contradiction that $m < n$. In this case we will prove by induction that $\mathbf{v}_1, \dots, \mathbf{v}_m \in V$ is a spanning set. Then since $m < n$ this implies that there exist coefficients $b_i \in \mathbb{F}$ such that

$$\begin{aligned} b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \dots + b_m\mathbf{v}_m &= \mathbf{v}_{m+1} \\ b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \dots + b_m\mathbf{v}_m - 1\mathbf{v}_{m+1} + 0\mathbf{v}_{m+2} + \dots + 0\mathbf{v}_n &= \mathbf{0}. \end{aligned}$$

And since $-1 \neq 0$ this contradicts the fact that the set $\mathbf{v}_1, \dots, \mathbf{v}_n$ is independent. Hence we conclude that $n \leq m$ as desired.

Proof. So let $\mathbf{u}_1, \dots, \mathbf{u}_m$ be spanning and let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be independent with $m < n$. We will show for all $k \in \{0, 1, \dots, m\}$ that it is possible to relabel the vectors \mathbf{u}_i so that

$$\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_m \in V \quad \text{is a spanning set.}$$

The statement is true when $k = 0$. So fix $0 \leq \ell < m$ and assume for induction that the statement is true when $k = \ell$. Then since $\mathbf{v}_1, \dots, \mathbf{v}_\ell, \mathbf{u}_{\ell+1}, \dots, \mathbf{u}_m$ is a spanning set there exist coefficients $a_i \in \mathbb{F}$ such that

$$\mathbf{v}_{\ell+1} = a_1\mathbf{v}_1 + \dots + a_\ell\mathbf{v}_\ell + a_{\ell+1}\mathbf{u}_{\ell+1} + \dots + a_m\mathbf{u}_m,$$

and since the \mathbf{v}_i are independent we know that the coefficients $a_{\ell+1}, \dots, a_m$ are not all zero. By relabeling the vectors \mathbf{u}_i we may assume that $a_{\ell+1} \neq 0$. Then since \mathbb{F} is a field we have

$$\mathbf{u}_{\ell+1} = -\frac{a_1}{a_{\ell+1}}\mathbf{v}_1 - \dots - \frac{a_\ell}{a_{\ell+1}}\mathbf{v}_\ell + \frac{1}{a_{\ell+1}}\mathbf{v}_{\ell+1} - \frac{a_{\ell+2}}{a_{\ell+1}}\mathbf{u}_{\ell+2} - \dots - \frac{a_m}{a_{\ell+1}}\mathbf{u}_m,$$

and it follows that $\mathbf{v}_1, \dots, \mathbf{v}_{k+1}, \mathbf{u}_{k+2}, \dots, \mathbf{u}_m$ is a spanning set. \square

(c) Suppose that $\mathbf{u}_1, \dots, \mathbf{u}_k \in V$ is a spanning set. If this set is not independent then there exists a relation

$$a_1\mathbf{u}_1 + \dots + a_k\mathbf{u}_k = \mathbf{0}$$

in which not all the coefficients are zero. By relabeling the vectors \mathbf{u}_i we may assume that $a_k \neq 0$. Then since \mathbb{F} is a field we have

$$\mathbf{u}_k = -\frac{a_1}{a_k}\mathbf{u}_1 - \frac{a_2}{a_k}\mathbf{u}_2 - \dots - \frac{a_{k-1}}{a_k}\mathbf{u}_{k-1},$$

and it follows that $\mathbf{u}_1, \dots, \mathbf{u}_{k-1}$ is a spanning set. By repeating this process as necessary we will obtain $1 \leq \ell < k$ such that $\mathbf{u}_1, \dots, \mathbf{u}_\ell$ is an independent spanning set, i.e., a basis.

(d) Let $\mathbf{u}_1, \dots, \mathbf{u}_m$ and $\mathbf{v}_1, \dots, \mathbf{v}_n$ be two bases for a vector space V . Since the \mathbf{u}_i are spanning and the \mathbf{v}_i are independent, part (b) says that $n \leq m$. Moreover, since the \mathbf{u}_i are independent and the \mathbf{v}_i are spanning, part (b) says that $m \leq n$. We conclude that $m = n$. \square

[Remark: This is the prototype for the concept of “dimension” in any area of mathematics. As you see, it is a subtle concept.]

3. Conjugacy Classes. Let G be a group and for all $a, b \in G$ define the following relation:

$$a \sim b \iff a = bg^{-1} \text{ for some } g \in G.$$

- (a) Prove that this is an equivalence relation, called *conjugacy*.
 (b) Compute the conjugacy classes for the group of symmetries of an equilateral triangle:

$$D_6 = \langle R, F \rangle = \{I, R, R^2, F, RF, R^2F\}.$$

Observe that conjugate elements “do the same thing” to the triangle.

- (c) Explicitly describe the conjugacy classes of the symmetric group S_n . [Hint: Let $f, g \in S_n$. Show that g sends i to j if and only if fgf^{-1} sends $f(i)$ to $f(j)$. What does this say about the cycle structure?]

(a) There are three things to show:

- **Reflexive.** For all $a \in G$ we have $a = \varepsilon a \varepsilon^{-1}$ and hence $a \sim a$.
- **Symmetric.** Assume that $a \sim b$ so that $a = bg^{-1}$ for some $g \in G$. Then we have $b = g^{-1}a(g^{-1})^{-1}$, which implies that $b \sim a$.
- **Transitive.** Assume that $a \sim b$ and $b \sim c$, so that $a = bg^{-1}$ and $b = hch^{-1}$ for some $g, h \in G$. It follows that

$$a = g(hch^{-1})g^{-1} = (gh)c(gh)^{-1},$$

and hence $a \sim c$. \square

(b) I’ll compute the conjugacy classes for the general dihedral group

$$D_{2n} = \langle R, F \rangle = \{I, R, \dots, R^{n-1}, F, RF, \dots, R^{n-1}F\}.$$

Recall the important fact that $RF = FR^{-1}$, and more generally that $R^kF = FR^{-k}$ for all $k \in \mathbb{Z}$. First let’s compute the conjugacy class of a rotation R^k (including the identity $R^0 = I$).

Conjugating by another rotation does nothing because the powers of R commute. Conjugating by a reflection $R^\ell F$ gives

$$(R^\ell F)R^k(R^\ell F)^{-1} = R^\ell F(R^k F)R^{-\ell} = R^\ell F(FR^{-k})R^{-\ell} = R^\ell R^{-k}R^{-\ell} = R^{-k}.$$

It follows that the conjugacy class of R^k is $\{R^k, R^{-k}\}$. These two elements “do the same thing.” Namely, they both “rotate by k/n of a full rotation.” Now we’ll compute the conjugacy class of a reflection $R^k F$. Conjugating by a rotation R^ℓ gives

$$R^\ell(R^k F)R^{-\ell} = R^\ell R^k R^\ell F = R^{k+2\ell} F,$$

and conjugating by a reflection $R^\ell F$ gives

$$\begin{aligned} (R^\ell F)(R^k F)(R^\ell F)^{-1} &= R^\ell F R^k F F R^{-\ell} \\ &= R^\ell F R^k R^{-\ell} \\ &= R^\ell F R^{k-\ell} \\ &= R^\ell R^{\ell-k} F \\ &= R^{2\ell-k} F. \end{aligned}$$

Thus there are two cases: If n is **odd** then there is one conjugacy class of reflections:

$$\{F, RF, R^2F, \dots, R^{n-1}F\}.$$

Note that each of these reflections “does the same thing.” Namely, each reflects the polygon across a line that connects a vertex to the midpoint of the opposite side. If n is **even** then there are two conjugacy classes:

$$\{F, R^2F, \dots, R^{n-2}F\} \quad \text{and} \quad \{RF, R^3F, \dots, R^{n-1}F\}.$$

One class reflects the polygon across a line through two opposite vertices, and the other class reflects across a line through the midpoints of two opposite sides. For example, if $n = 3$ (which is odd) then we obtain the following decomposition into conjugacy classes:

$$D_6 = \{I\} \cup \{R, R^2\} \cup \{F, RF, R^2F\}.$$

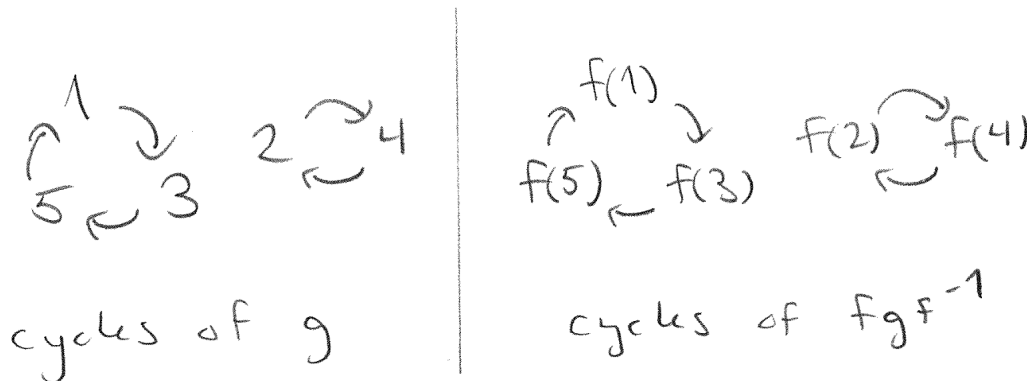
Unlike cosets, conjugacy classes **need not have the same size**.

(c) Consider any two permutations $f, g \in S_n$. Then for all $i, j \in \{1, 2, \dots, n\}$ we have

$$g(i) = j \Leftrightarrow gf^{-1}f(i) = f^{-1}f(j) \Leftrightarrow fgf^{-1}f(i) = ff^{-1}f(j) \Leftrightarrow (fgf^{-1})(f(i)) = f(j).$$

As a corollary, we see that conjugate permutations have the same “cycle structure,” i.e., they have the same number of cycles of each size.

Example: Let $g = (135)(24)$. Then for any $f \in S_5$ we have $fgf^{-1} = (f(1)f(3)f(5))(f(2)f(4))$:



4. Multiplication of Subgroups, Part II. Let G be a group and let $H, K \subseteq G$ be any two subgroups.

- (a) If at least one of H or K is normal, prove that $HK \subseteq G$ is a subgroup and hence that HK equals the join $H \vee K$. The converse is not true.
- (b) Prove that the multiplication function $\mu : H \times K \rightarrow G$ is a group isomorphism if and only if (1) H and K are both normal, (2) $H \wedge K = \{\varepsilon\}$ and (3) $H \vee K = G$. In this case we write

$$G = H \times K$$

and we say that G is the *internal direct product* of the subgroups H and K .

- (a) Let $H, K \subseteq G$ be subgroups with $H \trianglelefteq G$ normal. For all elements h_1k_2 and h_2k_2 in the set HK we want to show that $(h_1k_1)(h_2k_2)^{-1} = h_1k_1k_2^{-1}h_2^{-1}$ is also in HK . Now since H is normal we know that the left and right H -cosets generated by $k_1k_2^{-1}$ are equal:

$$k_1k_2^{-1}H = Hk_1k_2^{-1}$$

Then since the element $k_2k_1^{-1}h_2^{-1}$ is in the left coset it must also be in the right coset. In other words, there exists some element $h_3 \in H$ such that

$$k_1k_2^{-1}h_2^{-1} = h_3k_1k_2^{-1}.$$

We conclude that

$$(h_1k_1)(h_2k_2)^{-1} = h_1(k_1k_2^{-1}h_2^{-1}) = h_1(h_3k_1k_2^{-1}) = (h_1h_3)(k_1k_2^{-1}) \in HK$$

as desired. The proof for $K \trianglelefteq G$ is similar. □

[Remark: Alternatively, you could apply the result from a previous homework that $HK \subseteq G$ is a subgroup if and only if $HK = KH$. When I took abstract algebra as an undergraduate I

was asked to prove on the first (or second) exam that $H \trianglelefteq G$ implies $HK \subseteq G$ is a subgroup, and I got it wrong!]

If HK is a subgroup then I claim that $HK = H \vee K$. Indeed, since HK is a subgroup containing $H \cup K$ and since $H \vee K$ is the smallest subgroup containing $H \cup K$ we have $H \vee K \subseteq HK$. Conversely, for all $h \in H$ and $k \in K$ we have $h, k \in H \vee K$. Then since $H \vee K$ is closed under the group operation we have $hk \in H \vee K$ and it follows that $HK \subseteq H \vee K$. \square

(b) First note that (3) holds if and only if μ is surjective. Indeed, since $\text{im } \mu = HK$ we have that μ is surjective if and only if $HK = G$, which happens if and only if $H \vee K = G$. Furthermore, note that (2) holds if and only if μ is injective. Indeed, you proved this on a previous homework. If G is abelian then (1) always holds and there is nothing else to do.

So let's assume that G is non-abelian and assume that $\mu : H \times K \rightarrow G$ is a group isomorphism. As above this implies that (2) and (3) hold. To see that (1) holds, note for all $h \in H$ and $k \in K$ that

$$\begin{aligned} hk &= \mu(h, \varepsilon)\mu(\varepsilon, k) = \mu[(h, \varepsilon)(\varepsilon, k)] \\ &= \mu[(h, k)] \\ &= \mu[(\varepsilon, k)(h, \varepsilon)] \\ &= \mu(\varepsilon, k)\mu(h, \varepsilon) \\ &= kh. \end{aligned}$$

It follows from this that $H \trianglelefteq G$ and $K \trianglelefteq G$.

Conversely, suppose that (1), (2) and (3) hold. From (2) and (3) we know that $\mu : H \times K \rightarrow G$ is a bijection. To see that μ is a homomorphism, let $h \in H$ and $k \in K$ and consider the *commutator* element:

$$(hkh^{-1})k^{-1} = hkh^{-1}k^{-1} = h(khk^{-1}).$$

Since $K \trianglelefteq G$ we have $hkh^{-1} \in K$ and hence $(hkh^{-1})k^{-1} \in K$. But since $H \trianglelefteq G$ we also have $khk^{-1} \in H$ and hence $h(khk^{-1}) \in H$. It follows that $hkh^{-1}k^{-1} \in H \cap K = \{\varepsilon\}$ and hence

$$\begin{aligned} hkh^{-1}k^{-1} &= \varepsilon \\ hk &= kh. \end{aligned}$$

From a result on a previous homework this implies that μ is a homomorphism. \square

5. Euler's Rotation Theorem. Recall the definition of the special orthogonal group:

$$SO(3) = \{A \in \text{Mat}_3(\mathbb{R}) : A^T A = I \text{ and } \det(A) = 1\}.$$

We have seen that every element of this group is an isometry of \mathbb{R}^3 . Now you will show that every element of this group is a **rotation**.

- (a) Recall that there exists a nonzero vector $\mathbf{0} \neq \mathbf{u} \in \mathbb{R}^3$ satisfying $A\mathbf{u} = \lambda\mathbf{u}$ if and only if $\det(A - \lambda I) = 0$. Prove that there exists a unit vector $\mathbf{u} \in \mathbb{R}^3$ satisfying $A\mathbf{u} = \mathbf{u}$.
- (b) For all \mathbf{v} perpendicular to \mathbf{u} , prove that $A\mathbf{v}$ is perpendicular to \mathbf{u} .
- (c) Prove that there exists a matrix $B \in SO(3)$ and a real number $\theta \in \mathbb{R}$ such that

$$B^{-1}AB = \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{array} \right).$$

[Hint: Choose unit vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ so that $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are mutually perpendicular. These are the columns of B .] It follows from this that $\mathbf{x} \mapsto A\mathbf{x}$ is a rotation around the line $\mathbb{R}\mathbf{u} \subseteq \mathbb{R}^3$ by angle θ .

- (a) For all $n \times n$ matrices $A \in \text{Mat}_n$ recall that $\det(-A) = (-1)^n \det(A)$ and $\det(A^T) = \det(A)$. Thus for all $A \in SO(3)$ we have $A - I = A - A^T A = (I - A^T)A$ and hence

$$\begin{aligned} \det(A - I) &= \det([I - A^T]) \det(A) \\ &= \det([I - A]^T) \cdot 1 \\ &= \det(I - A) \\ &= (-1)^3 \det(A - I). \end{aligned}$$

Since $(-1)^3 = -1$ this implies that $\det(A - I) = 0$. Then since $\lambda = 1$ is an eigenvalue we conclude that there exists a unit vector $\mathbf{u} \in \mathbb{R}^3$ such that $A\mathbf{u} = \mathbf{u}$. Our goal is to show that the function $\mathbf{x} \mapsto A\mathbf{x}$ is a rotation around the line $\mathbb{R}\mathbf{u}$.

- (b) So consider any vector $\mathbf{v} \in \mathbb{R}^3$ that is perpendicular to \mathbf{u} , i.e., such that $\mathbf{v}^T \mathbf{u} = 0$. Since $A^T A = I$ this implies that

$$(A\mathbf{v})^T \mathbf{u} = (A\mathbf{v})^T (A\mathbf{u}) = \mathbf{v}^T (A^T A) \mathbf{u} = \mathbf{v}^T \mathbf{u} = 0,$$

hence the vector $A\mathbf{v}$ is also perpendicular to \mathbf{u} . In other words, the function $\mathbf{x} \mapsto A\mathbf{x}$ sends the plane $\mathbb{R}\mathbf{u}^\perp$ to itself. Since this function also preserves distances we know that the function $\mathbf{x} \mapsto A\mathbf{x}$ restricted to the plane $\mathbb{R}\mathbf{u}^\perp$ is a rotation or a reflection. And since $\det(A) = 1$, it must be a rotation.

- (c) To be specific, let \mathbf{v}, \mathbf{w} be perpendicular unit vectors in the plane $\mathbb{R}\mathbf{u}^\perp$ and let B be the matrix with columns $\mathbf{u}, \mathbf{v}, \mathbf{w}$. Since the columns of B are orthonormal we have $B^T B = I$ and by swapping \mathbf{v} and \mathbf{w} if necessary we can assume that $\det(B) = 1$. Then since $A, B \in SO(3)$ we also have $B^{-1}AB \in SO(3)$. Now observe that the first column of $B^{-1}AB$ is equal to \mathbf{e}_1 :

$$(B^{-1}AB)\mathbf{e}_1 = B^{-1}A\mathbf{u} = B^{-1}\mathbf{u} = \mathbf{e}_1.$$

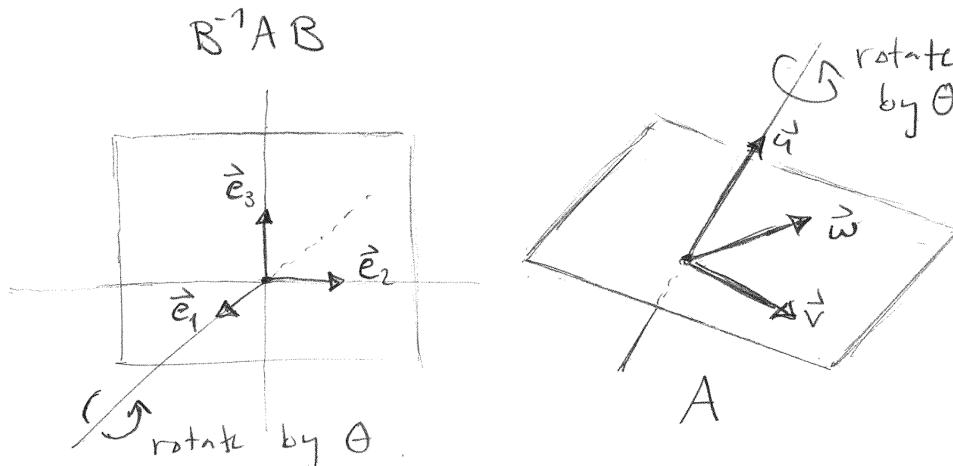
Since the columns of $B^{-1}AB \in SO(3)$ are orthonormal this implies that the second and third columns have zeroes in the first entry. In other words, there exists $A' \in \text{Mat}_2(\mathbb{R})$ such that

$$B^{-1}AB = \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & & A' \\ 0 & & \end{array} \right).$$

But since $B^{-1}AB$ has orthonormal columns we must have $A' \in O(2)$ and since $B^{-1}AB$ has determinant 1 we must have $A' \in SO(2)$. Finally, we conclude from Euler's Isomorphism (proved in class) that

$$A' = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad \text{for some angle } \theta.$$

Here's a picture:



[Remark: As a corollary of this, if R_1 and R_2 are rotations of \mathbb{R}^3 —hence are both elements of $SO(3)$ —then it follows that the composition $R_1R_2 \in SO(3)$ is also a rotation. This fact is **not obvious** to the human visual imagination.]

Week 9

The study of finite groups has always been concerned with “factoring” groups into smaller pieces. This comes directly from Galois Theory, where the goal is to “break down” the symmetries of a given polynomial equation. After discussing the basics of this theory we will finally be in a position to prove that the quintic equation is not solvable.⁷

⁷Assuming, of course, that Galois’ Theorem from Week 2 is true. You will have to wait until next semester if you want to see a proof of that.

Internal Multiplication of Groups. Let $(G, *, \varepsilon)$ be a group and let $H, K \subseteq G$ be any subgroups. Then we have a “multiplication function” from the Cartesian product set $H \times K$ into G :

$$\begin{aligned} \mu : H \times K &\rightarrow G \\ (h, k) &\mapsto h * k. \end{aligned}$$

You have already investigated the properties of this function on the homework. Let me recall the important points.

- The set-theoretic image of μ is called the *product set*:

$$HK := \text{im } \mu = \{h * k : h \in H, k \in K\}.$$

This set may or may not be a subgroup. Here is the most general thing we can say:

- The subset $HK \subseteq G$ is a subgroup if and only if $HK = H \vee K$

Proof. One direction is trivial. For the other direction, suppose that $HK \subseteq G$ is a subgroup. Note that HK contains the union $H \cup K$. But $H \vee K$ is by definition the smallest subgroup that contains the union $H \cup K$, hence we have $H \vee K \subseteq HK$. Conversely, consider any element $h * k \in HK$. Since h and k are in $H \cup K$, they are also in $H \vee K$. Then since $H \vee K$ is a subgroup we have $h * k \in H \vee K$ and hence $HK \subseteq H \vee K$. \square

Thus the function μ is surjective if and only if $HK = H \vee K = G$. When is it injective?

- The function μ is injective if and only if $H \cap K = H \wedge K = \{\varepsilon\}$.

Proof. If μ is injective then for all $g \in H \cap K$ we have $g * g^{-1} = \varepsilon * \varepsilon$, and hence $g = \varepsilon$. Conversely, let $H \cap K = \{\varepsilon\}$ and suppose that $h_1 * k_1 = h_2 * k_2$. Then we have $h_2^{-1} * h_1 = k_2 * k_1^{-1} \in H \cap K$, and it follows that $h_2^{-1} * h_1 = k_2 * k_1^{-1} = \varepsilon$, hence $h_1 = h_2$ and $k_1 = k_2$. \square

In summary, we have the following:

- The function μ is bijective if and only if $H \cap K = \{\varepsilon\}$ and $H \vee K = G$. In this case we say that the subgroups H and K are *complementary*.⁸

So let us assume that $H, K \subseteq G$ are complementary subgroups. Then every element $g \in G$ has a **unique factorization** of the form $g = h * k$ with $h \in H$ and $k \in K$. In other words, for all $h_1, h_2 \in H$ and $k_1, k_2 \in K$ there exist unique elements $h_3 \in H$ and $k_3 \in K$ such that

$$(h_1 * k_1) * (h_2 * k_2) = h_3 * k_3.$$

In general there is not much we can say about the elements h_3, k_3 . However, there are three particularly nice cases:

⁸Remark: A given subgroup can have many different complements. There is no uniqueness implied. For example, any two non-equal lines through the origin in $(\mathbb{R}^2, +, \mathbf{0})$ are complementary.

- If $H \trianglelefteq G$ is a normal subgroup then we have

$$h_3 = h_1 * (k_1 * h_2 * k_1^{-1}) \quad \text{and} \quad k_3 = k_1 * k_2.$$

Proof. For any $h_1, h_2 \in H$ and $k_1, k_2 \in K$ we have

$$(h_1 * k_1) * (h_2 * k_2) = (h_1 * (k_1 * h_2 * k_1^{-1})) * (k_1 * k_2)$$

and if $H \trianglelefteq G$ then we know that $(k_1 * h_2 * k_1^{-1}) \in H$. □

In this case we use the notation

$$G = H \rtimes K,$$

and we say that G is an *internal semidirect product* of H and K . Mnemonic: The triangle points to the normal subgroup.

- If $K \trianglelefteq G$ is a normal subgroup then we have

$$h_3 = h_1 * h_2 \quad \text{and} \quad k_3 = (h_2^{-1} * k_1 * h_2) * k_2.$$

Proof. For any $h_1, h_2 \in H$ and $k_1, k_2 \in K$ we have

$$(h_1 * k_1) * (h_2 * k_2) = (h_1 * h_2) * ((h_2^{-1} * k_1 * h_2) * k_2)$$

and if $K \trianglelefteq G$ then we know that $(h_2^{-1} * k_1 * h_2) \in K$. □

In this case G is still called an *internal semidirect product*, but now the triangle points to K :

$$G = H \ltimes K.$$

- If $H \trianglelefteq G$ and $K \trianglelefteq G$ are both normal then we have

$$h_3 = h_1 * h_2 \quad \text{and} \quad k_3 = k_1 * k_2.$$

Proof. For any $h \in H$ and $k \in K$ we have $k * h * k^{-1} \in H$ and $h * k * h^{-1} \in K$, so that

$$h * k * h^{-1} * k^{-1} \in H \cap K.$$

Since $H \cap K = \{\varepsilon\}$, this implies that $h * k * h^{-1} * k^{-1} = \varepsilon$ and hence $h * k = k * h$. Then for all $h_1, h_2 \in H$ and $k_1, k_2 \in K$ we have

$$(h_1 * k_1) * (h_2 * k_2) = (h_1 * h_2) * (k_1 * k_2).$$

□

In this case the subgroups H and K don't even see each other and we can really think of G as two independent groups sitting side by side. Technically: We say that G is an *internal direct product* of H and K , and we write

$$G = H \times K.^9$$

In any of these three cases it can be said that we have “factored” G into two subgroups, at least one of which is normal.

Next time I'll give some examples and explain the word “internal.”

Today I'll give some examples of internal direct and semidirect products. Then I'll define the concept of “external” products.

Example: Direct Sum of Abelian Groups. Let G be an abelian group and let $H, K \subseteq G$ be subgroups. Since **every** subgroup of an abelian group is normal we have

$$G = H \times K \iff \left\{ \begin{array}{l} HK = G \\ H \cap K = \{\varepsilon\} \end{array} \right\}.$$

If the group $(G, +, 0)$ is additive, we prefer to use the notation of *internal direct sum*:¹⁰

$$G = H \oplus K \iff \left\{ \begin{array}{l} H + K = G \\ H \cap K = \{0\} \end{array} \right\}.$$

This says that every element $g \in G$ has a unique decomposition of the form $g = h + k = k + h$, where $h \in H$ and $k \in K$.

Example: Fundamental Theorem of Finite Abelian Groups. It turns out that

every finite abelian group is a direct sum of cyclic subgroups.

Unfortunately, this theorem is difficult to prove. There are two big ideas needed for the proof:

- Chinese Remainder Theorem
- Smith Normal Form

⁹Clearly the notation $G = H \bowtie K$ would be better, but this would conflict with standard usage.

¹⁰This suggests that maybe “ \otimes ” would be a good notation for direct product of multiplicative groups. Sadly, that notation is used for a different purpose. Sometimes history saddles us with bad notation, such as the negatively charged electron.

You will investigate the Chinese Remainder Theorem on the next homework. The Smith Normal Form is beyond the scope of this course, but it might show up next semester. This presents a pedagogical dilemma: State the Fundamental Theorem now or wait until we can prove it? I choose to state it now.

Example: Basis of a Vector Space. Recall that a vector space consists of a field \mathbb{F} acting (by “scaling”) on an additive group $(V, +, \mathbf{0})$. We say that a subgroup $U \subseteq V$ is a subspace if it is closed under this scaling. Then the notion of direct sum applies without modification to subspaces. As an application, I claim that a set of nonzero vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n \in V - \{\mathbf{0}\}$ is a basis if and only if

$$V = \mathbb{F}\mathbf{u}_1 \oplus \mathbb{F}\mathbf{u}_2 \oplus \dots \oplus \mathbb{F}\mathbf{u}_n.$$

Proof. We define the direct sum of multiple groups by induction. The join condition is easy:

$$V = \mathbb{F}\mathbf{u}_1 + \mathbb{F}\mathbf{u}_2 + \dots + \mathbb{F}\mathbf{u}_n.$$

This literally says that $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n \in V$ is a spanning set. The meet condition is trickier. By induction it says that

$$\mathbb{F}\mathbf{u}_i \cap (\mathbb{F}\mathbf{u}_1 + \dots + \mathbb{F}\mathbf{u}_{i-1} + \mathbb{F}\mathbf{u}_{i+1} + \dots + \mathbb{F}\mathbf{u}_n) = \{\mathbf{0}\} \quad \text{for all } i.$$

Now suppose that $b_1\mathbf{u}_1 + \dots + b_n\mathbf{u}_n = \mathbf{0}$ for some coefficients $b_i \in \mathbb{F}$. Then for each i we have

$$b_i\mathbf{u}_i = -b_1\mathbf{u}_1 - \dots - b_{i-1}\mathbf{u}_{i-1} - b_{i+1}\mathbf{u}_{i+1} - \dots - b_n\mathbf{u}_n,$$

which by the above condition implies that $b_i\mathbf{u}_i = \mathbf{0}$ and hence $b_i = 0$. In other words, the set $\mathbf{u}_1, \dots, \mathbf{u}_n \in V$ is linearly independent. \square

Now we need a non-abelian example.

Example: Dihedral Groups. Consider the dihedral group of size $2n$:

$$D_{2n} = \langle R, F \rangle = \{I, R, \dots, R^{n-1}, F, RF, \dots, R^{n-1}F\}.$$

On a previous homework you showed that every element has the form $R^a F^b$ for some $a, b \in \mathbb{Z}$, which implies that

$$D_{2n} = \langle R \rangle \langle F \rangle = \langle F \rangle \langle R \rangle = \langle R \rangle \vee \langle F \rangle.$$

The easiest way to show that $\langle R \rangle \cap \langle F \rangle = \{I\}$ is to think of the representation where R is a rotation matrix and F is a reflection matrix, so that $\det(R) = 1$ and $\det(F) = -1$. Since $\det(R^a) = \det(R)^a = 1$ for all $a \in \mathbb{Z}$ this implies that F is not a power of R . And since $F^2 = I$ we conclude that no non-trivial power of F is equal to a power of R .

Thus we conclude that the multiplication map is a bijection:

$$\begin{aligned} \langle R \rangle \times \langle F \rangle &\longleftrightarrow D_{2n} \\ (R^a, F^b) &\mapsto R^a F^b. \end{aligned}$$

What kind of product is this? If $n = 2$ then it's a direct product. However if $n \geq 3$ then it's **not a direct product** because the elements of $\langle R \rangle$ and $\langle F \rangle$ don't commute:

$$FRF^{-1} = FRF = R^{-1} \neq R.$$

This implies indirectly that the subgroups $\langle R \rangle \subseteq D_{2n}$ and $\langle F \rangle \subseteq D_{2n}$ are not both normal. I claim that $\langle R \rangle$ is normal and $\langle F \rangle$ is not.

Proof. Assume that $n \geq 3$. Then since $R^2 \neq I$ we have

$$RFR^{-1} = RRF = R^2F \notin \langle F \rangle,$$

and hence $\langle F \rangle \subseteq D_{2n}$ is not a normal subgroup. To see that $\langle R \rangle \subseteq D_{2n}$ is normal we need to show that $gR^a g^{-1} \in \langle R \rangle$ for all $a \in \mathbb{Z}$. This is obvious when g is a power of R , so let's assume that $g = R^b F$ for some $b \in \mathbb{Z}$. Then we have

$$gR^a g^{-1} = (R^b F)R^a(R^b F)^{-1} = R^b F R^a F R^{-b} = R^b F F R^{-a} R^{-b} = R^{-a} \in \langle R \rangle.$$

□

It follows that the dihedral group is a semidirect product:

$$D_{2n} = \langle R \rangle \rtimes \langle F \rangle.$$

More precisely, the rule for multiplying elements is

$$(R^a F^b)(R^c F^d) = [R^a F^b (R^c) F^{-b}] [F^b F^d] = \begin{cases} (R^a R^{-c})(F^b F^d) & b \text{ odd,} \\ (R^a R^c)(F^b F^d) & b \text{ even.} \end{cases}$$

The whole structure is determined by the fact F acts on $\langle R \rangle$ by inversion:

$$FR^a F^{-1} = R^{-a}.$$

The geometric meaning behind this is that flipping the polygon reverses the senses of clockwise and counterclockwise. ///

More generally, we can define an abstract (“external”) product group whenever one group acts on another.

External Multiplication of Groups. Let $(H, *, \delta)$ and $(K, \bullet, \varepsilon)$ be abstract groups. Previously we assumed that H and K are subgroups of some “ambient” group G . Now there is no G , but we still want to construct a group that could be called the “product” of H and K . Specifically, we want to define a group operation on the Cartesian product set $H \times K$. Let's call this hypothetical operation \square , so that for all $h_1, h_2 \in H$ and $k_1, k_2 \in K$ we have

$$(h_1, k_1) \square (h_2, k_2) = (h_3, k_3)$$

for some unique elements h_3 and k_3 . We also want to require that the subsets

$$\begin{aligned}\tilde{H} &= \{(h, \varepsilon) : h \in H\} \subseteq H \times K \\ \tilde{K} &= \{(\delta, k) : k \in K\} \subseteq H \times K\end{aligned}$$

are subgroups isomorphic to H and K , respectively. It turns out that it is hopeless to solve this problem in general. However, there are two constructions that are particularly nice.

- **External Direct Product** The direct product structure is defined by

$$(h_1, k_1) \square (h_2, k_2) = (h_1 * h_2, k_1 \bullet k_2).$$

It is easy to check that $G := (H \times K, \square, (\delta, \varepsilon))$ is an abstract group. Furthermore, I claim that G is an internal direct product of the subgroups \tilde{H} and \tilde{K} .

Proof. Clearly we have $\tilde{H} \cap \tilde{K} = \{(\delta, \varepsilon)\}$ and $\tilde{H}\tilde{K} = G$. The fact that $\tilde{H} \trianglelefteq G$ and $\tilde{K} \trianglelefteq G$ are both normal follows from the fact that their elements commute:

$$(h, \varepsilon) \square (\delta, k) = (h, k) = (\delta, k) \square (h, \varepsilon) \quad \text{for all } h \in H \text{ and } k \in K.$$

□

The definition of the external direct product is so obvious¹¹ that we just use the Cartesian product notation:

$$H \times K = (H \times K, \square, (\delta, \varepsilon)).$$

- **External Semidirect Product.** Suppose that the abstract group $(K, \bullet, \varepsilon)$ acts on the abstract group $(H, *, \delta)$ by automorphisms. In other words, suppose that we have a group homomorphism

$$\theta : K \rightarrow \text{Aut}(H).$$

Then we can define the operation

$$(h_1, k_1) \square_\theta (h_2, k_2) = (h_1 * \theta_{k_1}(h_2), k_1 \bullet k_2).$$

It is relatively easy to check that $G := (H \times K, \square_\theta, (\delta, \varepsilon))$ is an abstract group and I will leave this as an optional exercise for the reader. The reason we call it semidirect is because this G is an internal semidirect product of its subgroups \tilde{H} and \tilde{K} .

Proof. I'll skip some details. The main point is that \tilde{H} is closed under conjugation by elements of \tilde{K} . To see this, note that for all $(h, \varepsilon) \in \tilde{H}$ and $(\delta, k) \in \tilde{K}$ we have

$$\begin{aligned}(\delta, k) \square_\theta (h, \varepsilon) \square_\theta (\delta, k)^{-1} &= (\delta, k) \square_\theta (h, \varepsilon) \square_\theta (\delta, k^{-1}) \\ &= (\delta, k) \square_\theta (h * \theta_\varepsilon(\delta), k^{-1}) \\ &= (\delta, k) \square_\theta (h * \delta, k^{-1})\end{aligned}$$

¹¹Another reason for this notation is the fact that the direct product is the “categorical product” in the category of groups. Convention says that categorical products are always denoted by \times .

$$\begin{aligned}
&= (\delta, k) \square_{\theta} (h, k^{-1}) \\
&= (\delta * \theta_k(h), k \bullet k^{-1}) \\
&= (\theta_k(h), \varepsilon) \in \tilde{H}.
\end{aligned}$$

□

In summary, given any action $\theta : K \rightarrow \text{Aut}(H)$ of one abstract group on another we have defined an abstract group G out of thin air, which contains isomorphic copies $\tilde{H}, \tilde{K} \subseteq G$, in which $\tilde{H} \trianglelefteq G$ is a normal subgroup, and in which the action of \tilde{K} on \tilde{H} by conjugation coincides with the abstract action of K on H . We call this G the *external semidirect product with respect to θ* and we use the notation

$$H \rtimes_{\theta} K = (H \times K, \square_{\theta}, (\delta, \varepsilon)).$$

///

Remarks:

- The external semidirect product is sometimes called a “twisted product,” and the homomorphism θ is sometimes called a “twist.” The use of the Greek character θ is traditional in this context.
- If H acts on K via $\theta : H \rightarrow \text{Aut}(K)$ then we can define a group $H \rtimes_{\theta} K$ via the operation

$$(h_1, k_1)_{\theta} \square (h_2, k_2) = (h_1 * h_2, \theta_{h_2}^{-1}(k_1) \bullet k_2).$$

All of the properties work out the same except that now \tilde{K} is the normal subgroup.

- Let $\text{triv} : K \rightarrow \text{Aut}(H)$ be the “trivial action” that sends each $k \in K$ to the identity function $\text{triv}_k = \text{id} : H \rightarrow H$. Then the semidirect product coincides with the direct product:

$$H \rtimes_{\text{triv}} K = H \times K.$$

- If H and K are abelian groups then the external direct product $H \times K$ is also abelian. However, a semidirect product $H \rtimes_{\theta} K$ need not be abelian. For example, the non-abelian dihedral group is a semidirect product of two abelian (cyclic) groups.
- Many authors of undergraduate algebra textbooks choose to omit the semidirect product on the grounds that it is too abstract. I agree that it’s abstract, but I prefer to keep it in because it is important for geometry and physics. We will see an interesting example next time.

Today we will discuss a very interesting example of a semidirect product. But first, here's a more basic example.

Example: Dihedral Groups Again. Consider the cyclic groups $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/n\mathbb{Z}$. You will show on the next homework that every automorphism of the group $\mathbb{Z}/n\mathbb{Z}$ has the form $k \mapsto ak \pmod n$ for some $a \in \mathbb{Z}$ satisfying $\gcd(a, n) = 1$. Now suppose we have a group homomorphism

$$\begin{aligned} \theta : \mathbb{Z}/2\mathbb{Z} &\rightarrow \text{Aut}(\mathbb{Z}/n\mathbb{Z}) \\ 0 &\mapsto \theta_0 \\ 1 &\mapsto \theta_1 \end{aligned}$$

By definition this consists of two automorphisms $\theta_0, \theta_1 \in \text{Aut}(\mathbb{Z}/n\mathbb{Z})$ with the property that

$$\theta_{a+b \pmod 2} = \theta_{a \pmod 2} \circ \theta_{b \pmod 2}.$$

This implies that $\theta_0 = \text{id}$ and $\theta_1^2 = \theta_2 = \theta_0 = \text{id}$. On the other hand we know that $\theta_1(k) = ak$ for some $a \in \mathbb{Z}$ and since $\theta_1^2 = \text{id}$ we must have $a^2 = 1$. Thus there are only two possible ways that $\mathbb{Z}/2\mathbb{Z}$ can act on $\mathbb{Z}/n\mathbb{Z}$ by automorphisms:

- The trivial action sends each element of $\mathbb{Z}/2\mathbb{Z}$ to the identity function $\mathbb{Z}/n\mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z}$.
- The nontrivial action $\theta : \mathbb{Z}/2\mathbb{Z} \rightarrow \text{Aut}(\mathbb{Z}/n\mathbb{Z})$ sends $0 \in \mathbb{Z}/2\mathbb{Z}$ to the identity function and sends $1 \in \mathbb{Z}/2\mathbb{Z}$ to the “inversion function” $k \mapsto -k$. In this case one can show that the semidirect product is isomorphic to the dihedral group:

$$(\mathbb{Z}/n\mathbb{Z}) \rtimes_{\theta} (\mathbb{Z}/2\mathbb{Z}) \cong D_{2n}.$$

[Exercise: Show that the function $(a, b) \mapsto R^a F^b$ is the desired group isomorphism.]

Now for the interesting example.

Example: Isometries of Euclidean Space. Recall that n -dimensional Euclidean space consists of the vector space \mathbb{R}^n together with the standard dot product $\langle -, - \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$. By an *isometry* of Euclidean space we mean any function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ that preserves the distance between points:

$$\|f(\mathbf{x}) - f(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\| \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Clearly the identity is an isometry and the composition of any two isometries is an isometry. It is less obvious, but it will follow from the analysis below that any isometry is invertible.

Thus we obtain a group

$$\text{Isom}(\mathbb{R}^n) = \{f : \mathbb{R}^n \rightarrow \mathbb{R}^n : f \text{ preserves distance}\}.$$

This group has two interesting subgroups:

- Let $\text{Isom}_{\mathbf{0}}(\mathbb{R}^n) \subseteq \text{Isom}(\mathbb{R}^n)$ denote the subset of isometries that fix the origin: $f(\mathbf{0}) = \mathbf{0}$. We saw on a previous homework that this subgroup is isomorphic to the group $O(n)$ of $n \times n$ orthogonal matrices. To be specific, for each $f \in \text{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ there exists a unique matrix $A \in O(n)$ such that

$$f(\mathbf{x}) = A\mathbf{x} \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

The hardest part of that proof was to show that any isometry that fixes the origin must be a linear function.

- Recall that each group acts on itself by translation. In the case of the additive group $(\mathbb{R}^n, +, \mathbf{0})$ we have a group homomorphism

$$\tau : \mathbb{R}^n \rightarrow \text{Aut}(\mathbb{R}^n)$$

which sends each vector $\mathbf{u} \in \mathbb{R}^n$ to the *translation function* $\tau_{\mathbf{u}}(\mathbf{x}) = \mathbf{x} + \mathbf{u} = \mathbf{u} + \mathbf{x}$. I claim that this translation is an isometry.

Proof. For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we have

$$\|\tau_{\mathbf{u}}(\mathbf{x}) - \tau_{\mathbf{u}}(\mathbf{y})\| = \|(\mathbf{x} + \mathbf{u}) - (\mathbf{y} + \mathbf{u})\| = \|\mathbf{x} - \mathbf{y}\|.$$

□

Thus we have a group homomorphism from $(\mathbb{R}^n, +, \mathbf{0})$ into the group of isometries:

$$\tau : \mathbb{R}^n \rightarrow \text{Isom}(\mathbb{R}^n).$$

Furthermore, I claim that this homomorphism is injective.

Proof. We will show that the kernel is trivial. So consider any vector $\mathbf{u} \in \mathbb{R}^n$ such that $\tau_{\mathbf{u}}$ is the identity function. Then in particular we must have $\mathbf{u} = \mathbf{0} + \mathbf{u} = \tau_{\mathbf{u}}(\mathbf{0}) = \mathbf{0}$. □

In conclusion, we find that the image of τ is a subgroup of $\text{Isom}(\mathbb{R}^n)$ which is isomorphic to the additive group $(\mathbb{R}^n, +, \mathbf{0})$. We will call this the *translation subgroup* and we will label it by $T(\mathbb{R}^n)$:

$$\mathbb{R}^n \cong \text{im } \tau =: T(\mathbb{R}^n) \subseteq \text{Isom}(\mathbb{R}^n).$$

Then we have the following theorem.

Theorem (Isometries of Euclidean Space). The group of isometries is a semidirect product of translations with the origin-fixing isometries:

$$\text{Isom}(\mathbb{R}^n) = T(\mathbb{R}^n) \rtimes \text{Isom}_{\mathbf{0}}(\mathbb{R}^n).$$

Proof. There are three things to check: (1) $T(\mathbb{R}^n)$ and $\text{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ meet at the identity, (2) $T(\mathbb{R}^n)$ and $\text{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ join to the full group, and (3) $T(\mathbb{R}^n)$ is a normal subgroup of $\text{Isom}(\mathbb{R}^n)$.

(1) To show that $T(\mathbb{R}^n) \cap \text{Isom}_{\mathbf{0}}(\mathbb{R}^n) = \{\text{id}\}$, suppose that $\tau_{\mathbf{u}}$ is a translation that fixes the origin. We saw above that this implies $\mathbf{u} = \mathbf{0}$ and hence $\tau_{\mathbf{u}} = \tau_{\mathbf{0}} = \text{id}$.

(2) To show that $T(\mathbb{R}^n) \circ \text{Isom}_{\mathbf{0}}(\mathbb{R}^n) = \text{Isom}(\mathbb{R}^n)$ consider any isometry $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and suppose that $f(\mathbf{0}) = \mathbf{u}$. Now define $g := \tau_{-\mathbf{u}} \circ f$ and observe that

$$g(\mathbf{0}) = \tau_{-\mathbf{u}}(f(\mathbf{0})) = \tau_{-\mathbf{u}}(\mathbf{u}) = \mathbf{u} - \mathbf{u} = \mathbf{0}.$$

It follows that $g \in \text{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ and hence

$$f = \tau_{\mathbf{u}} \circ g \in T(\mathbb{R}^n) \circ \text{Isom}_{\mathbf{0}}(\mathbb{R}^n).$$

(3) To show that $T(\mathbb{R}^n) \trianglelefteq \text{Isom}(\mathbb{R}^n)$ it is enough¹² to show that $T(\mathbb{R}^n)$ is closed under conjugation by elements of $\text{Isom}_{\mathbf{0}}(\mathbb{R}^n)$. So consider any $f \in \text{Isom}_{\mathbf{0}}(\mathbb{R}^n)$. The important fact (which was tricky to prove) is that f is a linear function. Therefore for any $\mathbf{x} \in \mathbb{R}^n$ we have

$$(f \circ \tau_{\mathbf{u}})(\mathbf{x}) = f(\mathbf{x} + \mathbf{u}) = f(\mathbf{x}) + f(\mathbf{u}) = \tau_{f(\mathbf{u})}(f(\mathbf{x})) = (\tau_{f(\mathbf{u})} \circ f)(\mathbf{x}).$$

It follows that $f \circ \tau_{\mathbf{u}} = \tau_{f(\mathbf{u})} \circ f$ and hence

$$f \circ \tau_{\mathbf{u}} \circ f^{-1} = \tau_{f(\mathbf{u})} \in T(\mathbb{R}^n).$$

□

In summary, every element of $\text{Isom}(\mathbb{R}^n)$ has a unique factorization of the form $\tau_{\mathbf{u}} \circ f$ where $\tau_{\mathbf{u}} \in T(\mathbb{R}^n)$ is a translation and $f \in \text{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ is an orthogonal linear function. In this language the group operation is given by

$$(\tau_{\mathbf{u}} \circ f) \circ (\tau_{\mathbf{v}} \circ g) = (\tau_{\mathbf{u}} \circ f \circ \tau_{\mathbf{v}} \circ f^{-1}) \circ (f \circ g) = (\tau_{\mathbf{u}} \circ \tau_{f(\mathbf{v})}) \circ (f \circ g).$$

There is also an external point of view. Let $\theta : O(n) \rightarrow \text{Aut}(\mathbb{R}^n)$ be the natural action of the group of orthogonal matrices on the vector space $(\mathbb{R}^n, +, \mathbf{0})$. This is defined by matrix multiplication:

$$\theta_A(\mathbf{x}) := A\mathbf{x} \quad \text{for all } A \in O(n) \text{ and } \mathbf{x} \in \mathbb{R}^n.$$

We can then form the external semidirect product

$$\mathbb{R}^n \rtimes_{\theta} O(n).$$

By the above theorem this semidirect product is isomorphic to the group of isometries of Euclidean space. ///

Finally, here's a less interesting version of the same construction.

Example: The General Affine Group. Let V be a vector space and let G be the group of all invertible functions $V \rightarrow V$. (These do not need to preserve any structure.) Inside this

¹²This follows from part (2).

group there is a subgroup $T(V) \subseteq G$ of translations and a subgroup $GL(V) \subseteq G$ of linear functions. By the same reasoning as above one can show that

$$f \circ \tau_{\mathbf{u}} = \tau_{f(\mathbf{u})} \circ f \quad \text{for all } \tau_{\mathbf{u}} \in T(V) \text{ and } f \in GL(V).$$

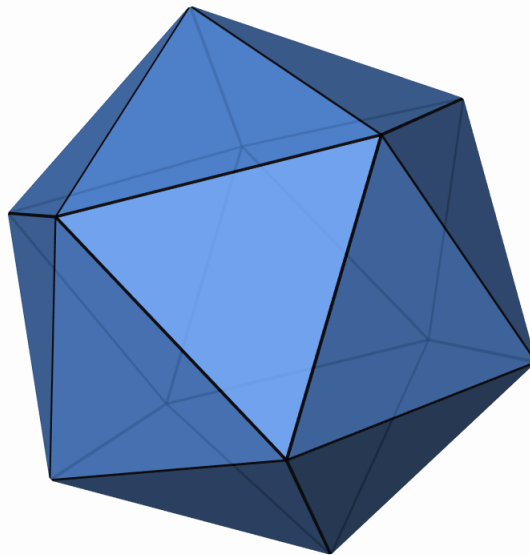
It follows that the product set $T(V) \circ GL(V) \subseteq G$ is a subgroup, which contains $T(V)$ as a normal subgroup. We call this the *general affine group* of V :

$$GA(V) := T(V) \circ GL(V) = T(V) \rtimes GL(V).$$

This construction is less interesting because it's not so clear why we should care about this kind of function (i.e., compositions of linear functions and translations).¹³

Week 10

For me this is the hardest part of the course, when I need to start thinking about tying up loose ends. Back in Week 2 I mentioned that the unsolvability of the quintic equation has something to do with the group of symmetries of the regular icosahedron. Let's return to that topic now. Just so we're all on the same page, here's a picture:¹⁴



Assume that the regular icosahedron is centered at the origin in \mathbb{R}^3 and let $I \subseteq SO(3)$ be the subgroup of rotations that leave the icosahedron invariant. We will prove below that this group has 60 elements and it satisfies the following special property:

$$\{\text{id}\} \subsetneq H \leq I \implies H = I.$$

¹³Given a vector space V , there is a technical way to “forget” which point is the origin. After doing this we call V an “affine vector space.” The group $GA(V)$ is simply the group of automorphisms of this structure.

¹⁴I made this picture using *KaleidoTile*, by Jeff Weeks.

This property has a name.

Definition of Simple Groups. We say that a group G is *simple* if it has no nontrivial normal subgroups. This implies that G cannot be decomposed as a direct or semidirect product of smaller groups. ///

Example: Simple Abelian Groups. Since every subgroup of an abelian group is normal, we see that an abelian group is simple if and only if it has **no non-trivial subgroups**. I claim that the only such groups are $\mathbb{Z}/p\mathbb{Z}$.

Theorem. Every simple abelian group has the form $\mathbb{Z}/p\mathbb{Z}$ for some prime $p \in \mathbb{Z}$.

Proof. Let $(G, *, \varepsilon)$ be a simple abelian group and consider any element $g \in G$. If $g \neq \varepsilon$ then we have $\{\varepsilon\} \subsetneq \langle g \rangle \subseteq G$, which implies that $G = \langle g \rangle$ is cyclic. If G were infinite then we would have $G \cong (\mathbb{Z}, +, 0)$, which is not simple. Therefore we must have $G \cong \mathbb{Z}/n\mathbb{Z}$ for some $n \geq 1$. But recall from the Fundamental Theorem of Cyclic Groups that the lattice of subgroups $\mathcal{L}(\mathbb{Z}/n\mathbb{Z})$ is isomorphic to the lattice of divisors $\text{Div}(n)$. It follows that $\mathbb{Z}/n\mathbb{Z}$ has no proper subgroup if and only if n has no proper divisor, i.e., if and only if n is prime. \square

[Remark: You will show on the homework that $\mathbb{Z}/p\mathbb{Z}$ is actually a **field**.]

In this sense we can think of simple groups as a generalization of prime numbers. It is much more difficult to find non-abelian simple groups. If you only know about small groups then you might suspect that there is no such thing. In fact, it turns out that the icosahedral group I of size 60 is the **smallest possible non-abelian simple group**.

Building on the analogy with prime numbers, it turns out that every group¹⁵ has a “unique decomposition” into simple factors.

The Jordan-Hölder Theorem and “Solvable” Groups. Let $(G, *, \varepsilon)$ be a group and consider a finite chain of subgroups:

$$G = G_0 \supsetneq G_1 \supsetneq G_2 \supsetneq \cdots \supsetneq G_\ell = \{\varepsilon\}.$$

We call this chain a *composition series* if it satisfies the following two conditions:

- $G_{i+1} \trianglelefteq G_i$ is normal for each i ,
- the quotient group G_i/G_{i+1} is simple for each i .

We can summarize these conditions by saying that $G_{i+1} \trianglelefteq G_i$ is a **maximal normal** subgroup for each i . To prove equivalence, one should check that the correspondence between subgroups of G_i/G_{i+1} and subgroups between G_i and G_{i+1} preserves normality. Then the quotient group

¹⁵Not literally every group, but it’s true for all finite groups and many infinite groups.

G_i/G_{i+1} has no non-trivial normal subgroup (i.e., is simple) if and only if there is no normal subgroup strictly between G_i and G_{i+1} (i.e., if G_{i+1} is maximal normal in G_i).

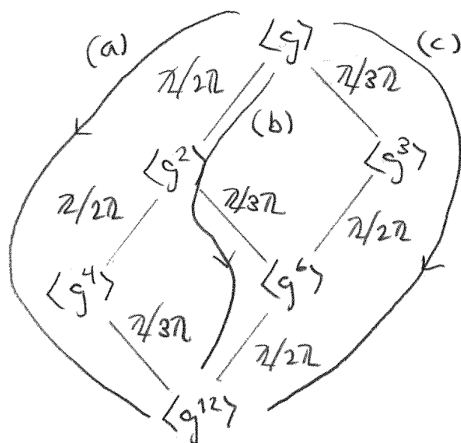
Under these conditions, the Jordan-Hölder Theorem says that the list of simple groups

$$G_1/G_0, \quad G_2/G_1, \quad \dots \quad G_{n-1}/G_{\ell}, \quad G_{\ell}/\{\varepsilon\} = G_{\ell}$$

is **unique** up to isomorphism and permutations. ///

Unfortunately the proof of this theorem is beyond the scope of the course, however you will prove a similar theorem for vector spaces on the homework. The unique simple groups G_{i+1}/G_i arising from a composition series are called the *composition factors* of the group G . If G is a **cyclic group** and if a prime p divides $\#G$ with multiplicity k , then the simple group $\mathbb{Z}/p\mathbb{Z}$ is a composition factor of G with multiplicity k . In this sense the Jordan-Hölder Theorem is a vast generalization of the Fundamental Theorem of Arithmetic.

Example: Composition Factors of $\mathbb{Z}/12\mathbb{Z}$. Let $\langle g \rangle \cong \mathbb{Z}/12\mathbb{Z}$ be a cyclic group of size 12. Since we know the subgroup lattice, it is easy to see that this group has exactly three different composition series, labeled (a), (b), (c) in the following picture. I have also labeled each edge in the diagram with the corresponding quotient group. Observe that the sequence of composition factors is the same for all three composition series:



Composition Factors :

- (a) $\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}, \mathbb{Z}/3\mathbb{Z}$
- (b) $\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}/3\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}$
- (c) $\mathbb{Z}/3\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}$

The big difference between integers and groups is that multiplying integers is easy, while “multiplying groups” can be arbitrarily complicated. Indeed, suppose that p is prime and let $f(p)$ be the number of different (non-isomorphic) groups having the composition factors

$$\mathbb{Z}/p\mathbb{Z}, \quad \mathbb{Z}/p\mathbb{Z} \quad \dots \quad \mathbb{Z}/p\mathbb{Z} \quad (k \text{ times}).$$

Graham Higman proved in 1960 that the number of such groups is really big:

$$f(p) \geq p^{2k^2(k-6)27}.$$

On the other hand, there is only one integer with the prime factors p, p, \dots, p (k times). Even though it is impossible to classify these so-called p -groups, we still say that these groups are “solvable” in the following technical sense.

Definition of Solvable Groups. Since every simple abelian group is isomorphic to $\mathbb{Z}/p\mathbb{Z}$ for some prime $p \in \mathbb{Z}$, we have the following equivalence:

$$\left\{ \begin{array}{l} G \text{ has abelian} \\ \text{composition factors} \end{array} \right\} \iff \left\{ \begin{array}{l} G \text{ has composition factors of the form} \\ \mathbb{Z}/p\mathbb{Z} \text{ for various prime numbers } p \end{array} \right\}.$$

Any group satisfying these conditions is called a *solvable group*.

And what is so “solvable” about these groups? To explain this, here is another restatement of Galois’ Theorem.

Galois’ Theorem Again. The general n -th degree polynomial equation is solvable by radicals if and only if the symmetric group S_n has abelian composition factors, i.e., if and only if the symmetric group S_n is a “solvable group.” ///

At this point I might as well go ahead and prove that S_n is **not** solvable for all $n \geq 5$. If you believe Galois’ Theorem then this fact implies that the general n -th degree equation is not solvable by radicals when $n \geq 5$. We will prove Galois’ Theorem next semester.

Theorem. The symmetric group S_n is not solvable when $n \geq 5$.

Proof. Let $n \geq 5$ and assume for contradiction that there exists a chain of subgroups

$$S_n = G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_r = \{\text{id}\}$$

such that each quotient group G_i/G_{i+1} exists and is abelian. Now let $C \subseteq S_n = G_0$ be the set of all 3-cycles. We will prove by induction that $C \subseteq \{\text{id}\} = G_r$, which is a contradiction.

So fix $0 \leq i < r$ and assume for induction that $C \subseteq G_i$. If $c_1, c_2 \in C$ are any two 3-cycles, then since G_i/G_{i+1} is abelian we have

$$\begin{aligned} (c_1 c_2 c_1^{-1} c_2^{-1} G_{i+1}) &= (c_1 G_{i+1})(c_2 G_{i+1})(c_1 G_{i+1})^{-1}(c_2 G_{i+1})^{-1} \\ &= (c_1 G_{i+1})(c_1 G_{i+1})^{-1}(c_2 G_{i+1})(c_2 G_{i+1})^{-1} \\ &= (\text{id } G_{i+1})(\text{id } G_{i+1}) \\ &= \text{id } G_{i+1} \\ &= G_{i+1}, \end{aligned}$$

which implies that $c_1 c_2 c_1^{-1} c_2^{-1} \in G_{i+1}$. Thus in order to show that $C \subseteq G_{i+1}$ it is enough to show that every 3-cycle $c \in C$ has the form $c = c_1 c_2 c_1^{-1} c_2^{-1}$ for some 3-cycles $c_1, c_2 \in C$. For

this we will use the fact that $n \geq 5$. To be specific, let $c = (ijk)$. Then for any numbers $\ell \neq m$ not in the set $\{i, j, k\}$ we have

$$(ijk) = (jkm)(ilj)(jkm)^{-1}(ilj)^{-1}.$$

[Exercise: Check this.]

□

Remarks:

- It is remarkable that the proof of the unsolvability of polynomial equations looks like this. Clearly this is the most efficient way to think about the problem.
- With more work, one can show that the alternating subgroup $A_n \subseteq S_n$ is actually **simple** when $n \geq 5$. (The proof is a bit hairy so we won't do it. In general it is difficult to prove that a non-abelian group is simple.) It follows from this that the composition factors of S_n are A_n and $S_n/A_n \cong \mathbb{Z}/2\mathbb{Z}$. The fact that A_n is non-abelian is the ultimate reason why S_n is not solvable.
- Thus there exists an infinite sequence A_5, A_6, A_7, \dots of non-abelian finite simple groups. We will see later that the group A_5 is isomorphic to the icosahedral group, which is why the icosahedron is related to the quintic equation.
- You will show on a future homework that the group A_4 has a normal subgroup of size 4. This is the ultimate reason why the quartic equation is solvable. You already know that A_3 and A_2 are solvable. In fact they are abelian.

I claimed last time that the icosahedral group $I \subseteq SO(3)$ has 60 elements and no non-trivial normal subgroups, but I didn't prove either of these statements. In order to count the elements we will use the method of "orbits" and "stabilizers." It turns out that the same method (applied to conjugacy classes) will also help us study the normal subgroups.

Definition of Orbits and Stabilizers. Let $(G, *, \varepsilon)$ be a group, let X be a set with structure and let $\varphi : G \rightarrow \text{Aut}(X)$ be a group homomorphism (i.e., an action of G on X). Then for any point $x \in X$ we define the following sets:

$$\begin{aligned} \text{Orb}_\varphi(x) &:= \{\varphi_g(x) : g \in G\} \subseteq X, \\ \text{Stab}_\varphi(x) &:= \{g \in G : \varphi_g(x) = x\} \subseteq G. \end{aligned}$$

When the specific action φ is understood we will just write $\text{Orb}(x)$ and $\text{Stab}(x)$. We can also view orbits as the equivalence classes of the following relation:

$$x \sim_\varphi y \iff \exists g \in G, \varphi_g(x) = y.$$

Let's verify that this relation is an equivalence.

Proof.

(E1) For all $x \in X$ we have $\varphi_\varepsilon(x) = x$ and hence $x \sim_\varphi x$.

(E2) Let $x, y \in X$ and assume that $x \sim_\varphi y$ so that $\varphi_g(x) = y$ for some $g \in G$. But then we have $\varphi_{g^{-1}}(y) = \varphi_g^{-1}(y) = x$, which implies that $y \sim_\varphi x$ because $g^{-1} \in G$.

(E3) Let $x, y, z \in X$ and assume that $x \sim_\varphi y$ and $y \sim_\varphi z$. This means that $\varphi_g(x) = y$ and $\varphi_h(y) = z$ for some $g, h \in G$. But then we have

$$\varphi_{h*g}(x) = \varphi_h(\varphi_g(x)) = \varphi_h(y) = z,$$

which implies that $x \sim_\varphi z$ because $h * g \in G$.

□

It follows that X is a disjoint union of the orbits:

$$X = \coprod_i \text{Orb}(x_i) \quad \text{for some arbitrary class representatives } x_i \in X.$$

///

If the set X has some nice structure (e.g., if it's a topological space or a manifold) then the orbits might also have this structure but it depends on the properties of the action φ . There is not much we can say in general. As for the stabilizer, it is always a subgroup of G .

Proof. For all $x \in X$ and $a, b \in \text{Stab}(x)$ we have $\varphi_a(x) = x$ and $\varphi_b(x) = x$, hence $\varphi_b^{-1}(x) = x$. But then since φ is a group homomorphism we have

$$\varphi_{a*b^{-1}}(x) = (\varphi_a \circ \varphi_b^{-1})(x) = \varphi_a(\varphi_b^{-1}(x)) = \varphi_a(x) = x,$$

and it follows that $a * b^{-1} \in \text{Stab}(x)$.

□

Unfortunately the subgroup $\text{Stab}(x) \subseteq G$ is generally **not normal**, but we still have a nice structure theorem for group actions, which is analogous to the First Isomorphism Theorem for group homomorphisms.

The Orbit-Stabilizer Theorem. Let $\varphi : G \rightarrow \text{Aut}(X)$ be a group action. Then for all $x \in X$ we have a bijection between points of the orbit and left cosets of the stabilizer:

$$\begin{aligned} \Phi : \text{Orb}(x) &\longrightarrow G / \text{Stab}(x) \\ \varphi_g(x) &\longmapsto g \text{Stab}(x). \end{aligned}$$

Proof. The function Φ is well-defined and injective because

$$\varphi_a(x) = \varphi_b(x) \iff \varphi_b^{-1}(\varphi_a(x)) = x$$

$$\begin{aligned} \iff x &= \varphi_{b^{-1}*a}(x) \\ \iff b^{-1} * a &\in \text{Stab}(x) \\ \iff a \text{Stab}(x) &= b \text{Stab}(x), \end{aligned}$$

and it is surjective by definition. □

It follows that X can be identified with a disjoint union of sets of cosets:

$$X = \coprod_i \text{Orb}(x_i) \iff \coprod_i G/\text{Stab}(x_i).$$

We will see below that this formula is often useful for counting. ///

[Remark: We could also define a bijection between points of the orbit and **right cosets** of the stabilizer. The reason I use left cosets is because the map $\Phi : \text{Orb}(x) \rightarrow G/\text{Stab}(x)$ “commutes” with the natural action of G on both sides. In other words, the bijection Φ is actually an *isomorphism of G -sets*. But we won’t use this extra structure.]

For example, let’s count the symmetries of an icosahedron.

Example: Counting the Symmetries of a Regular Icosahedron. Let $I \subseteq SO(3)$ be the group of rotational symmetries of a regular icosahedron centered at the origin in \mathbb{R}^3 . The Greek prefix *icos-* indicates that the icosahedron has 20 triangular faces. Consider the set

$$F = \{\text{faces of the icosahedron}\}.$$

The group I acts on the set F in the obvious way, and we say that this action is *transitive* since for any face $f \in F$ we have $\text{Orb}(f) = F$. (Indeed, the adjective “regular” in “regular icosahedron” indicates that every face/edge/vertex of the polyhedron looks the same up to symmetry.) Furthermore, the only symmetries that stabilize the triangle f are the three rotational symmetries through the center of the triangle. We conclude from the Orbit-Stabilizer Theorem and Lagrange’s Theorem that

$$\begin{aligned} \text{Orb}(f) = F &\leftrightarrow I/\text{Stab}(f) \\ \#F &= \#I/\#\text{Stab}(f) \\ 20 &= \#I/3 \\ \#I &= 60. \end{aligned}$$

Similarly, we have transitive actions of I on the set of edges E and the set of vertices V of the icosahedron. It is easy to see that for each edge $e \in E$ the stabilizer $\text{Stab}(e)$ is a (cyclic) group of size 2, and for each vertex v the stabilizer $\text{Stab}(v)$ is a cyclic group of size 5. (Look at the picture above.) Thus we obtain two more equations

$$\begin{aligned} \#\text{Orb}(e) &= \#I/\#\text{Stab}(e), & \#\text{Orb}(v) &= \#I/\#\text{Stab}(v), \\ \#E &= \#I/2, & \#V &= \#I/5. \end{aligned}$$

It follows from this that the number of edges of the icosahedron is $\#E = 60/2 = 30$ and the number of vertices is $\#V = 60/5 = 12$. I find this method much easier than counting the edges and vertices by hand. ///

Today's lecture will be a bit philosophical.

What is “group theory”? In retrospect, one could say that Carl Friedrich Gauss was doing group theory when he invented and studied the group $\mathbb{Z}/n\mathbb{Z}$ around 1800. However, as you know by now, the study of finite abelian groups is only a tiny (but important) part of the subject. The main definitions of the theory were only revealed with Galois' work on the (non-abelian) symmetric group S_n and his discovery of “normal subgroups.” From 1830 until 1870 the subject of group theory basically consisted of the study of S_n and finite abelian groups. The subject still had nothing to do with geometry.

Meanwhile, the discovery of non-Euclidean geometry inspired Sophus Lie and Felix Klein to study “geometric transformations.” Slowly they realized that the collection of all transformations of a geometry X forms an abstract group, which today we call the automorphism group $\text{Aut}(X)$. After reading Camille Jordan's 1870 book on permutations, they decided it would be worthwhile to develop some “Galois theory” of geometric transformations. This inspired Klein's famous *Erlangen Program* of 1872.

Definition of Transitive Actions. Let G be a group and let X be a set with structure. We say that an action $\varphi : G \rightarrow \text{Aut}(X)$ is *transitive* if it has only one orbit. In other words:

$$\text{Orb}_\varphi(x) = X \quad \text{for each point } x \in X.$$

Then from the Orbit-Stabilizer Theorem we obtain a bijection

$$X \quad \longleftrightarrow \quad G/\text{Stab}_\varphi(x) \quad \text{for each point } x \in X.$$

///

Klein's Erlangen Program concerns the case when X is a non-Euclidean geometry and G is the corresponding group of transformations, i.e., functions $f : X \rightarrow X$ that preserve the geometric structure. An essential feature is that any two points of a geometry should “look the same,” which means that the action $G \curvearrowright X$ should be transitive. Klein suggested that one could classify and study different geometries X by looking at the coset spaces G/H for various $H \subseteq G$. Today we use the word *homogeneous space* instead of the old-fashioned *non-Euclidean geometry*. From this point on, the study of geometry was slowly integrated into group theory. By the 1920s even physicists had reluctantly switched to the new group theoretic language.

Before giving an example of a “non-Euclidean geometry,” let me recall what we know about Euclidean geometry.

Example: Euclidean Space. Let $X = (\mathbb{R}^n, \langle -, - \rangle)$ be n -dimensional Euclidean space and let $\text{Isom}(\mathbb{R}^n)$ be the group of isometries, i.e., functions $f : X \rightarrow X$ that preserve distance. Clearly the action of $\text{Isom}(\mathbb{R}^n)$ on X is transitive. (Indeed, for any points $\mathbf{x}, \mathbf{y} \in X$ the translation $\tau_{\mathbf{y}-\mathbf{x}} \in \text{Isom}(\mathbb{R}^n)$ sends \mathbf{x} to \mathbf{y} .) Thus for any point $\mathbf{x} \in X$ we obtain a bijection

$$X \longleftrightarrow \text{Isom}(\mathbb{R}^n)/\text{Isom}_{\mathbf{x}}(\mathbb{R}^n),$$

where $\text{Isom}_{\mathbf{x}}(\mathbb{R}^n) := \{f \in \text{Isom}(\mathbb{R}^n) : f(\mathbf{x}) = \mathbf{x}\}$ is the stabilizer of \mathbf{x} . In particular, we already know that $\text{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ is isomorphic to the orthogonal group. Hence we obtain a bijection:

$$\mathbb{R}^n \longleftrightarrow \text{Isom}(\mathbb{R}^n)/O(n).$$

But this is not so interesting because it follows from the group isomorphism

$$\text{Isom}(\mathbb{R}^n) \cong (\mathbb{R}^n, +, \mathbf{0}) \rtimes O(n)$$

which we proved above. ///

Here’s something new.

Example: Real Projective Space. The basic idea of projective geometry is that any two lines in a plane should meet at a unique point. Lines which were called “parallel” in Euclidean geometry now intersect at some ideal “point at infinity,” and the collection of all points at infinity forms the “line at infinity” for this plane. In the modern “analytic” treatment, we define the set

$$\mathbb{P}^{n-1}(\mathbb{R}) = \{\text{lines through the origin in } \mathbb{R}^n\}.$$

In other words, a “point” in $(n-1)$ -dimensional projective space corresponds to a “line through the origin” in n -dimensional Euclidean space. In order to get some concrete representation of this set, we observe that the orthogonal group $O(n)$ acts transitively on $\mathbb{P}^{n-1}(\mathbb{R})$. (Indeed, given two lines $\ell, \ell' \subseteq \mathbb{R}^n$ intersecting at $\mathbf{0}$, we can send ℓ to ℓ' by rotating the plane that they generate, and this rotation can be realized as an orthogonal matrix.) Furthermore, I claim that for any line $\ell \in \mathbb{P}^{n-1}(\mathbb{R})$ the stabilizer is isomorphic to a direct product $\text{Stab}(\ell) \cong O(1) \times O(n-1)$, hence we obtain a bijection

$$\mathbb{P}^{n-1}(\mathbb{R}) \longleftrightarrow \frac{O(n)}{O(1) \times O(n-1)}.$$

Actually I will prove something more general than this. For any vector subspace $U \subseteq \mathbb{R}^n$ let $U^\perp \subseteq \mathbb{R}^n$ be the *orthogonal subspace* defined by

$$U^\perp := \{\mathbf{v} \in \mathbb{R}^n : \langle \mathbf{u}, \mathbf{v} \rangle = 0 \text{ for all } \mathbf{u} \in U\}.$$

Then for any orthogonal matrix $A \in O(n)$ I claim that

$$A \text{ stabilizes } U \iff A \text{ stabilizes } U^\perp.$$

Proof. Suppose that $A \in O(n)$ stabilizes U . Then for all $\mathbf{u} \in U$ and $\mathbf{v} \in U^\perp$ we have $A\mathbf{u} \in U$ and $A^{-1} = A^T$, hence

$$\langle \mathbf{u}, A^{-1}\mathbf{v} \rangle = \langle \mathbf{u}, A^T\mathbf{v} \rangle = \mathbf{u}^T(A^T\mathbf{v}) = (A\mathbf{u})^T\mathbf{v} = \langle A\mathbf{u}, \mathbf{v} \rangle = 0.$$

It follows that $A^{-1}\mathbf{v} \in U^\perp$ for all $\mathbf{v} \in U^\perp$ and hence $A^{-1}U^\perp \subseteq U^\perp$ is a vector subspace. But since $A^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is injective, it follows from the Rank-Nullity Theorem (proved on the next homework) that $A^{-1}U^\perp$ and U^\perp have the same dimension, hence $A^{-1}U^\perp = U^\perp$. Finally, since every element $\mathbf{v} \in U^\perp$ has the form $\mathbf{v} = A^{-1}\mathbf{v}'$ for some $\mathbf{v}' \in U^\perp$, we conclude that $A\mathbf{v} = \mathbf{v}' \in U^\perp$ as desired. The other direction is similar. \square

[Remark: Note that the proof used finite-dimensionality. Indeed, this result is not true for infinite dimensional vector spaces.]

If $U \subseteq \mathbb{R}^n$ is a k -dimensional subspace then we can choose an orthonormal basis $\mathbf{u}_1, \dots, \mathbf{u}_n \in V$ such that $\mathbf{u}_1, \dots, \mathbf{u}_k$ is a basis for U and $\mathbf{u}_{k+1}, \dots, \mathbf{u}_n$ is a basis for U^\perp . If $B \in O(n)$ is the matrix whose i -th column is \mathbf{u}_i then for any $A \in \text{Stab}(U) \subseteq O(n)$ it follows from the result just proved that

$$B^{-1}AB = \left(\begin{array}{c|c} A' & 0 \\ \hline 0 & A'' \end{array} \right) \text{ for some } A' \in O(k) \text{ and } A'' \in O(n-k).$$

Finally, the map $A \mapsto (A', A'')$ defines a group isomorphism $\text{Stab}(U) \cong O(k) \times O(n-k)$ and we obtain a bijection

$$\left\{ \begin{array}{l} k\text{-dimensional} \\ \text{subspaces of } \mathbb{R}^n \end{array} \right\} \longleftrightarrow \frac{O(n)}{O(k) \times O(n-k)}.$$

The case $k = 1$ corresponds to projective space. ///

Remarks:

- It would take us too far afield to discuss what this has to do with “points at infinity.” One hundred years ago it was common for every undergraduate math major to take a course in synthetic projective geometry. Sadly, the analytic version the subject is so technical that it is usually only studied by graduate students.
- The set $O(n)/[O(k) \times O(n-k)]$ is called a *Grassmann manifold* or a *Grassman variety*. It is important for the study of vector bundles in physics.

- I understand that this example was challenging. It will not be on the exam. An easier version of the same argument shows that:

$$\left\{ \begin{array}{l} \text{subsets of size } k \text{ from} \\ \text{the set } \{1, \dots, n\} \end{array} \right\} \longleftrightarrow \frac{S_n}{S_k \times S_{n-k}}.$$

Note that this is related to the binomial coefficients.

To end the lecture I will discuss some easier (but still philosophical) examples.

Definition of Free and Regular Actions. Let G be a group and let X be a set with structure. We say that an action $\varphi : G \rightarrow \text{Aut}(X)$ is *free* if each stabilizer is trivial:

$$\text{Stab}_\varphi(x) = \{\varepsilon\} \quad \text{for each point } x \in X.$$

In this case, every orbit is in bijection with G :

$$\text{Orb}_\varphi(x) \longleftrightarrow G/\{\varepsilon\} = G.$$

If there is only one orbit (i.e., if the action is also transitive) then we say that the action is *regular*.¹⁶ In this case, each point $x \in X$ defines a bijection between X and G :

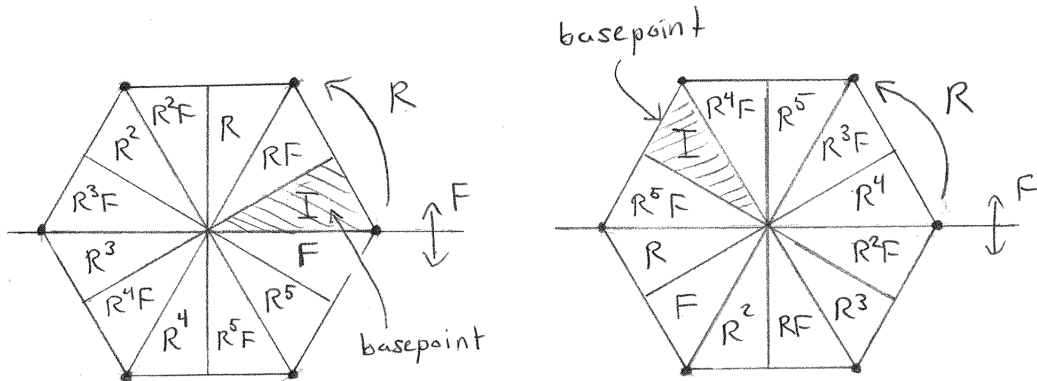
$$\begin{array}{l} X \longleftrightarrow G \\ \varphi_g(x) \longleftrightarrow g. \end{array}$$

///

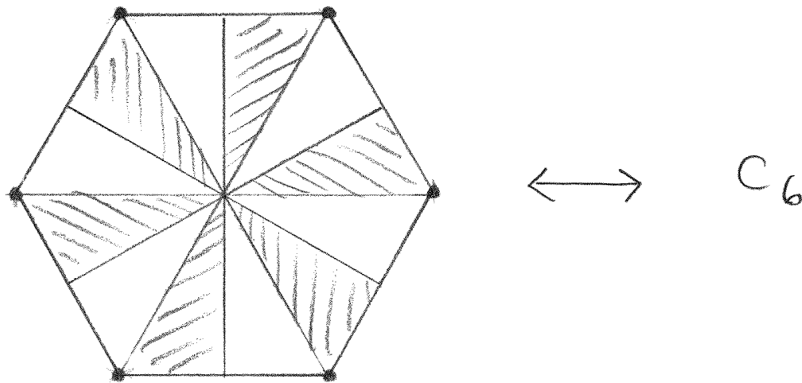
Example: Dihedral and Cyclic Groups. The dihedral group D_{2n} acts by symmetries on a regular n -sided polygon. This induces transitive actions of D_{2n} on the n -vertices and the n -edges of the polygon. But is there some set of objects on which the group acts freely?

Answer: Divide the polygon into n isocoles triangles from the center and then divide each of these into 2 right triangles. Let X be the resulting set of $2n$ triangles. Then $D_{2n} \curvearrowright X$ is a regular action, hence for each arbitrary choice of “basepoint” $x \in X$ we obtain a bijection $D_{2n} \leftrightarrow X$. Here are two choices of basepoint when $n = 6$:

¹⁶I don’t like this word too much. I’ll probably just say *free and transitive*. The term *simply-transitive* is also common. The fanciest way to describe a regular action $G \curvearrowright X$ is to say that X is a G -torsor.



Note that the two bijections are quite different. We can obtain a geometric model for the cyclic group $C_{2n} = \langle R \rangle \subseteq D_{2n}$ by shading half of the triangles. For example, the group C_6 acts freely and transitively on the six shaded triangles in the following picture:



///

If $G \curvearrowright X$ is a regular action, then after choosing a basepoint $x \in X$ we can think of X as a group isomorphic to G , with identity element x . However, no basepoint is better than any other. Thus, in some sense, we can think of X as a version of G where we have forgotten which element is the identity. Sometimes this is useful.

Definition of Affine Space. When René Descartes invented Cartesian coordinates (in 1637) his intention was to model the real world. However, there is one big problem: The Cartesian space \mathbb{R}^3 has an origin but the real world does not. Can we fix this problem?

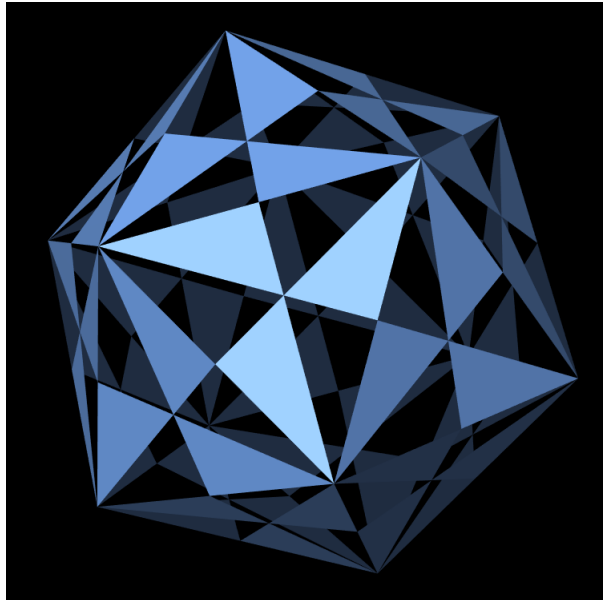
Answer: Let X be a set and let $\varphi : (\mathbb{R}^n, +, \mathbf{0}) \rightarrow \text{Perm}(X)$ be a regular action. Then for any arbitrary basepoint $x \in X$ we obtain a bijection $X \leftrightarrow \mathbb{R}^n$ identifying $x \in X$ with $\mathbf{0} \in \mathbb{R}^n$. The

pair (X, φ) is called *affine n -dimensional space*. This is a better model for the real world. ///

And here is one last example.

Example: Rotations and Reflections of an Icosahedron. Let $I \subseteq SO(3)$ be the group of rotational symmetries of a regular icosahedron centered at the origin in \mathbb{R}^3 . Can we find some set of 60 things on which this group acts regularly?

Answer: Consider the “barycentric subdivision” of the icosahedron. This is defined by dividing each edge at the midpoint and dividing each triangular face into six right triangles. Then color all of the triangles with two alternating colors, as in the following picture:



The group I acts freely and transitively on the set of blue triangles. If we choose an arbitrary triangle to play the role of the identity element then we obtain a bijection

$$\{\text{blue triangles}\} \longleftrightarrow I.$$

And what about the other 60 “empty” triangles?¹⁷ Let $\hat{I} \subseteq O(3)$ be the group of rotation **and** reflection symmetries of the icosahedron, which contains the rotations $I \subseteq \hat{I}$ as a subgroup. Then the group \hat{I} acts freely and transitively on the set of all 120 triangles, and it follows that

$$\#\hat{I} = 120.$$

[Remark: Note that $I \subseteq \hat{I}$ is analogous to the cyclic subgroup $C_n \subseteq D_{2n}$ of the dihedral group. More generally, for any shape $X \subseteq \mathbb{R}^n$ in Euclidean space we have a group of symmetries $\text{Sym}(X) \subseteq O(n)$ and an “alternating subgroup” $\text{Alt}(X) = \text{Sym}(X) \cap SO(n)$.]

¹⁷ *KaleidoTile* won’t let me use two colors.

Problem Set 5

1. Quotient Rings. Let $(R, +, \times, 0, 1)$ be a *commutative ring*. Technically: This means that (1) $(R, +, 0)$ is an abelian group, (2) $(R, \times, 1)$ is a commutative monoid (abelian group without inverses), and (3) for all $a, b, c \in R$ we have $a(b + c) = ab + ac$.

(a) Let $I \subseteq R$ be an additive subgroup and recall that “addition of cosets” is well-defined:

$$(a + I) + (b + I) = (a + b) + I.$$

Thus we obtain the quotient group $(R/I, +, 0 + I)$. Now suppose that for all $a \in R$ and $b \in I$ we have $ab \in I$. (Jargon: We say that $I \subseteq R$ is an *ideal*.) In this case prove that the following “multiplication of cosets” is well-defined:

$$(a + I)(b + I) = (ab) + I.$$

It follows that $(R/I, +, \times, 0 + I, 1 + I)$ is a ring, called the *quotient ring*. [You do not need to check all the details.]

(b) Apply part (a) to show that $\mathbb{Z}/n\mathbb{Z}$ is a ring.

(a) Let $I \subseteq R$ be an ideal and assume that we have $a + I = a' + I$ and $b + I = b' + I$. By definition this means that $a - a' \in I$ and $b - b' \in I$, and since I is an ideal this implies that $a(b - b') \in I$ and $(a - a')b \in I$. It follows that

$$ab - a'b' = ab - ab' + ab' - a'b' = a(b - b') + (a - a')b \in I,$$

and hence $ab + I = a'b' + I$.

(b) We will show that $n\mathbb{Z} \subseteq \mathbb{Z}$ is an ideal. Indeed, for any $a \in \mathbb{Z}$ and $b = nk \in n\mathbb{Z}$ we have

$$ab = a(nk) = n(ak) \in n\mathbb{Z}.$$

[Remark: We will say much more about this next semester.]

2. The Fermat-Euler-Lagrange Theorem, Part II. Let $(R, +, \times, 0, 1)$ be a ring and let $R^\times \subseteq R$ denote the subset of elements that have multiplicative inverses. We call $(R^\times, \times, 1)$ the *group of units*.

(a) For all $n \in \mathbb{Z}$ prove that $(\mathbb{Z}/n\mathbb{Z})^\times = \{a + n\mathbb{Z} : \gcd(a, n) = 1\}$. [Hint: If $\gcd(a, n) = 1$ then we have $a\mathbb{Z} + n\mathbb{Z} = 1\mathbb{Z}$, hence there exist integers $x, y \in \mathbb{Z}$ with $ax + ny = 1$. This is sometimes called *Bézout’s Identity*.]

(b) **Euler’s Totient Theorem.** Euler’s totient function is defined by $\phi(n) := \#(\mathbb{Z}/n\mathbb{Z})^\times$. For all $a \in \mathbb{Z}$ with $\gcd(a, n) = 1$ prove that

$$a^{\phi(n)} = 1 \pmod{n}.$$

(c) **Fermat's Little Theorem.** If $p \in \mathbb{Z}$ is prime and $p \nmid a$ prove that

$$a^{p-1} = 1 \pmod{p}.$$

(a) Recall that $a\mathbb{Z} + n\mathbb{Z} = \gcd(a, n)\mathbb{Z}$, and note that the following conditions are equivalent:

$$\begin{aligned} (a + n\mathbb{Z} \text{ is invertible}) &\iff \exists x \in \mathbb{Z}, (a + n\mathbb{Z})(x + n\mathbb{Z}) = 1 + n\mathbb{Z} \\ &\iff \exists x \in \mathbb{Z}, ax + n\mathbb{Z} = 1 + n\mathbb{Z} \\ &\iff \exists x \in \mathbb{Z}, 1 - ax \in n\mathbb{Z} \\ &\iff \exists x \in \mathbb{Z}, 1 \in ax + n\mathbb{Z} \\ &\iff 1 \in a\mathbb{Z} + n\mathbb{Z} \\ &\iff 1\mathbb{Z} = a\mathbb{Z} + n\mathbb{Z} \\ &\iff 1 = \gcd(a, n). \end{aligned}$$

(b) **Euler's Totient Theorem.** You proved on a previous homework that if $(G, *, \varepsilon)$ is a finite abelian group then $g^{\#G} = \varepsilon$ for all $g \in G$, and I proved in class that the same result holds for non-abelian groups. Now consider the multiplicative group $G = (\mathbb{Z}/n\mathbb{Z})^\times$ and let $\phi(n) := \#(\mathbb{Z}/n\mathbb{Z})^\times$. From part (a) we know that

$$(\mathbb{Z}/n\mathbb{Z})^\times = \{a + n\mathbb{Z} : \gcd(a, n) = 1\}.$$

So consider any $a \in \mathbb{Z}$ with $\gcd(a, n) = 1$. Then our theorem says that

$$a^{\phi(n)} + n\mathbb{Z} = (a + n\mathbb{Z})^{\phi(n)} = 1 + n\mathbb{Z}.$$

In other words, we have $a^{\phi(n)} = 1 \pmod{n}$.

(c) **Fermat's Little Theorem.** For example, let p be prime. Then since $\gcd(k, p) = 1$ for all $1 \leq k \leq p-1$ we conclude that $\phi(p) = p-1$. It follows from (b) that for all $a \in \mathbb{Z}$ with $\gcd(a, p) = 1$ we have $a^{p-1} = 1 \pmod{p}$. It only remains to observe that

$$\gcd(a, p) = 1 \iff p \nmid a.$$

3. Chinese Remainder Theorem. In this problem I will use the shorthand notation $[a]_n := a + n\mathbb{Z}$. Now fix some $m, n \in \mathbb{Z}$ with $\gcd(m, n) = 1$ and consider the function

$$\begin{aligned} \varphi: \mathbb{Z}/mn\mathbb{Z} &\rightarrow \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \\ [a]_{mn} &\mapsto ([a]_m, [a]_n). \end{aligned}$$

(a) Prove that φ is **well-defined**. That is, for all $a, a' \in \mathbb{Z}$ prove that

$$[a]_{mn} = [a']_{mn} \quad \text{implies} \quad [a]_m = [a']_m \quad \text{and} \quad [a]_n = [a']_n.$$

- (b) For all $c \in \mathbb{Z}$ prove that $m|c$ and $n|c$ together imply $(mn)|c$. [Hint: There exist $x, y \in \mathbb{Z}$ such that $mx + ny = 1$.] Use this conclude that φ is **injective**.
- (c) Prove that φ is **surjective**. [Big Hint: Given $([a]_m, [b]_n)$ we want to find $c \in \mathbb{Z}$ such that $[a]_m = [c]_m$ and $[b]_n = [c]_n$. Try $c := any + bmx$.]
- (d) Prove that φ restricts to a bijection

$$\varphi : (\mathbb{Z}/mn\mathbb{Z})^\times \longleftrightarrow (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times.$$

[Hint: Use the fact that $\gcd(k, \ell) = 1$ if and only if there exist integers $x, y \in \mathbb{Z}$ such that $kx + \ell y = 1$.] It follows that Euler's totient is multiplicative: $\phi(mn) = \phi(m)\phi(n)$.

(a) Assume that $[a]_{mn} = [a']_{mn}$, so that $a - a' \in mn\mathbb{Z}$. Since $mn\mathbb{Z} \subseteq m\mathbb{Z} \cap n\mathbb{Z}$ this implies that $a - a' \in m\mathbb{Z}$, hence $[a]_m = [a']_m$, and $a - a' \in n\mathbb{Z}$, hence $[a]_n = [a']_n$.

(b) Let $c \in \mathbb{Z}$ and assume that $m|c$ and $n|c$. Say $c = mm'$ and $c = nn'$ for some $m', n' \in \mathbb{Z}$. Then since $\gcd(m, n) = 1$ we have $mx + ny = 1$ for some $x, y \in \mathbb{Z}$. It follows that

$$\begin{aligned} 1 &= mx + ny, \\ c &= cmx + cny = nn'mx + mm'ny = mn(n'x + m'y), \end{aligned}$$

and hence $(mn)|c$. To see that φ is injective, suppose that we have $([a]_m, [a]_n) = ([b]_m, [b]_n)$ for some $a, b \in \mathbb{Z}$. Since $[a]_m = [b]_m$ we have $m|(a - b)$ and since $[a]_n = [b]_n$ we have $n|(a - b)$. Then since $\gcd(m, n) = 1$, the above result tells us that $mn|(a - b)$ and hence $[a]_{mn} = [b]_{mn}$ as desired.

(c) It is easy to prove that φ is surjective, but it is not so easy to find a formula for the inverse. To see that φ is surjective, note that

- $\text{im } \varphi \subseteq (\mathbb{Z}/m \times \mathbb{Z}/n\mathbb{Z})$,
- $\#(\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}) = \#(\mathbb{Z}/n\mathbb{Z}) \cdot \#(\mathbb{Z}/m\mathbb{Z}) = mn$,
- and $\#\text{im } \varphi = \#(\mathbb{Z}/mn\mathbb{Z}) = mn$ because φ is injective.

It follows that $\text{im } \varphi = \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$.

Now for the tricky part. For any element $([a]_m, [b]_n)$ our goal is to find some $c \in \mathbb{Z}$ such that $\varphi([c]_{mn}) = ([a]_m, [b]_n)$. Since $\gcd(m, n) = 1$ we know that there exist some $x, y \in \mathbb{Z}$ such that $mx + ny = 1$. Now define $c := any + bmx$ and observe that

$$\begin{aligned} [c]_m &= [any + bmx]_m = [any]_m = [a(1 - mx)]_m = [a - amx]_m = [a]_m, \\ [c]_n &= [any + bmx]_n = [bmx]_n = [b(1 - ny)]_n = [b - bny]_n = [b]_n. \end{aligned}$$

It follows that $\varphi([c]_{mn}) = ([c]_m, [c]_n) = ([a]_m, [b]_n)$ and hence

$$\varphi^{-1}([a]_m, [b]_n) = [c]_{mn} = [any + bmx]_{mn}.$$

(d) From Problem 2 we know that $[a]_n$ is invertible if and only if $\gcd(a, n) = 1$. Thus we need to show that

$$\gcd(a, mn) = 1 \iff \gcd(a, m) = 1 \text{ and } \gcd(a, n) = 1.$$

But we also know that $\gcd(a, b) = 1$ if and only if there exist some $x, y \in \mathbb{Z}$ with $ax + by = 1$. So let $\gcd(a, mn) = 1$. Then there exist $x, y \in \mathbb{Z}$ with $ax + mny = 1$. But then we have

$$ax + m(ny) = 1 \implies \gcd(a, m) = 1$$

and

$$ax + n(my) = 1 \implies \gcd(a, n) = 1.$$

Conversely, let $\gcd(a, m) = \gcd(a, n) = 1$, so there exist integers $x, y, x', y' \in \mathbb{Z}$ with $ax + my = 1$ and $ax' + ny' = 1$. But then we have

$$\begin{aligned} (ax + my)(ax' + ny') &= 1 \\ a(axx' + xny' + myx') + mn(yy') &= 1, \end{aligned}$$

and it follows that $\gcd(a, mn) = 1$. We have shown that φ restricts to a bijection

$$\varphi : (\mathbb{Z}/mn\mathbb{Z})^\times \longleftrightarrow (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times.$$

In particular, this tells us that $\phi(mn) = \phi(m)\phi(n)$. □

[Remark: For distinct primes p, q this result says that $\phi(pq) = \phi(p)\phi(q) = (p-1)(q-1)$. Then Euler's Totient Theorem says that for any integers $a, k \in \mathbb{Z}$ with $\gcd(a, pq) = 1$ (i.e., with $p \nmid a$ and $q \nmid a$) we have

$$\begin{aligned} a^{(p-1)(q-1)} &= 1 \pmod{pq}, \\ a^{(p-1)(q-1)k} &= 1 \pmod{pq}, \\ a^{(p-1)(q-1)k+1} &= a \pmod{pq}. \end{aligned}$$

In fact, one can show that the third equation is still true even when $\gcd(a, pq) \neq 1$. This equation is the foundation of the RSA Cryptosystem.]

[Another Remark: The Chinese Remainder Theorem was named by Leonard Dickson in 1929. The name refers to the fact that this result was known in 3rd century China,¹⁸ over 1000 years before it was rediscovered in Europe by Euler (1743) and Gauss (1801). The original application is to solve a system of simultaneous linear congruences. For example, suppose we want to find all $c \in \mathbb{Z}$ such that $c = 2 \pmod{3}$ and $c = 3 \pmod{5}$. That is, we want to solve the following system of two linear congruences:

$$\left\{ \begin{array}{l} [c]_3 = [2]_3 \\ [c]_5 = [3]_5 \end{array} \right\}$$

¹⁸To be specific, it was discovered by Sunzi. Some authors have tried to rename the result as Sunzi's Theorem but it seems to be too late.

In the language of Problem 3(c) we have $(a, b) = (2, 3)$ and $(m, n) = (3, 5)$. Since m and n are coprime we can find some $x, y \in \mathbb{Z}$ such that $mx + ny = 1$. By trial and error I found $(x, y) = (-3, 2)$. (For larger numbers I would use the Euclidean Algorithm.) The theorem/method then tells us that $c = any + bmx = 2 \cdot 5 \cdot 2 + 3 \cdot 3 \cdot (-3) = 20 - 27 = -7$ is the unique solution mod $mn = 15$. In other words:

$$\left\{ \begin{array}{l} [c]_3 = [2]_3 \\ [c]_5 = [3]_5 \end{array} \right\} \iff [c]_{15} = [-7]_{15} = [8]_{15}.$$

This method can be extended to many simultaneous congruences by induction.]

4. Automorphisms of a Cyclic Group. For all integers $n \in \mathbb{Z}$ prove that

$$\text{Aut}(\mathbb{Z}/n\mathbb{Z}) \cong (\mathbb{Z}/n\mathbb{Z})^\times.$$

[Hint: Show that any automorphism $\varphi : \mathbb{Z}/n\mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z}$ has the form $\varphi_a([k]_n) := [ak]_n$ for some integer $a \in \mathbb{Z}$ satisfying $\gcd(a, n) = 1$.]

Proof. Let $\varphi : \mathbb{Z}/n\mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z}$ be any automorphism and suppose that $\varphi([1]_n) = [a]_n$ for some integer $0 \leq a < n$. Since φ is a group homomorphism, I will prove by induction that $\varphi([k]_n) = [ak]_n$ for all $k \in \mathbb{Z}$. Indeed, the statement is true for $k = 0$ and $k = 1$. If it's true for some k then it's also true for $k + 1$ because

$$\varphi([k + 1]_n) = \varphi([k]_n + [1]_n) = \varphi([k]_n) + \varphi([1]_n) = [ak]_n + [a]_n = [a(k + 1)]_n.$$

And if the statement is true for k then it's also true for $-k$ because

$$\varphi([-k]_n) = \varphi(-[k]_n) = -\varphi([k]_n) = -[ak]_n = [-ak]_n = [a(-k)]_n.$$

We have shown that every group homomorphism $\varphi : \mathbb{Z}/n\mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z}$ has the form $\varphi_a([k]_n) := [ak]_n$ for some integer $0 \leq a < n$. It remains to show that φ_a is invertible (i.e., an automorphism) precisely when $\gcd(a, n) = 1$. Indeed, if $\gcd(a, n) = 1$ then there exists $x \in \mathbb{Z}$ with $[ax]_n = [1]_n$ and it follows that $\varphi_a^{-1} = \varphi_x$. Conversely, if $\gcd(a, n) \neq 1$ then we have seen that $[1]_n$ is not in the image of φ_a , hence φ is not surjective.

Now I claim that the function

$$\begin{array}{ccc} \varphi : (\mathbb{Z}/n\mathbb{Z})^\times & \rightarrow & \text{Aut}(\mathbb{Z}/n\mathbb{Z}) \\ [a]_n & \mapsto & \varphi_a \end{array}$$

is a group isomorphism. Indeed, we showed above that this function is surjective. It is well-defined because multiplication mod n is well defined, and it is injective because $\varphi_a = \varphi_b$ implies $[a]_n = \varphi_a([1]_n) = \varphi_b([1]_n) = [b]_n$. Finally, this function is a group homomorphism since for all $a, b, k \in \mathbb{Z}$ we have

$$\varphi_{ab}([k]_n) = [(ab)k]_n = [a(bk)]_n = \varphi_a(\varphi_b([k]_n)) = (\varphi_a \circ \varphi_b)([k]_n).$$

□

5. Matrix Representation of Isometries. Consider the following set of matrices:

$$G = \left\{ \left(\begin{array}{ccc|c} A & & & \mathbf{u} \\ 0 & \cdots & 0 & 1 \end{array} \right) : A \in O(n) \text{ and } \mathbf{u} \in \mathbb{R}^n \right\} \subseteq \text{Mat}_{n+1}(\mathbb{R}).$$

- (a) Prove that $G \subseteq \text{Mat}_{n+1}(\mathbb{R})$ is a subgroup. [Hint: Block multiplication.]
 (b) Use results from class to prove that G is isomorphic to the group $\text{Isom}(\mathbb{R}^n)$ of isometries of n -dimensional Euclidean space.

(a) Clearly the identity is in G . To show that G is closed under multiplication, consider any $A, B \in O(n)$ and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Then using block multiplication gives

$$\left(\begin{array}{ccc|c} A & & & \mathbf{u} \\ 0 & \cdots & 0 & 1 \end{array} \right) \left(\begin{array}{ccc|c} B & & & \mathbf{v} \\ 0 & \cdots & 0 & 1 \end{array} \right) = \left(\begin{array}{ccc|c} AB & & & A\mathbf{v} + \mathbf{u} \\ 0 & \cdots & 0 & 1 \end{array} \right) \in G.$$

We can also use this formula to compute the inverse. Indeed, if $AB = I$ and $A\mathbf{v} + \mathbf{u} = \mathbf{0}$ then we must have $B = A^{-1}$ and $\mathbf{v} = -A^{-1}\mathbf{u}$. It follows that

$$\left(\begin{array}{ccc|c} A & & & \mathbf{u} \\ 0 & \cdots & 0 & 1 \end{array} \right)^{-1} = \left(\begin{array}{ccc|c} A^{-1} & & & -A^{-1}\mathbf{u} \\ 0 & \cdots & 0 & 1 \end{array} \right) \in G.$$

(b) In class we saw that $\text{Isom}(\mathbb{R}^n) = T(\mathbb{R}^n) \times \text{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ is an internal semidirect product, where $T(\mathbb{R}^n) = \{\tau_{\mathbf{u}} : \mathbf{u} \in \mathbb{R}^n\}$ is the subgroup of translations and $\text{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ is the subgroup of isometries that fix $\mathbf{0} \in \mathbb{R}^n$. Furthermore, we know that each element $f \in \text{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ has the form $f(\mathbf{x}) = A\mathbf{x}$ for some unique orthogonal matrix $A \in O(n)$. Let's call this function $f_A(\mathbf{x}) := A\mathbf{x}$. In summary, each element of $\text{Isom}(\mathbb{R}^n)$ has the form $\tau_{\mathbf{u}} \circ f_A$ for some unique vector $\mathbf{u} \in \mathbb{R}^n$ and matrix $A \in O(n)$ and the semidirect product structure is given by

$$(\tau_{\mathbf{u}} \circ f_A) \circ (\tau_{\mathbf{v}} \circ f_B) = (\tau_{\mathbf{u}} \circ \tau_{A\mathbf{v}}) \circ (f_A \circ f_B) = \tau_{\mathbf{u} + A\mathbf{v}} \circ f_{AB}.$$

Now consider the bijection $\varphi : \text{Isom}(\mathbb{R}^n) \rightarrow G$ defined by

$$\tau_{\mathbf{u}} \circ f_A \mapsto \left(\begin{array}{ccc|c} A & & & \mathbf{u} \\ 0 & \cdots & 0 & 1 \end{array} \right).$$

I claim that this is a group isomorphism. Indeed, for all $A, B \in O(n)$ and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ we have

$$\varphi(\tau_{\mathbf{u}} \circ f_A) \varphi(\tau_{\mathbf{v}} \circ f_B) = \left(\begin{array}{ccc|c} A & & & \mathbf{u} \\ 0 & \cdots & 0 & 1 \end{array} \right) \left(\begin{array}{ccc|c} B & & & \mathbf{v} \\ 0 & \cdots & 0 & 1 \end{array} \right)$$

$$\begin{aligned}
&= \left(\begin{array}{ccc|c} AB & & & A\mathbf{v} + \mathbf{u} \\ 0 & \cdots & 0 & 1 \end{array} \right) \\
&= \varphi(\tau_{A\mathbf{v} + \mathbf{u}} \circ f_{AB}) \\
&= \varphi((\tau_{\mathbf{u}} \circ f_A) \circ (\tau_{\mathbf{v}} \circ f_B)).
\end{aligned}$$

□

[Remark: An *affine function* $\mathbb{R}^n \rightarrow \mathbb{R}^n$ has the form $\mathbf{x} \mapsto A\mathbf{x} + \mathbf{u}$ for some matrix $A \in \text{Mat}_n(\mathbb{R})$ and some vector \mathbf{u} . The same trick can be used to represent affine functions as $(n+1) \times (n+1)$ matrices.]

6. Second and Third Isomorphism Theorems.

- (a) Let $H, K \subseteq G$ be subgroups with $K \trianglelefteq G$ normal. We already know that $HK \subseteq G$ is a subgroup. Prove that $K \trianglelefteq HK$ is a normal subgroup and the map $h \mapsto hK$ defines a surjective group homomorphism $H \rightarrow (HK)/K$ with kernel $H \cap K$. It follows that

$$\frac{H}{H \cap K} \cong \frac{HK}{K}.$$

- (b) Now consider another normal subgroup $N \trianglelefteq G$ such that $N \subseteq K$. Prove that $N \trianglelefteq K$ is normal and that the map $gN \mapsto gK$ defines a surjective group homomorphism $G/N \rightarrow G/K$ with kernel K/N . It follows that

$$\frac{G/N}{K/N} \cong \frac{G}{K}.$$

- (a) Let $K \trianglelefteq G$. Since $gkg^{-1} \in K$ for all $k \in K$ and $g \in G$, the same is true for all $k \in K$ and $g \in HK$, hence $K \trianglelefteq HK$. Now consider the function

$$\begin{aligned}
\varphi : H &\rightarrow HK/K \\
h &\mapsto hK.
\end{aligned}$$

This is a group homomorphism by definition of coset multiplication. To see that φ is surjective, consider any element $hk \in HK$ and note that $(hk)K = hK$ because $h^{-1}(hk) = k \in K$. It follows that $(hk)K = \varphi(h)$ for some $h \in H$. Finally, we will show that $\ker \varphi = H \cap K$. Indeed, if $h \in H \cap K$ then we have $\varphi(h) = hK = K$. Conversely, if $\varphi(h) = K$ for some $h \in H$ then we must have $hK = K$ and hence $h \in H \cap K$. Now the First Isomorphism Theorem says

$$\frac{H}{H \cap K} = \frac{H}{\ker \varphi} \cong \text{im } \varphi = \frac{HK}{K}.$$

□

(b) Let $N \subseteq K$ with $N \trianglelefteq G$. For the same reason as above, this implies that $N \trianglelefteq K$, hence the quotient group K/N is defined. Now consider the function

$$\begin{aligned} \varphi : G/N &\rightarrow G/K \\ gN &\mapsto gK. \end{aligned}$$

To see that this is well-defined, observe that

$$g_1N = g_2N \implies g_1^{-1}g_2 \in N \implies g_1^{-1}g_2 \in K \implies g_1K = g_2K.$$

Then φ is a surjective group homomorphism by definition. Finally, we will show that $\ker \varphi = K/N$. Indeed, if $kN \in K/N$ then we have $\varphi(kN) = kK = K$. Conversely, if $\varphi(gN) = gK = K$ then we must have $g \in K$ and hence $gN \in K/N$. Now the First Isomorphism Theorem says

$$\frac{G/N}{K/N} = \frac{G/N}{\ker \varphi} \cong \text{im } \varphi = \frac{G}{K}.$$

□

[Remark: There is not much to do here once you know the definition of the maps. If G is finite then the Third Isomorphism Theorem doesn't tell us anything new about cardinality, but the Second Isomorphism Theorem and Lagrange's Theorem tell us that

$$\#(HK) = \frac{\#H \cdot \#K}{\#(H \cap K)}.$$

It turns out that this formula is still true even when $HK \subseteq G$ is **not** a subgroup. You will prove a generalization of this on the next homework using the Orbit-Stabilizer Theorem.]

7. Dimension of a Vector Space, Part II.

Let V be a vector space over a field \mathbb{F} .

- (a) Let $\mathbf{u}_1, \dots, \mathbf{u}_n \in V$ be a basis and consider the subspaces $V_k := \mathbb{F}(\mathbf{u}_1, \dots, \mathbf{u}_k) \subseteq V$. Prove for all $0 \leq k < n$ that there is no subspace U satisfying

$$V_k \subsetneq U \subsetneq V_{k+1}.$$

- (b) Conversely, suppose that we have a maximal chain of subspaces

$$\{\mathbf{0}\} = V_0 \subsetneq V_1 \subsetneq \dots \subsetneq V_n = V.$$

Prove by induction that V_k has a basis of size k , hence $\dim(V_k) = k$. Parts (a) and (b) together show that **dimension** equals the **length** of a maximal chain of subspaces

- (c) If $U \subseteq V$ is a subspace you may assume that the quotient group V/U is a vector space. Prove that $\dim(V/U) = m$ if and only if there exists a maximal chain of subspaces

$$U = V_0 \subsetneq V_1 \subsetneq \dots \subsetneq V_m = V.$$

[Hint: You may assume that the Correspondence Theorem and the First Isomorphism Theorem still hold after replacing the word "subgroup" with "subspace."¹⁹

¹⁹For that matter, the Second and Third Isomorphism Theorems also hold.

(d) Prove that $\dim(V) = \dim(U) + \dim(V/U)$. [Hint: Combine (a), (b) and (c).]

(e) **Rank-Nullity Theorem.** For any linear function $\varphi : V \rightarrow W$ prove that

$$\dim(V) = \dim(\ker \varphi) + \dim(\operatorname{im} \varphi).$$

(a) Suppose that we have $V_k \subsetneq U \subseteq V_{k+1}$ for some subspace U . Choose some vector $\mathbf{u} \in U - V_k$. Since $\mathbf{u} \in V_{k+1}$ there exist scalars $a_1, \dots, a_{k+1} \in \mathbb{F}$ such that

$$\mathbf{u} = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + \dots + a_{k+1} \mathbf{u}_{k+1}.$$

But since $\mathbf{u} \notin V_k$ we know that $a_{k+1} \neq 0$. It follows that

$$\mathbf{u}_{k+1} = \mathbf{u} - \frac{a_1}{a_{k+1}} \mathbf{u}_1 - \frac{a_2}{a_{k+1}} \mathbf{u}_2 - \dots - \frac{a_k}{a_{k+1}} \mathbf{u}_k \in U,$$

which implies that $V_{k+1} \subseteq U$ and hence $U = V_{k+1}$.

(b) For the base case we will show that $\dim(V_1) = 1$. To do this, choose any vector $\mathbf{u} \in V_1 - \{\mathbf{0}\}$. Then since $\{\mathbf{0}\} \subsetneq \mathbb{F}\mathbf{u} \subseteq V_1$ and since the chain is maximal, we must have $\mathbb{F}\mathbf{u} = V_1$ which implies that \mathbf{u} is a basis for V_1 .

Now assume for induction that $\mathbf{u}_1, \dots, \mathbf{u}_k \in V_k$ is a basis and choose an arbitrary vector $\mathbf{u}_{k+1} \in V_{k+1} - V_k$. I claim that $\mathbf{u}_1, \dots, \mathbf{u}_{k+1}$ is a basis for V_{k+1} . To see this that this is a **spanning set**, observe that

$$V_k \subsetneq \mathbb{F}(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{k+1}) \subseteq V_{k+1}.$$

Then since the chain is maximal we must have $V_{k+1} = \mathbb{F}(\mathbf{u}_1, \dots, \mathbf{u}_{k+1})$. To see that the set is **independent**, consider any scalars $a_1, \dots, a_{k+1} \in \mathbb{F}$ such that

$$a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + \dots + a_{k+1} \mathbf{u}_{k+1} = \mathbf{0}.$$

If $a_{k+1} \neq 0$ then we obtain $\mathbf{u}_{k+1} = -(a_1/a_{k+1})\mathbf{u}_1 - \dots - (a_k/a_{k+1})\mathbf{u}_k \in V_k$, which is a contradiction. Therefore we must have $a_{k+1} = 0$. But then since $\mathbf{u}_1, \dots, \mathbf{u}_k$ is an independent set we must also have $a_1 = a_2 = \dots = a_k = 0$. It follows that $\mathbf{u}_1, \dots, \mathbf{u}_{k+1} \in V_{k+1}$ is a basis and hence that $\dim(V_{k+1}) = k + 1$.

[Remark: Putting (a) and (b) together tells us that the dimension of a (finite-dimensional) vector space V equals the length of any maximal chain in the lattice of subspaces $\mathcal{L}(V)$. In particular, any two maximal chains of subspaces have the same length. This is a specific example of the Jordan-Hölder Theorem for modules over a ring.]

(c) Let $U \subseteq V$ be any subspace. The Correspondence Theorem for subspaces says that the map $W \mapsto W/U$ is a poset isomorphism from the lattice of subspaces $U \subseteq W \subseteq V$ to the lattice of subspaces of V/U :

$$\begin{array}{ccc} \mathcal{L}(V, U) & \xrightarrow{\sim} & \mathcal{L}(V/U) \\ W & \mapsto & W/U. \end{array}$$

It follows that a maximal chain in the poset $\mathcal{L}(V/U)$ has the same length as a maximal chain in the poset $\mathcal{L}(V, U)$.

(d) Suppose that $\dim(U) = k$ and $\dim(V/U) = \ell$. From parts (a), (b), (c) we know that there exist maximal chains of subspaces

$$\{\mathbf{0}\} = U_0 \subsetneq U_1 \subsetneq \cdots \subsetneq U_k = U \quad \text{and} \quad U = V_0 \subsetneq V_1 \subsetneq \cdots \subsetneq V_\ell = V.$$

But then

$$\{\mathbf{0}\} = U_0 \subsetneq U_1 \subsetneq \cdots \subsetneq U_k = V_0 \subsetneq V_1 \subsetneq \cdots \subsetneq V_\ell = V$$

is also a maximal chain of subspaces and it follows from part (b) that

$$\dim(V) = k + \ell = \dim(U) + \dim(V/U).$$

(e) **Rank-Nullity Theorem.** Let $\varphi : V \rightarrow W$ be any linear function. Then the First Isomorphism Theorem gives us an isomorphism of vector spaces

$$V / \ker \varphi \cong \text{im } \varphi$$

and it follows from part (d) that

$$\dim(\text{im } \varphi) = \dim(V / \ker \varphi) = \dim(V) - \dim(\ker \varphi).$$

□

[Remark: In elementary linear algebra the subspaces $\ker \varphi \subseteq V$ and $\text{im } \varphi \subseteq W$ are called the *nullspace* and the *range* of the linear function φ . The dimensions $\dim(\ker \varphi)$ and $\dim(\text{im } \varphi)$ are called the *nullity* and the *rank* of φ . Hence the name of the theorem. The elementary proof (which is quite different from our proof) uses the Reduced Row Echelon Form of the corresponding $\dim(W) \times \dim(V)$ matrix $[\varphi]$ to show that

$$\begin{aligned} \dim(\text{im } \varphi) &= \#(\text{pivot columns in RREF of } [\varphi]) \\ \dim(\ker \varphi) &= \#(\text{non-pivot columns in RREF of } [\varphi]), \end{aligned}$$

and hence

$$\dim(\text{im } \varphi) + \dim(\ker \varphi) = \#(\text{columns in } [\varphi]) = \dim(V).$$

The most important consequence of this theorem says that if $\dim(V) = \dim(W)$ (i.e., if $[\varphi]$ is a **square** matrix) then we have

$$\begin{aligned} (\varphi \text{ is injective}) &\iff \ker \varphi = \{\mathbf{0}\} \\ &\iff \dim(\ker \varphi) = 0 \\ &\iff \dim(\text{im } \varphi) = \dim(V) \\ &\iff \dim(\text{im } \varphi) = \dim(W) \\ &\iff \text{im } \varphi = W \\ &\iff (\varphi \text{ is surjective}). \end{aligned}$$

This can be used to show that that $AB = I$ implies $BA = I$ for any square matrices A, B over a field. Of course, we already had a slightly easier proof of that fact.]

Week 11

This week we will apply the Orbit-Stabilizer Theorem to a group acting on itself. Recall that a group G acts on itself in two basic ways:

- **Translation.** For any $g \in G$ we define the function $\tau_g : G \rightarrow G$ by $\tau_g(a) := ga$. Then one can show that the map $g \mapsto \tau_g$ defines a group homomorphism

$$\tau : G \rightarrow \text{Perm}(G).$$

- **Conjugation.** For any $g \in G$ we define the function $\kappa_g : G \rightarrow G$ by $\kappa_g(a) := gag^{-1}$. Then one can show that the map $g \mapsto \kappa_g$ defines a group homomorphism

$$\kappa : G \rightarrow \text{Aut}(G).$$

First let's deal with translation.

Orbit-Stabilizer for Translation. We already know that the kernel of τ is trivial, which implies that G is isomorphic to the group of permutations $\text{im } \tau \subseteq \text{Perm}(G)$. (Jargon: We say that τ is a *faithful action*. This is another way to state Cayley's Theorem.) Now I claim that translation is free and transitive.²⁰

Proof.

- **Free.** For all $a \in G$ we want to show that $\text{Stab}_\tau(a) \subseteq G$ is the trivial group. So consider any element $g \in \text{Stab}_\tau(a)$. By definition we have $a = \tau_g(a) = ga$, and multiplying by a^{-1} on the right gives $g = \varepsilon$. We conclude that $\text{Stab}_\tau(a) = \{\varepsilon\}$ for all $a \in G$.
- **Transitive.** For all $a, b \in G$ we want to show that there exists some group element $g \in G$ with $\tau_g(a) = b$. Simply take $g = ba^{-1}$. Then we have

$$\tau_g(a) = \tau_{ba^{-1}}(a) = (ba^{-1})a = b.$$

□

Now that we know the orbits and stabilizers, let's see what the Orbit-Stabilizer Theorem tells us. For each $a \in G$ we have $\text{Orb}_\tau(a) = G$ and $G/\text{Stab}_\tau(a) = G/\{\varepsilon\} = G$. Thus we obtain a bijection from G to itself:

$$\begin{aligned} G = \text{Orb}_\tau(a) &\longleftrightarrow G/\text{Stab}_\tau(a) = G \\ ga &\longleftrightarrow g. \end{aligned}$$

Note that we can explicitly describe the bijection $G = G/\text{Stab}_\tau(a) \rightarrow \text{Orb}_\tau(a) = G$ as **multiplication on the right by a** . It's a bit interesting that multiplication on the right

²⁰We already proved this for the abelian group $G = (\mathbb{R}^n, +, \mathbf{0})$. Now we'll show that it holds in general.

comes into play (since the action is by left multiplication). Otherwise, there's not much going on here. ///

Orbit-Stabilizer for conjugation is much more interesting.

Orbit-Stabilizer for Conjugation: The Class Equation. Recall that the kernel of the conjugation action is the set of group elements that commute with everything. We call this the *center* [Z is for *Zentrum*] of G :

$$\begin{aligned} Z(G) &:= \ker \kappa = \{g \in G : \kappa_g = \text{id}\} \\ &= \{g \in G : \kappa_g(a) = a \text{ for all } a \in G\} \\ &= \{g \in G : gag^{-1} = a \text{ for all } a \in G\} \\ &= \{g \in G : ga = ag \text{ for all } a \in G\}. \end{aligned}$$

Being a kernel, we know that the center $Z(G) \trianglelefteq G$ is a normal subgroup. Observe that $Z(G) = G$ if and only if G is abelian. The orbits and stabilizers also have special names:

- **Conjugacy Classes.** For all $a \in G$ we define the *conjugacy class* [K is for *Klasse*]:

$$K(a) := \text{Orb}_\kappa(a) = \{gag^{-1} : g \in G\}.$$

- **Centralizers.** For all $a \in G$ we define the *centralizer* [Z is for *Zentrum* again]:

$$Z(a) := \text{Stab}_\kappa(a) = \{g \in G : gag^{-1} = a\} = \{g \in G : ga = ag\}.$$

Thus for each group element $a \in G$ the Orbit-Stabilizer Theorem gives us a bijection between elements of the conjugacy class $K(a)$ and the set $G/Z(a)$ of left cosets of the centralizer:

$$\begin{aligned} K(a) &\longleftrightarrow G/Z(a) \\ gag^{-1} &\longleftrightarrow gZ(a). \end{aligned}$$

We will combine these facts to obtain a useful formula. First, observe for all $a \in G$ that

$$K(a) = \{a\} \iff Z(a) = G \iff a \in Z(G).$$

This suggests that we should collect the singleton conjugacy classes together. If $a_1, a_2, \dots, a_k \in G$ is an arbitrary system of conjugacy class representatives, then we obtain a disjoint union:

$$\begin{aligned} G = \coprod K(a_i) &= \coprod_{K(a_i)=\{a_i\}} \{a_i\} \cup \coprod_{K(a_i) \neq \{a_i\}} K(a_i) \\ &= \coprod_{a_i \in Z(G)} \{a_i\} \cup \coprod_{K(a_i) \neq \{a_i\}} K(a_i) \\ &= Z(G) \cup \coprod_{K(a_i) \neq \{a_i\}} K(a_i). \end{aligned}$$

And if G is finite then we apply the Orbit-Stabilizer Theorem to obtain

$$\begin{aligned}\#G &= \#Z(G) + \sum_{K(a_i) \neq \{a_i\}} \#K(a_i) \\ \#G &= \#Z(G) + \sum_{Z(a_i) \neq G} \#G/\#Z(a_i).\end{aligned}$$

This last formula is called the *class equation*. It is surprisingly useful. ///

The main application of the “class equation” is to study how the **size** of a finite group affects its **structure**. This general topic is called “Sylow theory.” I have decided not to go very far in this direction; the next theorem will give you just a taste.

Theorem (Groups of size p^2). Let $p \in \mathbb{Z}$ be prime and let G be a group of size p^2 . Then:

- (1) G is abelian,
- (2) G is isomorphic to $\mathbb{Z}/p^2\mathbb{Z}$ or $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.

///

To prove this we need two basic lemmas.

Lemma 1. Any group of size p is cyclic.

Proof. Let $\#G = p$ be prime and consider any non-identity element $\varepsilon \neq g \in G$. By Lagrange’s Theorem, the cyclic subgroup $\langle g \rangle \subseteq G$ has size dividing p . Since p is prime this means that $\#\langle g \rangle = 1$ or $\#\langle g \rangle = p$. But since $g \neq \varepsilon$ we know that $\langle g \rangle \neq \{\varepsilon\}$, and it follows that $\#\langle g \rangle = p$. Finally, since $\langle g \rangle \subseteq G$ and $\#\langle g \rangle = \#G$ we conclude that $G = \langle g \rangle$. □

Lemma 2. If the quotient group $G/Z(G)$ is cyclic then G is abelian.

Proof. Recall that $Z(G) \trianglelefteq G$ is a normal subgroup. Assume that the quotient $G/Z(G)$ is a cyclic group. This means there exists an element $g \in G$ such that every left coset of $Z(G)$ has the form $(gZ(G))^k = g^kZ(G)$ for some $k \in \mathbb{Z}$. Then since the cosets cover G it follows that every element of G has the form g^kz for some $k \in \mathbb{Z}$ and $z \in Z(G)$. Finally, if $g^{k_1}z_1$ and $g^{k_2}z_2$ are any two elements of G then since z_1, z_2 commute with everything, and since

$$g^{k_1}g^{k_2} = g^{k_1+k_2} = g^{k_2+k_1} = g^{k_2}g^{k_1},$$

we conclude that

$$(g^{k_1} z_1)(g^{k_2} z_2) = g^{k_1} g^{k_2} z_1 z_2 = g^{k_2} g^{k_1} z_2 z_1 = (g^{k_2} z_2)(g^{k_1} z_1).$$

□

Proof of the Theorem. Let $p \in \mathbb{Z}$ be prime and let G be a group of size p^2 .

(1) To prove that G is abelian, we consider the class equation:

$$p^2 = \#G = \#Z(G) + \sum_{Z(a_i) \neq G} \#G/\#Z(a_i)$$

Let $Z(a_i) \subseteq G$ be any centralizer. From Lagrange's Theorem we know that $\#Z(a_i)$ divides $\#G = p^2$, which implies that $\#Z(a_i) \in \{1, p, p^2\}$. But if $Z(a_i) \neq G$ then we must have $\#Z(a_i) \in \{1, p\}$ and hence $\#G/\#Z(a_i) \in \{p, p^2\}$. Thus p divides the sum

$$\sum_{Z(a_i) \neq G} \#G/\#Z(a_i),$$

which implies that p divides the size of the center:

$$\#Z(G) = p^2 - \sum_{Z(a_i) \neq G} \#G/\#Z(a_i) = p^2 - (\text{some multiple of } p).$$

Since $Z(G) \trianglelefteq G$ is a subgroup, Lagrange tells us that $\#Z(G) \in \{1, p, p^2\}$ and the previous formula tells us that $\#Z(G) \in \{p, p^2\}$. Thus there are two possible cases:

- If $\#Z(G) = p^2$ then we have $Z(G) = G$ which implies that G is abelian as desired.
- If $\#Z(G) = p$ then G is not abelian because $Z(G) \neq G$. I will show that **this case is impossible**. Indeed, if $\#Z(G) = p$ then we must have $\#(G/Z(G)) = \#G/\#Z(G) = p^2/p = p$. But then Lemma 1 says that $G/Z(G)$ is cyclic and Lemma 2 says that G is abelian. Contradiction.

(2) Now we will prove that $G \cong \mathbb{Z}/p^2\mathbb{Z}$ or $G \cong \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. For each non-identity element $g \in G - \{\varepsilon\}$ we know from Lagrange's Theorem that the order $\#\langle g \rangle \neq 1$ divides $\#G = p^2$ and hence $\#\langle g \rangle \in \{p, p^2\}$. Now there are two cases:

- Suppose that there exists some element $g \in G - \{\varepsilon\}$ such that $\#\langle g \rangle = p^2$. Then we conclude that $G = \langle g \rangle \cong \mathbb{Z}/p^2\mathbb{Z}$.
- Otherwise we must have $\#\langle g \rangle = p$ for all $g \in G - \{\varepsilon\}$. So choose some arbitrary element $g \in G - \{\varepsilon\}$ and then choose an arbitrary element $h \in G - \langle g \rangle$. I claim that G is an internal direct product:

$$G = \langle g \rangle \times \langle h \rangle \cong \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}.$$

To see this, first note that $\langle g \rangle \cap \langle h \rangle \subseteq \langle g \rangle$ is a subgroup. Thus by Lagrange we have

$$\#(\langle g \rangle \cap \langle h \rangle) \in \{1, p\}.$$

If $\#(\langle g \rangle \cap \langle h \rangle) = p$ then we have $\langle g \rangle \cap \langle h \rangle = \langle g \rangle$ which contradicts the fact that $h \notin \langle g \rangle$. Therefore $\langle g \rangle \cap \langle h \rangle = \{\varepsilon\}$. Now consider the multiplication map

$$\begin{aligned} \mu : \langle g \rangle \times \langle h \rangle &\rightarrow G \\ (g^k, h^\ell) &\mapsto g^k h^\ell. \end{aligned}$$

Since $\langle g \rangle \cap \langle h \rangle = \{\varepsilon\}$ we know that μ is injective, hence the image $\langle g \rangle \langle h \rangle = \text{im } \mu \subseteq G$ has size $\#(\langle g \rangle \times \langle h \rangle) = p^2$, which implies that $G = \langle g \rangle \langle h \rangle$. Finally, since G is abelian we know that each of $\langle g \rangle \trianglelefteq G$ and $\langle h \rangle \trianglelefteq G$ is normal.

□

Remarks:

- The first part of the theorem fails for higher powers of p . For example, not every group of size 2^3 is abelian. Proof: D_8 is not abelian.
- For **abelian** groups of size p^k , the second part of the theorem still holds. That is, any abelian group of size p^k is a direct product of cyclic groups. The different ways to do this correspond to the different partitions of the integer k . For example, here are the non-isomorphic abelian groups of size p^4 :

- $\mathbb{Z}/p^4\mathbb{Z}$,
- $\mathbb{Z}/p^3\mathbb{Z} \times \mathbb{Z}/p$,
- $\mathbb{Z}/p^2\mathbb{Z} \times \mathbb{Z}/p^2\mathbb{Z}$,
- $\mathbb{Z}/p^2\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$,
- $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.

This result is rightly seen as a theorem of advanced linear algebra, which is outside the scope of this course. The correct proof uses the Smith Normal Form of a matrix over \mathbb{Z} .

- You might wonder if there is a formula for counting these abelian p -groups. Let $P(k)$ be the number of ways to partition the integer k , i.e., the number of non-isomorphic abelian groups of size p^k . Hardy and Ramanujan proved in 1918 that this number satisfies

$$P(k) \sim \frac{1}{4k\sqrt{3}} \cdot \exp\left(\pi\sqrt{\frac{2k}{3}}\right) \quad \text{as } k \rightarrow \infty.$$

There is no closed formula.

- It is less difficult to prove that every finite abelian group is a direct product of abelian p -groups. You will prove this on the next homework, assuming that the famous “Sylow Theorems” are true.

I will just state the Sylow Theorems without proof and give a slick application. These are named for the Norwegian mathematician Ludwig Sylow (1832–1918). One can view these theorems as a partial converse to Lagrange’s Theorem.

The Sylow Theorems (1872). Let G be a finite group of size $\#G = p^k m$ where p is prime and $\gcd(p, m) = 1$. Then:

- (1) There exists at least one subgroup $H \subseteq G$ of size $\#H = p^k$.
- (2) Any two subgroups H_1, H_2 of size p^k are conjugate. In other words, there exists some $g \in G$ such that $H_2 = gH_1g^{-1}$. It follows that a subgroup of size p^k is normal if and only if it is unique.
- (3) Let n_p be the number of subgroups of size p^k . Then $n_p | m$ and $n_p \equiv 1 \pmod{p}$.

///

You can find a proof in any advanced textbook on group theory. Here’s a cute application.

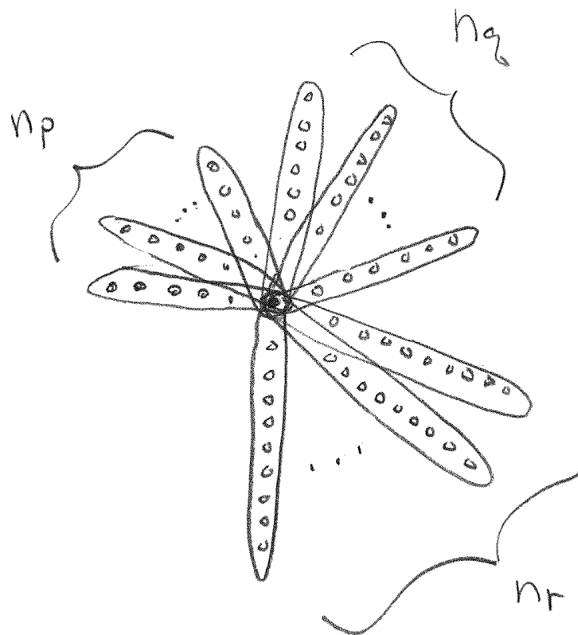
Application of Sylow. Let p, q, r be distinct primes. Any group of size pqr is **not** simple.

Proof. Suppose that $\#G = pqr$ with $p < q < r$ prime. Let n_p, n_q, n_r be the numbers of subgroups of size p, q, r , respectively. If any of n_p, n_q, n_r equals 1 then from Sylow (2) we obtain a non-trivial normal subgroup. So assume for contradiction that $n_p, n_q, n_r > 1$. Then from Sylow (3) we have $n_r = pq$, $n_q \in \{r, pr\}$ and $n_p \in \{q, r, qr\}$, hence $n_q \geq r$ and $n_p \geq q$. By Lagrange’s Theorem we see that any two subgroups with sizes in $\{p, q, r\}$ intersect trivially (i.e., they are equal or they intersect at the identity). By counting the elements of these subgroups we obtain

$$\begin{aligned}
 pqr = \#G &\geq n_p(p-1) + n_q(q-1) + n_r(r-1) + 1 \\
 &\geq q(p-1) + r(q-1) + pq(r-1) + 1 \\
 &= (pq - q) + (qr - q) + (pqr - pq) + 1 \\
 &= pqr + (qr - q - r + 1) \\
 &= pqr + (q-1)(r-1),
 \end{aligned}$$

and hence $0 \geq (q-1)(r-1)$. This contradicts the fact that $(q-1) > 0$ and $(r-1) > 0$. \square

Here is a picture of the counting method we used:



Total # Dots:

$$\begin{aligned}
 & n_p(p-1) \\
 & + n_q(q-1) \\
 & + n_r(r-1) \\
 & + 1
 \end{aligned}$$

[Remark: Using similar tricks with Sylow theory, one can show that no non-abelian group of size < 60 is simple. Some people think these tricks make good exam problems but I don't agree. I prefer to ask about generalities.]

Week 12

To end this course, I want to complete our discussion of the quintic equation and the icosahedral group. Namely, I will prove that the group $I \subseteq SO(3)$ of rotational symmetries of a regular icosahedron is a **simple group**.²¹ This is related to the solvability of the quintic equation because of an “accidental isomorphism” with the alternating group A_5 :

$$I \cong A_5.$$

Before discussing the group A_5 , I will state a general theorem about alternating groups.

If G is a finite group, recall that a *composition series* consists of a chain of subgroups

$$G = G_0 \supsetneq G_1 \supsetneq \cdots \supsetneq G_\ell = \{\varepsilon\}$$

in which each quotient group G_i/G_{i+1} exists and is **simple** (i.e., has no non-trivial normal subgroup). Recall from the Jordan-Hölder Theorem that the list of simple groups $\{G_i/G_{i+1}\}_i$

²¹In fact, the infinite group $SO(3)$ is also simple, but this is beyond the scope of the course.

is the same (up to isomorphism and permutation) for any two composition series of G . These unique simple groups are called the *composition factors* of G . Recall further that the group G is called *solvable* when its composition factors are abelian (i.e., have the form $\mathbb{Z}/p\mathbb{Z}$ for various primes p). We have already proved that the symmetric group S_n is not solvable when $n \geq 5$. Now we will be more specific.

Theorem (Composition Factors of S_n). Let $n \geq 5$ and consider the symmetric group S_n . Let $A_n \subseteq S_n$ be the alternating subgroup. Then we have:

- $A_n \trianglelefteq S_n$ is the only non-trivial normal subgroup of S_n ,
- A_n is simple.

It follows from this that the group S_n has a unique composition series:

$$S_n \supsetneq A_n \supsetneq \{\text{id}\}.$$

Hence the simple composition factors are $S_n/A_n \cong \mathbb{Z}/2\mathbb{Z}$ and $A_n/\{\text{id}\} \cong A_n$. ///

Remarks:

- Now we see that the group S_n is not solvable (for $n \geq 5$) because the composition factor A_n is not abelian. It's just an accident of nature that all simple groups smaller than A_5 are abelian.
- One might even say that the group S_n is “almost simple,” except for the piddly factor of $\mathbb{Z}/2\mathbb{Z}$. After all, $\#A_n = n!/2$ is **much** larger than $\#\mathbb{Z}/2\mathbb{Z} = 2$.
- Similar theorems say that the matrix groups $GL_n(\mathbb{F})$, $O(n)$ and $U(n)$ are “almost simple” (i.e., simple except for a piddly quotient). However these results are quite involved and are never proved in undergraduate courses.
- Some undergraduate books do give a proof that A_n is simple (for $n \geq 5$), but this proof is also not very nice. Michael Artin only included the proof in the **second** edition of his book. I think it's fair to omit the proof entirely. In this course we will only discuss the special case $n = 5$.

Now we will prove the theorem in the case $n = 5$. In other words, we will prove that

- A_5 is the only non-trivial normal subgroup of S_5 ,
- A_5 is simple.

Both of these proofs will use the following trick.

Trick. Let $N \trianglelefteq G$ be a normal subgroup. Then N is a union of conjugacy classes.

Proof. Let $\kappa : G \rightarrow \text{Aut}(G)$ be the conjugation action consider any element $n \in N$. Then since N is normal, we have $gng^{-1} \in N$ for all $g \in G$ and hence

$$\text{Orb}_\kappa(n) = \{gng^{-1} : g \in G\} \subseteq N.$$

It follows that

$$\bigcup_{n \in N} \text{Orb}_\kappa(n) \subseteq N.$$

On the other hand, we obviously have

$$N \subseteq \bigcup_{n \in N} \text{Orb}_\kappa(n)$$

because $n \in \text{Orb}_\kappa(n)$ for all $n \in N$. □

The reason this trick is useful is because sometimes we can compute the sizes of all the conjugacy classes. Then by using Lagrange's Theorem we can dramatically narrow the search for normal subgroups. To see how this works, let's compute the sizes of the conjugacy classes in the symmetric group. We might as well do this for general n .

Theorem (Sizes of Conjugacy Classes in S_n). Let $\kappa : S_n \rightarrow \text{Aut}(S_n)$ be the conjugation action and consider any permutation $f \in S_n$. Recall that the conjugacy class $K(f) := \text{Orb}_\kappa(f)$ consists of all permutations that have the same "cycle structure" as f (i.e., the same number of cycles of each length). To be specific, let's say that the cycle decomposition of f contains m_i cycles of length i , for each $i \in \{1, 2, \dots, n\}$. Then we have

$$\#K(f) = \frac{n!}{1^{m_1} m_1! 2^{m_2} m_2! \dots n^{m_n} m_n!}.$$

///

Before proving this, let's test some simple examples. Note that the identity permutation $\text{id} \in S_n$ has $m_1 = n$ cycles of length 1 and $m_i = 0$ cycles of length i for each $i \in \{2, 3, \dots, n\}$. Thus the formula gives

$$\#K(\text{id}) = \frac{n!}{1^n n! 2^0 0! \dots n^0 0!} = \frac{n!}{n!} = 1.$$

This is correct because the identity is only conjugate to itself. Next, let's count the conjugacy class of transpositions (2-cycles), which has $m_1 = n - 2$, $m_2 = 1$ and $m_i = 0$ for $i \in \{3, \dots, n\}$. If $t \in S_n$ is any transposition then the formula gives

$$\#K(t) = \frac{n!}{1^{n-2} (n-2)! 2^1 1! 3^0 0! \dots n^0 0!} = \frac{n!}{2(n-2)!} = \frac{n!}{2!(n-2)!} = \binom{n}{2}.$$

This is correct because each transposition $(ij) \in S_n$ corresponds to a choice of two elements $i \neq j$ from the set $\{1, 2, \dots, n\}$.

Proof of the Theorem. We will use the Orbit-Stabilizer Theorem. Let $f \in S_n$ and recall that the stabilizer under conjugation is called the *centralizer*:

$$Z(f) := \text{Stab}_\kappa(f) = \{g \in S_n : gfg^{-1} = f\}.$$

Now suppose that the permutation f has m_i cycles of length i for each $i \in \{1, 2, \dots, n\}$. By Orbit-Stabilizer we have $\#K(f) = \#S_n/\#Z(f) = n!/\#Z(f)$, thus our goal is to prove that

$$\#Z(f) = 1^{m_1}m_1!2^{m_2}m_2!\cdots n^{m_n}m_n!.$$

To see this, suppose that (j_1, j_2, \dots, j_i) is one of the cycles of f . This means that

$$f(j_1) = j_2, \quad f(j_2) = j_3, \quad \cdots \quad f(j_{m-1}) = j_m \quad \text{and} \quad f(j_m) = j_1.$$

Then for any $g \in S_n$ we see that $(g(j_1), g(j_2), \dots, g(j_i))$ is a cycle of gfg^{-1} . (You proved this on a previous homework.) If $g \in Z(f)$ (i.e., if $gfg^{-1} = f$) then this cycle must equal one of the cycles of f . If there is only one cycle of length i (i.e., if $m_i = 1$) then we must have

$$(j_1, j_2, \dots, j_i) = (g(j_1), g(j_2), \dots, g(j_i)).$$

In this case there are exactly i ways to choose the values $g(j_1), g(j_2), \dots, g(j_i) \in \{j_1, j_2, \dots, j_i\}$ since we are only allowed to rotate the cycle. If $m_i > 1$ then we can also permute the various i -cycles. There are $m_i!$ ways to do this and then there are $i \cdot i \cdot i \cdots i = i^{m_i}$ ways to rotate each of the cycles. Hence there are $i^{m_i}m_i!$ different ways to choose the values inside the i -cycles of gfg^{-1} . Since the choices for different values of i are independent, the total number of ways to choose a permutation $g \in Z(f)$ is

$$\#Z(f) = \prod_{i=1}^n \#(\text{ways to fill the } i\text{-cycles}) = \prod_{i=1}^n i^{m_i}m_i!.$$

□

The notation in that proof is terrible. Hopefully an example will be more convincing.

Example: Conjugacy Classes and Normal Subgroups of S_5 . Recall that the conjugacy classes of S_5 consist of permutations with the same number of i -cycles for each i . There are two equivalent ways to encode this “cycle structure.” First, since the order of the cycles doesn’t matter we will record the lengths of the cycles in a vector $\lambda = \lambda_1\lambda_2\lambda_3\lambda_4\lambda_5$, where

- $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq \lambda_5 \geq 0$,
- $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 5$.

Thus the set of possible vectors λ is $\{50000, 41000, 32000, 31100, 22100, 21110, 11111\}$. Second, we will write $m = m_1m_2m_3m_4m_5$, where m_i is the number of cycles of length i . These numbers must satisfy

- $m_i \geq 0$ for all $i \in \{1, 2, 3, 4, 5\}$,
- $1m_1 + 2m_2 + 3m_3 + 4m_4 + 5m_5 = 5$.

Thus the set of possible vectors m is $\{00001, 10010, 01100, 20100, 12000, 31000, 50000\}$. Now we have the following table recording the sizes of the conjugacy classes and centralizers in S_5 :

λ	m	$\#Z$	$\#K$
50000	00001	$5^1 \cdot 1! = 5$	$120/5 = 24$
41000	10010	$1^1 \cdot 1! \cdot 4^1 \cdot 1! = 4$	$120/4 = 30$
32000	01100	$2^1 \cdot 1! \cdot 3^1 \cdot 1! = 6$	$120/6 = 20$
31100	20100	$1^2 \cdot 2! \cdot 3^1 \cdot 1! = 6$	$120/6 = 20$
22100	12000	$1^1 \cdot 1! \cdot 2^2 \cdot 2! = 8$	$120/8 = 15$
21110	31000	$1^3 \cdot 3! \cdot 2^1 \cdot 1! = 12$	$120/12 = 10$
11111	50000	$1^5 \cdot 5! = 120$	$120/120 = 1$

Note that $\lambda = 11111$ corresponds to the conjugacy class $\{\text{id}\}$ and $\lambda = 21110$ corresponds to the conjugacy class of 2-cycles $\{(12), (13), \dots, (45)\}$, which has $\binom{5}{2} = 10$ elements. For the rest of the calculation, we can tell that we didn't make a mistake because the sizes of the conjugacy classes add up to the size of the group:

$$24 + 30 + 20 + 20 + 15 + 10 + 1 = 120 = 5! = \#S_5.$$

I find this example more convincing than the general proof above. That's often how it goes with combinatorics. Now let's use this information to find all the normal subgroups.

Theorem. The alternating group $A_5 \trianglelefteq S_5$ is the only non-trivial normal subgroup of S_5 .

Proof. Recall that a normal subgroup $N \trianglelefteq S_5$ is a union of conjugacy classes, which must include the class $\{\text{id}\}$. We also know from Lagrange's Theorem that $\#N$ divides $\#S_5$. Let $K_\lambda \subseteq S_5$ be the conjugacy class with cycle type λ . Then combining all of these restrictions leaves only three possible normal subgroups:

$$\begin{aligned} N &= K_{11111} \cup K_{22100} \cup K_{50000}, \\ N' &= K_{11111} \cup K_{22100} \cup K_{31100} \cup K_{50000}, \\ N'' &= K_{11111} \cup K_{22100} \cup K_{32000} \cup K_{50000}. \end{aligned}$$

It is easy to check that $N' = A_5$ and that the sets $N, N'' \subseteq S_5$ are **not subgroups**. It follows that $A_5 \trianglelefteq S_5$ is the only non-trivial normal subgroup of S_5 . \square

Obviously this proof won't work for higher values of n . For general n we should first prove that A_n is simple, then use that fact to prove that S_n has no other normal subgroups. (See

the homework.) We will only prove this for $n = 5$ because I don't know a nice general proof.²² The conjugacy classes of A_5 are a bit tricky to describe so we will use the following strategy:

- (1) Prove that the icosahedral group $I \subseteq SO(3)$ is simple.
- (2) Then prove that $I \cong A_5$. This isomorphism is just a lucky accident, resulting from the fact that 60 is a relatively small number. Felix Klein made a big deal of this lucky accident in his *Lectures on the Icosahedron* (1888).

Theorem. The icosahedral group $I \subseteq SO(3)$ is simple.

Proof. We will use the fact that the conjugacy classes have geometric meaning. Recall that two invertible matrices $A, B \in GL_n(\mathbb{R})$ are conjugate if and only if they represent the same linear function after a change of basis. More specifically, two matrices $A, B \in I$ are conjugate in I if and only if they represent the same function after a rotational symmetry of the icosahedron. Thus we obtain the following table of conjugacy classes:

description of the conjugacy class	number of elements
{id}	1
{rotate by $\pm 2\pi/5$ around a vertex}	12
{rotate by $\pm 4\pi/5$ around a vertex}	12
{rotate by π around an edge}	15
{rotate by $\pm 2\pi/3$ around a face}	20

We know that we didn't make a mistake because

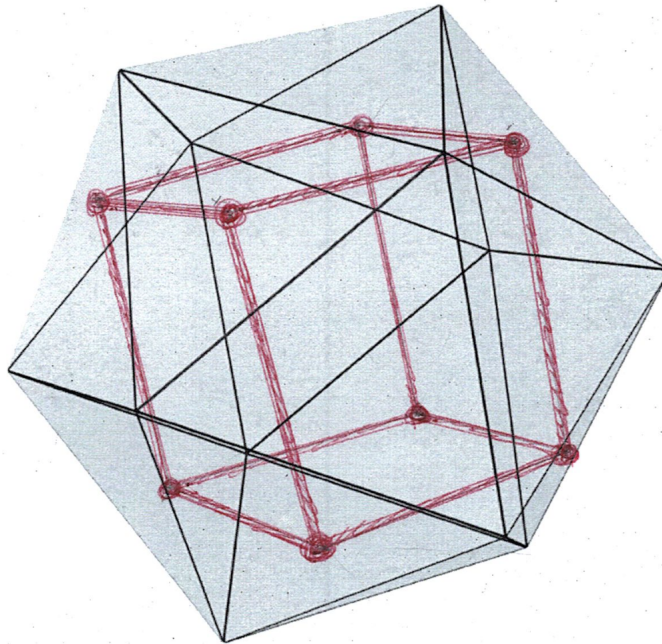
$$1 + 12 + 12 + 15 + 20 = 60 = \#I.$$

Now let's look for normal subgroups. Recall that any normal subgroup $N \trianglelefteq I$ must be a union of conjugacy classes, which must include the identity class {id}. Furthermore, we know from Lagrange that $\#N$ divides $\#I$. It is easy to check that there is no non-trivial solution to this combinatorial problem. \square

Theorem. The icosahedral group $I \subseteq SO(3)$ is isomorphic to the alternating group $A_5 \subseteq S_5$.

Proof. I will follow the proof from Artin's book. The proof relies on the strange fact that 5 cubes can be inscribed in a regular icosahedron. To be specific, consider the midpoints of the 20 triangular faces. It turns out that certain collections of 8 midpoints can be joined up into a cube, and that there are exactly 5 different ways to do this. Here is one of the cubes:

²²See the second edition of Artin for a not-nice proof.



Now any isometry $f \in I$ will send cubes to cubes, therefore we obtain a group homomorphism

$$\varphi : I \rightarrow \text{Perm}(\{5 \text{ cubes}\}) \cong S_5.$$

Since $\ker \varphi \trianglelefteq I$ is a normal subgroup and since I is **simple**, we must have $\ker \varphi = \{\text{id}\}$ or $\ker \varphi = I$. But the second choice is impossible because clearly some element of I moves the cubes. Therefore we have $\ker \varphi = \{\text{id}\}$ and hence φ is injective. It follows from the First Isomorphism Theorem that I is isomorphic to its image:

$$I \cong I/\ker \varphi \cong \text{im } \varphi \subseteq S_5.$$

Now it only remains to show that $\text{im } \varphi = A_5$. To do this we consider the determinant homomorphism $\det : S_n \rightarrow \{\pm 1\}$ and recall that $A_5 = \ker(\det)$. Consider the composition of these homomorphisms:

$$\det \circ \varphi : I \rightarrow \{\pm 1\}.$$

Again, since $\ker(\det \circ \varphi) \trianglelefteq I$ is a normal subgroup and since I is **simple**, we must have $\ker(\det \circ \varphi) = \{\text{id}\}$ or $\ker(\det \circ \varphi) = I$. This time the first choice is impossible because I has more elements than $\{\pm 1\}$. It follows that $\ker(\det \circ \varphi) = I$ and hence

$$\text{im } \varphi \subseteq \ker(\det) = A_5 \subseteq S_5.$$

Finally, since $\#\text{im } \varphi = \#I = \#A_5 = 60$, we conclude that $\text{im } \varphi = A_5$ as desired. \square

This concludes our study of S_5 and the icosahedron. Next semester we will see what this has to do with the general quintic equation.

But I don't want to end it there. Let me just say a few final words about simple groups. Recall from the Jordan-Hölder Theorem that every finite group G has a unique collection of simple composition factors. These are something like the “prime factors” of the group. This suggests a strategy for classifying all finite groups, which is sometimes called *Hölder's Program* because the project was begun by Otto Hölder (1859–1937):

- Classify all finite simple groups.
- Describe all ways of putting them together.

The second problem is far too difficult to have a nice solution. The first problem, on the other hand, turns out to be solvable. After 100 years of intense work by generations of group theorists, the full classification of finite simple groups was announced by Daniel Gorenstein in 1983. The details are complicated but the general outline is easy to describe.

Theorem (The Classification of Finite Simple Groups). There exist three infinite families of finite simple groups:

- Cyclic groups $\mathbb{Z}/p\mathbb{Z}$ for p prime.
- Alternating groups A_n for $n \geq 5$.
- Groups related to $GL_n(\mathbb{Z}/p\mathbb{Z})$. This includes finite versions of the orthogonal and unitary groups, together with a few strange families that we need not mention.²³

On top of this, there are exactly 26 so-called “sporadic groups,” which are not related to any of the infinite families. The largest of these is the *Monster group* \mathbb{M} which has approximately 8×10^{53} elements. ///

The amount of work involved in the classification is mind-boggling. The original proof was spread over tens of thousands of journal pages. Right now some group theorists are working on a “second generation proof,” which is estimated to fill about 5000 pages. It's fair to say that the mathematical community is far from understanding all the details.

As for infinite groups: If we had a few more weeks, I would like to discuss the relationship between the continuous groups $SU(2)$ and $SO(3)$. This is beautiful topic related to geometry and physics.²⁴ Of course, we do have another whole semester together, but that semester will be devoted to a completely different topic (rings, fields and polynomials). See you then.

²³Fine, I'll mention them. There is one more “classical” family $Sp(n)$ coming from the quaternions and then five “exceptional” families called G_2, F_4, E_6, E_7, E_8 .

²⁴I strongly recommend John Stillwell's *Naive Lie Theory*.

Problem Set 6

1. The Alternating Group A_4 is Not Simple. Recall that $A_4 \subseteq S_4$ is the subgroup of permutations of $\{1, 2, 3, 4\}$ which can be expressed as the product of an even number of transpositions.

(a) Prove that the following set is a normal subgroup:

$$V = \{\text{id}, (12)(34), (13)(24), (14)(23)\} \trianglelefteq A_4.$$

It follows that A_4 is not a simple group.

(b) Furthermore, prove that $V \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. The letter V is for *Klein's Viergruppe*. [Once upon a time it was surprising that not every abelian group is cyclic.]

(a) First note that V is a subset of A_4 since each element can be expressed by an even number of transpositions. Next observe that V is a union of two conjugacy classes in S_4 :

$$\begin{aligned} V &= \{\text{id}\} \cup \{(12)(34), (13)(24), (14)(23)\} \\ &= K_{1111} \cup K_{22} \\ &= \{\text{permutations with four one-cycles}\} \cup \{\text{permutations with two 2-cycles}\}. \end{aligned}$$

Therefore if $V \subseteq A_4$ is a subgroup then it will necessarily be normal. So let's check:

- **Identity.** We have $\text{id} \in V$ by definition.
- **Inverse.** Note that every element $f \in V$ has order one or two, hence $f^{-1} = f \in V$.
- **Closed.** This is the hardest thing to check and it only works because of a lucky accident of small numbers:

$$\begin{aligned} (12)(34) \circ (13)(24) &= (14)(23), \\ (12)(34) \circ (14)(23) &= (13)(24), \\ (13)(24) \circ (14)(23) &= (12)(34). \end{aligned}$$

(b) We know that there are exactly two groups of size p^2 for any prime p . Thus to show that $V \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ we only need to show that $V \not\cong \mathbb{Z}/4\mathbb{Z}$. And this is true because V contains no element of order 4. Alternatively, one can check directly that V is the internal direct product of any two distinct non-trivial subgroups, say $\langle (12)(34) \rangle$ and $\langle (13)(24) \rangle$.

2. Primary Factorization of a Finite Abelian Group. Let G be finite abelian group.

(a) Suppose that there exist subgroups $H, K \subseteq G$ such that $\#G = \#H \cdot \#K$ and $\gcd(\#H, \#K) = 1$. In this case, prove that G is an internal direct product:

$$G = H \times K.$$

- (b) Now suppose that $\#G = p_1^{e_1} \cdots p_n^{e_n}$ for distinct primes p_1, \dots, p_n . The Sylow Theorems tell us that for each i there exists a unique subgroup $H_i \subseteq G$ of size $\#H_i = p_i^{e_i}$. Use part (a) and induction to prove that G is the direct product of these subgroups:

$$G = H_1 \times H_2 \times \cdots \times H_n.$$

This is called the *primary factorization* of G . It is also true that each *primary factor* H_i is a product of cyclic subgroups but this is harder to prove.

- (c) In the special case that G is **cyclic**, prove that

$$G \cong \frac{\mathbb{Z}}{p_1^{e_1}\mathbb{Z}} \times \frac{\mathbb{Z}}{p_2^{e_2}\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{p_n^{e_n}\mathbb{Z}}.$$

This is a non-constructive version of the Chinese Remainder Theorem.

- (a) Let G be a finite abelian group and let $H, K \subseteq G$ be subgroups with $\gcd(\#H, \#K) = 1$ and $\#H \cdot \#K = \#G$. Since $H \cap K$ is a subgroup of H and K , Lagrange's Theorem tells us that $\#(H \cap K)$ is a common divisor of $\#H$ and $\#K$, hence $\#(H \cap K) = \pm 1$. It follows that $\#(H \cap K) = 1$ and hence $H \cap K = \{\varepsilon\}$.

This implies that the multiplication map $\mu : H \times K \rightarrow G$ is injective and hence

$$\#HK = \#\text{im } \mu = \#(H \times K) = \#H \cdot \#K = \#G.$$

But since $HK \subseteq G$, this implies that $G = HK$. Finally, since G is abelian we have $H \trianglelefteq G$ and $K \trianglelefteq G$, which implies that $G = H \times K$ is a direct product.

- (b) Now suppose that $\#G = p_1^{e_1} p_2^{e_2} \cdots p_n^{e_n}$ for distinct primes p_1, p_2, \dots, p_n . By Sylow there exists a unique subgroup $H_i \subseteq G$ of size $\#H_i = p_i^{e_i}$ for each i . Now consider the multiplication map from the external direct product:

$$\mu : H_1 \times H_2 \times \cdots \times H_n \longrightarrow G$$

Since G is abelian we know that μ is a group homomorphism. We will use induction on n to prove that μ is actually a **group isomorphism**.

The base case $n = 2$ follows from part (a). Now let $n \geq 2$ and assume for induction that the result holds for n . We will prove that the result still holds for $n+1$. So let $\#G = p_1^{e_1} \cdots p_n^{e_n} p_{n+1}^{e_{n+1}}$ for some distinct primes p_1, \dots, p_n, p_{n+1} and consider the multiplication homomorphism from the external direct product of Sylow subgroups:

$$\mu : H_1 \times \cdots \times H_n \times H_{n+1} \longrightarrow G.$$

To show that μ is bijective, let $G' := H_1 \cdots H_n \subseteq G$ be the image of the first n factors, so we obtain a **surjective** group homomorphism:

$$\mu' : H_1 \times \cdots \times H_n \longrightarrow G' \subseteq G.$$

Now the First Isomorphism Theorem and Lagrange tell us that

$$\#G' = \#\text{im } \mu' = \frac{\#(H_1 \times \cdots \times H_n)}{\#\ker \mu'} = \frac{p_1^{e_1} \cdots p_n^{e_n}}{\#\ker \mu'} = p_1^{d_1} \cdots p_n^{d_n}$$

for some exponents $0 \leq d_i \leq e_i$. But since $H_i \subseteq G'$ is a subgroup for each $i \in \{1, \dots, n\}$, Lagrange tells us that $d_i = e_i$ and hence $\#G' = p_1^{e_1} \cdots p_n^{e_n}$.

We have shown that H_1, \dots, H_n are the Sylow subgroups of G' , so the induction hypothesis tells us that μ' is a bijection. Finally, we conclude that μ is a **composition of two bijections**, where the second follows from part (a):

$$\mu : (H_1 \times \cdots \times H_n) \times H_{n+1} \xrightarrow{\mu' \times \text{id}} G' \times H_{n+1} \longrightarrow G.$$

Hence μ is a bijection. □

[Remark: This problem was tricky because I never told you the definition for the *internal direct product of several subgroups* $H_1, \dots, H_n \subseteq G$. There are different ways to define this. The easiest way is to require that the multiplication map

$$\mu : H_1 \times \cdots \times H_n \longrightarrow G$$

is a group isomorphism.]

(c) This part is much easier. If G is cyclic then we know from the Fundamental Theorem of Cyclic Groups that every subgroup of G is cyclic. Hence each primary subgroup is cyclic. □

3. Lagrange vs. Rank-Nullity. Let $p \in \mathbb{Z}$ be prime. You showed on the previous homework that every nonzero element of the ring $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$ has a multiplicative inverse. In other words, \mathbb{F}_p is a field of size p .

- (a) Let V be an n -dimensional vector space over \mathbb{F}_p . Prove that $\#V = p^n$.
- (b) Now let $U \subseteq V$ be a k -dimensional subspace. Show that Lagrange's Theorem and the Rank-Nullity Theorem give you the same information about this subspace.

(a) Choose a basis $\mathcal{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subseteq V$ and consider the function $\varphi_{\mathcal{U}} : \mathbb{F}_p^n \rightarrow V$ defined by

$$\varphi_{\mathcal{U}}(a_1, a_2, \dots, a_n) := a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n.$$

This function is surjective because \mathcal{U} is a spanning set and injective because \mathcal{U} is an independent set. We conclude that $\varphi_{\mathcal{U}}$ is a bijection and hence

$$\#V = \#\mathbb{F}_p^n = (\#\mathbb{F}_p)^n = p^n.$$

[Remark: In fact, the function $\varphi_{\mathcal{U}}$ also preserves addition and scalar multiplication. Hence it defines an isomorphism of vector spaces $V \cong \mathbb{F}_p^n$.]

(b) Now let $U \subseteq V$ be a k -dimensional subspace. We will compute the cardinality of the quotient space V/U in two ways.

- (1) Note that V/U is (in particular) a quotient of abelian groups. Then from part (a) and Lagrange's Theorem we have

$$\#(V/U) = \#V/\#U = p^n/p^k = p^{n-k}.$$

- (2) We know from the previous homework that $\dim(V/U) = \dim(V) - \dim(U) = n - k$. (This is what I mean by the Rank-Nullity Theorem.) Then from part (a) we have

$$\#(V/U) = p^{\dim(V/U)} = p^{n-k}.$$

Note that we get the same answer.

4. Double Cosets. Let G be a group and let $H, K \subseteq G$ be any subgroups. For each pair $(h, k) \in H \times K$ consider the function $\varphi_{(h,k)}(g) := h g k^{-1}$.

- (a) Prove that this defines a group homomorphism $\varphi : H \times K \rightarrow \text{Perm}(G)$.
 (b) For each $g \in G$, prove that the orbit satisfies

$$\text{Orb}_\varphi(g) = HgK := \{h g k : h \in H, k \in K\}.$$

These orbits are called *double cosets*. Unlike single cosets, we will see that double cosets do not all have the same size.

- (c) We also have a group action $\psi : H \rightarrow \text{Perm}(G/K)$ defined by $\psi_h(gK) := (hg)K$. (Don't bother to prove this.) For all $g \in G$ prove that HgK is the disjoint union of the cosets in the ψ -orbit of gK :

$$HgK = \coprod_{C \in \text{Orb}_\psi(gK)} C.$$

- (d) For all $g \in G$ prove that $\text{Stab}_\psi(gK) = H \cap gKg^{-1}$, where $gKg^{-1} := \{gkg^{-1} : k \in K\}$.
 (e) Combine (c) and (d) with Lagrange's Theorem and Orbit-Stabilizer to conclude that

$$\#HgK = \frac{\#H \cdot \#K}{\#(H \cap gKg^{-1})}.$$

(a) Let $H, K \subseteq G$ be subgroups, and for each pair $(h, k) \in H \times K$ consider the function $\varphi_{(h,k)} : G \rightarrow G$ defined by $\varphi_{(h,k)}(g) := h g k^{-1}$. I claim that this defines a group homomorphism

$$\varphi : H \times K \rightarrow \text{Perm}(G).$$

There are two things to check:

- For all $(h, k) \in H \times K$ I claim that $\varphi_{h,k} : G \rightarrow G$ is invertible with inverse $\varphi_{(h^{-1}, k^{-1})}$. Indeed, for all $g \in G$ we have

$$(\varphi_{(h,k)} \circ \varphi_{(h^{-1}, k^{-1})})(g) = h(h^{-1}g(k^{-1})^{-1})k^{-1} = (hh^{-1})g(kk^{-1}) = g$$

and

$$(\varphi_{(h^{-1}, k^{-1})} \circ \varphi_{(h,k)})(g) = h^{-1}(hgk^{-1})(k^{-1})^{-1} = (h^{-1}h)g(k^{-1}k) = g.$$

Hence $\varphi_{(h,k)} \in \text{Perm}(G)$.

- For all $(h_1, k_1), (h_2, k_2) \in H \times K$ and $g \in G$ we have

$$(\varphi_{(h_1, k_1)} \circ \varphi_{(h_2, k_2)})(g) = h_1(h_2gk_2^{-1})k_1^{-1} = (h_1h_2)g(k_1k_2)^{-1} = \varphi_{(h_1h_2, k_1k_2)}(g).$$

It follows that φ is a homomorphism from the external direct product $H \times K$ to G .
[Remark: This is why we need to apply k^{-1} on the right instead of k .]

(b) I just want you to observe that $\text{Orb}_\varphi(g) = HgK$. There's not really anything to check. The point is that these double cosets partition the group G .

(c) Assume that $\psi_h(gK) := (hg)K$ defines a group homomorphism $\psi : H \rightarrow \text{Perm}(G/K)$. [Remark: We do not assume that $K \subseteq G$ is normal, so G/K is only a set.] Now fix $g \in G$ and consider the orbit $\text{Orb}_\psi(gK) \subseteq G/K$. By definition this is a set of left cosets of K . I claim that we have a disjoint union:

$$HgK = \coprod_{C \in \text{Orb}_\psi(gK)} C.$$

There are three things to check:

- Since the left cosets of K partition G , we know that any two non-equal cosets are disjoint. It follows that

$$\bigcup_{C \in \text{Orb}_\psi(gK)} C = \coprod_{C \in \text{Orb}_\psi(gK)} C.$$

- Any $C \in \text{Orb}_\psi(gK)$ has the form $C = (hg)K$ for some $h \in H$. Since $(hg)K \subseteq HgK$ we conclude that $C \subseteq HgK$ and hence

$$\bigcup_{C \in \text{Orb}_\psi(gK)} C \subseteq HgK.$$

- Any element $hgz \in HgK$ is contained in the coset $C := (hg)K \subseteq \text{Orb}_\psi(gK)$, hence

$$HgK \subseteq \bigcup_{C \in \text{Orb}_\psi(gK)} C.$$

(d) For any $g \in G$ I claim that $\text{Stab}_\psi(gK) = H \cap gKg^{-1}$. Indeed, suppose that $gkg^{-1} \in H$ for some $k \in K$. Then we have

$$\psi_{gkg^{-1}}(gK) = (gkg^{-1}g)K = (gk)K = gK$$

and hence $gkg^{-1} \in \text{Stab}_\psi(gK)$. Conversely, if $h \in \text{Stab}_\psi(gK)$ then we must have

$$(hg)K = gK \implies g^t hg \in K \implies h \in gKg^{-1}.$$

(e) Finally, suppose that H and K are finite. Since all left cosets of K have size $\#K$ we conclude from part (c) that

$$\#HgK = \sum_{C \in \text{Orb}_\psi(gK)} \#C = \sum_{C \in \text{Orb}_\psi(gK)} \#K = \#\text{Orb}_\psi(gK) \cdot \#K.$$

Then from part (d), Orbit-Stabilizer and Lagrange we have

$$\#HgK = \#\text{Orb}_\psi(gK) \cdot \#K = \frac{\#H}{\#\text{Stab}_\psi(gK)} \cdot \#K = \frac{\#H \cdot \#K}{\#(H \cap gKg^{-1})}.$$

□

[Remark: In the special case $g = \varepsilon$ we obtain the formula

$$\#HK = \frac{\#H \cdot \#K}{\#(H \cap K)}.$$

We already proved this using the Second Isomorphism Theorem when one of H or K is normal. But now we have a proof that works for any H and K . That's nice.]

5. Burnside's Lemma. Let $\varphi : G \rightarrow \text{Perm}(X)$ be a group action, and let X/G denote the set of orbits. For each $g \in G$, let $\text{Fix}_\varphi(g)$ denote the set of elements fixed by g :

$$\text{Fix}_\varphi(g) := \{x \in X : \varphi_g(x) = x\} \subseteq X.$$

(a) Count the elements of the set $\{(g, x) \in G \times X : \varphi_g(x) = x\}$ in two ways to prove that

$$\sum_{g \in G} \#\text{Fix}_\varphi(g) = \sum_{x \in X} \#\text{Stab}_\varphi(x).$$

(b) Use Orbit-Stabilizer to obtain a formula for the number of orbits:

$$\#(X/G) = \frac{1}{\#G} \sum_{g \in G} \#\text{Fix}_\varphi(g).$$

- (c) Application: Consider a “bracelet” (circular string of beads) containing 6 beads. There are k possible colors for the beads, and we regard two bracelets to be the same if they are equivalent up to dihedral symmetry. Use the formula in part (b) to compute the number of different bracelets. [Hint: The dihedral group D_{12} acts on a set X of size k^6 . You want to compute the number of orbits: $\#(X/D_{12})$. To get started I’ll tell you that $\#\text{Fix}(R) = k$ and $\#\text{Fix}(R^2) = k^2$.]

(a) **Double Counting.** Suppose that G and X are both finite. We will count the elements of the set $S = \{(g, x) \in G \times X : \varphi_g(x) = x\}$ in two different ways:

- (1) For each $g \in G$ there are $\#\text{Fix}_\varphi(g)$ elements $x \in X$ such that $\varphi_g(x) = x$. Hence

$$\#S = \sum_{g \in G} \#\text{Fix}_\varphi(g).$$

- (2) For each $x \in X$ there are $\#\text{Stab}_\varphi(x)$ elements $g \in X$ such that $\varphi_g(x) = x$. Hence

$$\#S = \sum_{x \in X} \#\text{Stab}_\varphi(x).$$

Alternatively: Think of S as a matrix with rows indexed by elements of G and columns indexed by elements of X . We put 1 in the (g, x) entry of S if $\varphi_g(x) = x$ and put 0 otherwise. Then (1) sums the elements of S row-by-row and (2) sums the elements of S column-by-column.

(b) Let $X/G = \{\mathcal{O}_1, \dots, \mathcal{O}_n\}$ be the set of orbits, so that $\#(X/G) = n$. Furthermore, note that for all $x \in \mathcal{O}_i$ we have $\text{Orb}_\varphi(x) = \mathcal{O}_i$ and hence $\#\text{Orb}_\varphi(x) = \#\mathcal{O}_i$. Now applying Orbit-Stabilizer to part (a) gives



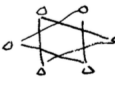

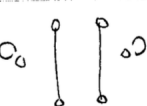

$$\begin{aligned} \sum_{g \in G} \#\text{Fix}_\varphi(g) &= \sum_{x \in X} \#\text{Stab}_\varphi(x) \\ &= \sum_{x \in X} (\#G / \#\text{Orb}_\varphi(x)) \\ &= \#G \cdot \sum_{x \in X} (1 / \#\text{Orb}_\varphi(x)) \\ &= \#G \cdot \sum_{i=1}^n \left(\sum_{x \in \mathcal{O}_i} 1 / \#\text{Orb}_\varphi(x) \right) \\ &= \#G \cdot \sum_{i=1}^n \left(\sum_{x \in \mathcal{O}_i} 1 / \#\mathcal{O}_i \right) \\ &= \#G \cdot \sum_{i=1}^n (1) \end{aligned}$$

$$\begin{aligned}
&= \#G \cdot n \\
&= \#G \cdot \#(X/G).
\end{aligned}$$

□

(c) **Application: Bracelets.** Consider the vertices of a fixed hexagon. Suppose we have k colors and let X be the set of all colorings of the vertices, so that $\#X = k^6$. Now consider the natural action $D_{12} \curvearrowright X$ of the dihedral group by rotation and reflection. We will use Burnside's Lemma to compute the number of orbits.

Since each element $g \in D_{12}$ permutes the vertices of the hexagon we can think of it as an element of S_6 . Furthermore, note that a given coloring $x \in X$ is fixed by g if and only if the vertices in each cycle have the same color. In other words, we have $\#\text{Fix}(g) = k^{\#(\text{cycles})}$. Here is a table showing all the possibilities:

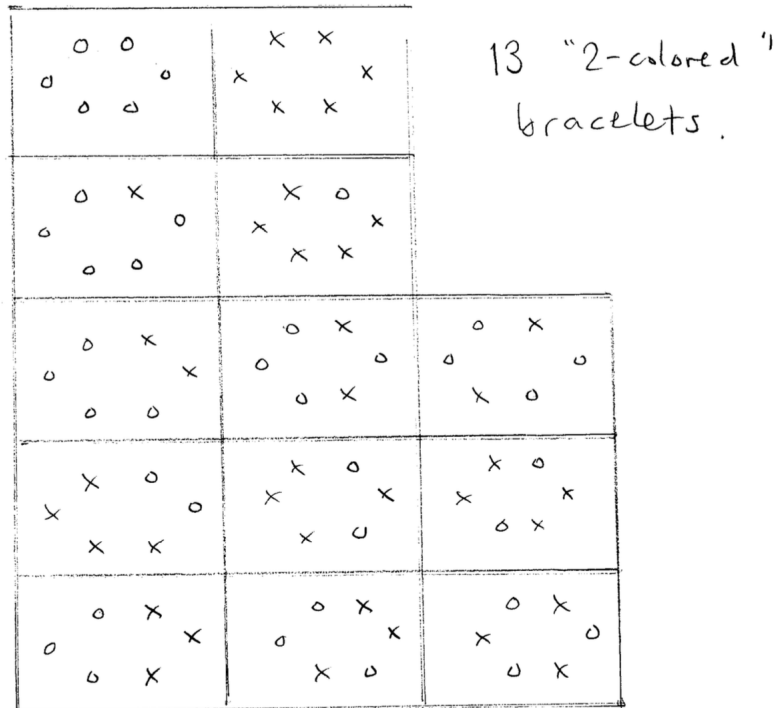
$g \in D_{12}$	cycles	$\#\text{Fix}(g)$
I		k^6
R, R^5		k^1
R^2, R^4		k^2
R^3		k^3
F, R^2F, R^4F		k^4
RF, R^3F, R^5F		k^3

By summing over all 12 group elements, Burnside's Lemma tells us that

$$\begin{aligned}
\#(\text{bracelets}) &= \#(X/D_{12}) \\
&= \frac{1}{\#D_{12}} \sum_{g \in D_{12}} \#\text{Fix}(g)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{12} \sum_{g \in D_{12}} k^{\#(\text{cycles})} \\
&= \frac{1}{12} (1k^6 + 2k + 2k^2 + 1k^3 + 3k^4 + 3k^3) \\
&= \frac{1}{12} k(k+1)(k^4 - k^3 + 4k^2 + 2).
\end{aligned}$$

For example, when $k = 2$ we get $\#(X/D_{12}) = 13$ different bracelets. Here they are:



6. Normal Subgroups of S_n . Assuming that A_n is simple (which is true for $n \geq 5$) you will prove that A_n is the only non-trivial normal subgroup of S_n .

- (a) For $n \geq 3$, prove that the center of S_n is trivial: $Z(S_n) = \{\text{id}\}$. [Hint: For any $\text{id} \neq g \in S_n$, prove that there exists some $f \in S_n$ such that $fgf^{-1} \neq g$.]
- (b) Suppose that $N \trianglelefteq S_n$ is a normal subgroup not equal to $\{\text{id}\}$ or S_n . Use the fact that A_n is simple to prove that $N = A_n$ or $\#N = 2$. [Hint: Consider $N \cap A_n \trianglelefteq A_n$.]
- (c) Continuing from (b), if $\#N = 2$ then we must have $N = \{\text{id}, \tau\}$ for some $\tau \in S_n$ such that $\tau \neq \text{id}$ and $\tau^2 = \text{id}$. Prove that $\tau \in Z(S_n)$ and get a contradiction.

(a) Let $n \geq 3$ and consider any non-identity permutation $\text{id} \neq g \in S_n$. We will prove that there exists a permutation $f \in S_n$ such that $fgf^{-1} \neq g$, and hence $g \notin Z(S_n)$. To do this we

will use the fact that fgf^{-1} has the same cycle structure as g , where the symbols $1, 2, \dots, n$ have been replaced by the permuted symbols f_1, f_2, \dots, f_n , respectively. There are two cases:

- Suppose g has at least two cycles, one of which is not a singleton. Choose any symbols i, j from these two cycles and define the transposition $f = (ij)$. Then $fgf^{-1} \neq g$.
- Otherwise, since $g \neq \text{id}$ we know that g is an n -cycle. In this case let $i := g_1$ and define the transposition $f = (1i)$. Then since $n \geq 3$ we have $fgf^{-1} \neq g$.

We conclude that $Z(S_n) = \{\text{id}\}$.

(b) Suppose that $N \trianglelefteq S_n$ is a normal subgroup not equal to $\{\text{id}\}$ or S_n . Assuming that A_n is a simple group, we will prove that $N = A_n$ or $\#N = 2$. To see this, define $N' := N \cap A_n$. Then since $N' \trianglelefteq A_n$ is normal and since A_n is simple we have two cases:

- (1) $N' = A_n$,
- (2) $N' = \{\text{id}\}$.

In case (1) we have $N \cap A_n = A_n$, which implies $A_n \subseteq N$. Now consider the subgroup $N/A_n \subsetneq S_n/A_n$. Since $S_n/A_n \cong \mathbb{Z}/2\mathbb{Z}$ has no non-trivial subgroups we must have $\#(N/A_n) = 1$, which implies that $\#N = \#A_n$ and hence $N = A_n$. In case (2) we have $N \cap A_n = \{\text{id}\}$, which implies that

$$\#NA_n = \frac{\#N \cdot \#A_n}{\#(N \cap A_n)} = \#N \cdot \#A_n.$$

But since at least one of N and A_n is normal (in fact, they are both normal) we know that $NA_n \subseteq S_n$ is a subgroup satisfying

$$A_n \subsetneq NA_n \subseteq S_n.$$

Finally, the same reasoning as in case (1) shows that $NA_n = S_n$, and hence

$$\begin{aligned} \#NA_n &= \#S_n \\ \#N \cdot \#A_n &= \#S_n \\ \#N &= 2. \end{aligned}$$

(c) It only remains to show that $N \trianglelefteq S_n$ and $\#N = 2$ lead to a contradiction. So assume that $N = \{\text{id}, \tau\}$ with $\tau \neq \text{id}$. If $N \trianglelefteq S_n$ then for any $f \in S_n$ we have $f\tau f^{-1} \in N$. But then the fact that $\tau \neq \text{id}$ implies that $f\tau f^{-1} \neq \text{id}$, and hence $f\tau f^{-1} = \tau$. We conclude that $\tau \in Z(S_n)$, which contradicts (a). (Of course, we assume that $n \geq 3$.) \square

[Remark: We have shown that **if** A_n is simple and **if** $n \geq 3$, then A_n is the **only** non-trivial normal subgroup of S_n . It is easy to check that A_3 is simple, and you showed above in Problem 1 that A_4 is **not** simple. It turns out to be true that A_n is simple for all $n \geq 5$, but, again, I don't want to prove that here. Look up a proof if you want.]

7. Gaussian Binomial Coefficients. Let p be prime and consider the field $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$.

(a) For all $n \geq 0$ we define the p -factorial:

$$[n]_p! := \prod_{i=1}^n \frac{p^i - 1}{p - 1} = \prod_{i=1}^n (1 + p + p^2 + \cdots + p^{i-1}) \in \mathbb{Z}.$$

Prove that $\#GL_n(\mathbb{F}_p) = p^{\binom{n}{2}} \cdot (p-1)^n \cdot [n]_p!$. [Hint: The columns of an invertible matrix are just an ordered basis for the vector space \mathbb{F}_p^n . Argue that there are $p^n - 1$ ways to choose the first basis vector, then $p^n - p$ ways to choose the second basis vector, etc., so that $\#GL_n(\mathbb{F}_p) = \prod_{i=0}^{n-1} (p^n - p^i)$.]

(b) Let X be the set of all k -dimensional subspaces of \mathbb{F}_p^n . The group $GL_n(\mathbb{F}_p)$ acts on X in the obvious way. For any k -dimensional subspace $U \in X$, prove that the stabilizer of U is isomorphic to the following subgroup of $GL_n(\mathbb{F}_p)$:

$$\left\{ \left(\begin{array}{c|c} A & C \\ \hline 0 & B \end{array} \right) : A \in GL_k(\mathbb{F}_p), B \in GL_{n-k}(\mathbb{F}_p), C \in \text{Mat}_{k \times (n-k)}(\mathbb{F}_p) \right\}.$$

[Hint: Choose a basis $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ for \mathbb{F}_p^n such that $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ is a basis for U .]

(c) Combine parts (a) and (b) with the Orbit-Stabilizer Theorem to prove that

$$\#X = \frac{[n]_p!}{[k]_p! \cdot [n-k]_p!}.$$

This is called a *Gaussian binomial coefficient*.

(a) The hint pretty much does it. By looking at the columns we have a bijection between elements of $GL_n(\mathbb{F}_p)$ and ordered bases for \mathbb{F}_p^n . From Problem 3(a) there are $p^n - 1$ ways to choose the first (nonzero) basis vector \mathbf{u}_1 . Then the next basis vector \mathbf{u}_2 must lie outside the line $\mathbb{F}_p(\mathbf{u}_1)$. Since the line has size p , there are $p^n - p$ ways to choose \mathbf{u}_2 . Then the third basis vector \mathbf{u}_3 must lie outside the plane $\mathbb{F}_p(\mathbf{u}_1, \mathbf{u}_2)$. Since the plane has size p^2 , there are $p^n - p^2$ ways to choose \mathbf{u}_3 . Continuing in this way we have

$$\begin{aligned} \#GL_n(\mathbb{F}_p) &= (p^n - 1)(p^n - p)(p^n - p^2) \cdots (p^n - p^{n-1}) \\ &= p^{\binom{n}{2}} \cdot (p^n - 1)(p^{n-1} - 1)(p^{n-2} - 1) \cdots (p - 1) \\ &= p^{\binom{n}{2}} \cdot (p - 1)^n \cdot \frac{(p^n - 1)}{(p - 1)} \frac{(p^{n-1} - 1)}{(p - 1)} \frac{(p^{n-2} - 1)}{(p - 1)} \cdots \frac{(p - 1)}{(p - 1)} \\ &= p^{\binom{n}{2}} \cdot (p - 1)^n \cdot [n]_p!. \end{aligned}$$

(b) Now let X be the set of all k -dimensional subspaces of \mathbb{F}_p^n . For any matrix $M \in GL_n(\mathbb{F}_p)$ and subspace $U \in X$ we define

$$MU := \{M\mathbf{x} : \mathbf{x} \in U\}.$$

It is easy to check that $MU \subseteq \mathbb{F}_p^n$ is a subspace and if $\mathbf{u}_1, \dots, \mathbf{u}_k$ is a basis for U then $M\mathbf{u}_1, M\mathbf{u}_2, \dots, M\mathbf{u}_k$ is a basis for MU , hence $MU \in X$. Furthermore, let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ be a basis for some arbitrary subspace $V \in X$ and extend the vectors \mathbf{u}_i and \mathbf{v}_i to bases for \mathbb{F}_p^n . Then the invertible linear function defined by $\mathbf{u}_i \mapsto \mathbf{v}_i$ is represented by some matrix $M \in GL_n(\mathbb{F}_p)$, so that $MU = V$. We conclude that $GL_n(\mathbb{F}_p)$ acts transitively on the set X .

It only remains to compute the stabilizer. So fix some $U \in X$ with basis $\mathbf{u}_1, \dots, \mathbf{u}_k$ and extend this to a basis $\mathbf{u}_1, \dots, \mathbf{u}_k, \dots, \mathbf{u}_n$ for \mathbb{F}_p^n . Let $N \in GL_n(\mathbb{F}_p)$ be the matrix with i -th column equal to \mathbf{u}_i . Then for any $M \in \text{Stab}(U)$ we have $M\mathbf{x} \in U$ for all $\mathbf{x} \in U$ and it follows that

$$N^{-1}MN = \left(\begin{array}{c|c} A & C \\ \hline 0 & B \end{array} \right) \in GL_n(\mathbb{F}_p)$$

for some matrices $A \in GL_k(\mathbb{F}_p), B \in GL_{n-k}(\mathbb{F}_p), C \in \text{Mat}_{k \times (n-k)}(\mathbb{F}_p)$. Conversely, for any matrices A, B, C of this form one can check that

$$N \left(\begin{array}{c|c} A & C \\ \hline 0 & B \end{array} \right) N^{-1} \in \text{Stab}(U).$$

Since conjugation by N is an automorphism of $GL_n(\mathbb{F}_p)$, it follows that $\text{Stab}(U)$ is isomorphic to the group of all such matrices. In particular, we obtain a bijection:

$$\text{Stab}(U) \longleftrightarrow GL_k(\mathbb{F}_p) \times GL_{n-k}(\mathbb{F}_p) \times \text{Mat}_{k \times (n-k)}(\mathbb{F}_p).$$

[Remark: In fact, I claim that the stabilizer is a semidirect product:

$$\text{Stab}(U) \cong (\text{Mat}_{k \times (n-k)}(\mathbb{F}_p), +, 0) \rtimes [GL_k(\mathbb{F}_p) \times GL_{n-k}(\mathbb{F}_p)].$$

What is the action?]

(c) Finally, combining parts (a) and (b) with the Orbit-Stabilizer Theorem gives

$$\begin{aligned} \#X &= \#\text{Orb}(U) \\ &= \frac{\#GL_n(\mathbb{F}_p)}{\#\text{Stab}(U)} \\ &= \frac{\#GL_n(\mathbb{F}_p)}{\#(GL_k(\mathbb{F}_p) \times GL_{n-k}(\mathbb{F}_p) \times \text{Mat}_{k \times (n-k)}(\mathbb{F}_p))} \\ &= \frac{\#GL_n(\mathbb{F}_p)}{\#GL_k(\mathbb{F}_p) \cdot \#GL_{n-k}(\mathbb{F}_p) \cdot \#\text{Mat}_{k \times (n-k)}(\mathbb{F}_p)} \\ &= \frac{p^{\binom{n}{2}} \cdot \cancel{(p-1)^n} \cdot [n]_p!}{p^{\binom{k}{2}} \cdot \cancel{(p-1)^k} \cdot [k]_p! \cdot p^{\binom{n-k}{2}} \cdot \cancel{(p-1)^{n-k}} \cdot [n-k]_p! \cdot \cancel{p^{k(n-k)}}} \\ &= \frac{[n]_p!}{[k]_p! \cdot [n-k]_p!}. \end{aligned}$$

□

[Remark: If we treat p as a formal variable then one can check that

$$\frac{[n]_p!}{[k]_p! \cdot [n-k]_p!} \longrightarrow \frac{n!}{k!(n-k)!} \quad \text{as} \quad p \longrightarrow 1.$$

This suggests that a “ k -subset of an n -subset” is somehow the same thing as a “ k -dimensional subspace of an n -dimensional vector space over the field $\mathbb{Z}/1\mathbb{Z}$.” Unfortunately that makes no sense because $\mathbb{Z}/1\mathbb{Z} \cong \{0\}$, so every vector space over $\mathbb{Z}/1\mathbb{Z}$ has size 1. Strange.]