Algebra: 1830–1930

Drew Armstrong

Version: September, 2023

For Moira and Miles

Contents

Preface

This textbook intended for a two-semester sequence in abstract algebra. The first semester covers group theory and the second semester covers field theory. The book is suitable for undergraduate students with some prior experience of integers, polynomials, vectors and matrices. It would be helpful to have some experience with abstract vector spaces, but we do not assume this. Necessary results of linear algebra will be developed in the Exercises.

Complex numbers are a trickier issue. In my experience, American students are not exposed early enough to complex numbers. In more elementary courses I am happy to remediate this. However, the pace of this course won't allow it. If students are not familiar with Euler's formula

$$e^{i\theta} = \cos\theta + i\sin\theta$$

then they will have to learn it on the fly.

The distinguishing feature of this textbook is that we **take history seriously**. This shows up in two ways. First, we use the historical development of Galois theory as a framework to organize the choice of topics. Second, we attempt to track down original sources for all of the main ideas. The dates of the title refer to Galois' *Memoir on the conditions for solvability of equations by radicals* (1830) and van der Waerden's textbook on *Modern algebra* (1930). However, we will also reach as far back as Lagrange's *Reflections on the algebraic resolution of equations* (1770) and as far forward as Artin's Notre Dame lectures on Galois theory (1942). It just so happens that almost all of the concepts that we regard as "undergraduate abstract algebra" developed during this time period.

Though our choice of topics is based on the development of algebra before 1930, we freely employ more recent notation in the proofs and in the logical arrangement of the results:

- We use standard set-theoretic terminology. For example, we use Bourbaki's notation for injective, surjective and bijective functions.
- We use arrows to denote functions. According to Mac Lane (1971, pg 29), the arrow notation was developed around 1940 in the papers of Hurewicz.

- We use modern notations for linear algebra. The history of linear algebra notation is difficult to track down, but the most significant source is surely Hermann Weyl's work on quantum mechanics.
- We use posets and lattices to organize collections of subgroups, etc. These concepts were introduced by Dedekind in the 1800s, but were only standardized in the 1940s by Birkhoff and Ore.
- In particular, we use Ore's (1944) concept of abstract Galois connections to organize the various "correspondence theorems" for subgroups, etc. This is the most peculiar feature of our book, since the concept is not completely standard. Apart from making some proofs more elegant, we find this to be good preparation for the modern version of Galois' theorem, which is due to Artin (1942).
- We make sparing use of commutative diagrams to explain the "universal properties" of polynomials and fields of fractions. This is mainly to clear up the ontological status of splitting fields.

First Semester: Group Theory

Week 1:

This first section is mainly for motivation. The ideas discussed here are quite tricky and we will not completely understand them until the end of the second semester. We will begin the logical development of the subject in Week 2.

1.1 What is Algebra?

The subject of algebra has changed over time. Here is a historical sketch:

$$\text{pre-modern algebra} \;\Big|_{1830}\; \text{transition period} \;\Big|_{1930}\; \text{modern algebra}$$

Prior to 1830 the word "algebra" meant the study of equations. After 1930 the word "algebra" refers to the study of "abstract structure" in mathematics. In this course I will tell the story of the transition between these two eras.

To begin, here is a quick review of "pre-modern algebra".

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

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

I assume that you 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

$$\begin{aligned} x^2 + \frac{b}{a}x + \frac{c}{a} &= 0 \\ x^2 + \frac{b}{a}x &= -\frac{c}{a}. \end{aligned}$$

Now there is 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:

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

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

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

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

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:

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

This algorithm was known to several ancient civilizations. It came to modern Europe via al-Khwarizmi's (∼820) *Compendious Book on Calculation by Completion and Balancing*. The terms "completion" [Arabic: al-jabr] and "balancing" [Arabic: al-muqabala] in the title refer to the operations of adding or subtracting the same quantity from each side of an equation. The historical significance of this work is illustrated by the fact that our word "algorithm" is a corruption of al-Khwarizmi's name, and our word "algebra" comes from the word "al-jabr" in the title of the work.

Al-Khwarizmi expressed the problem of quadratic equations in terms of geometry, and the solution is achieved by literally completing a square. Actually, since negative numbers have no geometric meaning, al-Khwarizmi treats three separate cases and only the easiest case can be solved by completing a square.

The modern symbolic formula has only one case because we accept negative numbers and square roots of negative numbers.

Now we come to a result 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: Cardano's Formula. Let a, b, c, d be any numbers with $a \neq 0$. Our goal is to find all numbers x such that

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

First we divide both sides by a and then we substitute $x \mapsto b/(3a)$ to obtain the so-called "depressed form" of the equation:

$$x^3 + 3px + 2q = 0,$$

where[1]

$$p = \frac{3ac - b^2}{9a^2} \quad \text{and} \quad q = \frac{27a^2d - 9abc + 2b^3}{54a^3}.$$

Finally, "Cardano's Formula" tells us that

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

Unfortunately, as with the quadratic formula, this is not really a "formula" because every nonzero complex number actually has **three** cube roots. This means that there might be **nine** different ways to choose the cube roots, while the original equation has only **three** solutions. Below we will discuss Lagrange's solution to this puzzle. ///

The difficulty of interpreting Cardano's formula was the historical motivation for complex numbers. To see why, let us consider the equation $x^3 - 15x - 4 = 0$. This equation has a real root $x = 4$, but Cardano's formula tells us that

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

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 $2 - \sqrt{-121}$. [**Exercise:** Verify this by cubing the expressions $2 \pm \sqrt{-1}$.] That is, the only way to get to the real solution $x = 4$ is by going through the so-called "imaginary numbers". In modern notation we write

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

[1] These complicated expressions are one of the reasons why the cubic equation is not studied in high school.

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

The general cubic formula was first published by Gerolamo Cardano in the *Ars Magna* (1545), although he did not discover the formula himself.[2] This work also includes a general solution to the quartic equation which was discovered by Cardano's student Lodovico Ferrari. After this intense burst of activity, progress stalled on the following question.

Question: Does there exist a Quintic Formula? 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:

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

In other words: Is the general quintic equation "solvable by radicals" ?

The next breakthrough occurred in the early 1700s when Abraham de Moivre discovered a special class of solvable quintics.

Example: De Moivre's Quintic Equation. I assume that you are familiar with *de Moivre's formula*:

$$(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta).$$

This can be proved for integer exponents by induction. Furthermore, Euler (1748) showed that this identity extends to real exponents by applying the power series expansions of the exponential and trigonometric functions:

$$\begin{aligned} e^{it} &:= 1 + (it) + \frac{1}{2!}(it)^2 + \frac{1}{3!}(it)^3 + \cdots \\ &= \left(1 - \frac{1}{2!}t^2 + \frac{1}{4!}t^4 - \cdots\right) + i\left(t - \frac{1}{3!}t^3 + \frac{1}{5!}t^5 - \cdots\right) \\ &= \cos t + i\sin t. \end{aligned}$$

Today we rightly view this as the fundamental theorem of trigonometry, but de Moivre (1707) originally viewed it as merely as a clever trick used to obtain radical solutions for a certain family of polynomials. I will show you how the works in the case of quintic equations. On the one hand, we can expand the binomial $(\cos\theta + i\sin\theta)^5$ and compare real parts to obtain

$$\cos(5\theta) = \cos^5\theta - 10\cos^3\theta\sin^2\theta + 5\cos\theta\sin^4\theta$$

[2]The full story is very colorful and has ween well told elsewhere.

$$\begin{aligned}&= \cos^5\theta - 10\cos^3\theta(1-\cos^2\theta) + 5\cos\theta(1-\cos^2\theta)^2\\ &= 16\cos^5\theta - 20\cos^3\theta + 5\cos\theta,\end{aligned}$$

and hence

$$\begin{aligned}2\cos(5\theta) &= 32\cos^5\theta - 40\cos^3\theta + 10\cos\theta\\ &= (2\cos\theta)^5 - 5(2\cos\theta)^3 + 5(2\cos\theta).\end{aligned}$$

On the other hand, if we temporarily ignore the subtleties involving roots of complex numbers, then we can write

$$\begin{aligned}\cos\theta + i\sin\theta &= \sqrt[5]{\cos(5\theta) + i\sin(5\theta)}\\ &= \sqrt[5]{\cos(5\theta) + i\sqrt{1-\cos^2(5\theta)}}\\ &= \sqrt[5]{\cos(5\theta) + \sqrt{\cos^2(5\theta)-1}},\end{aligned}$$

and hence

$$\begin{aligned}2\cos\theta &= (\cos\theta + i\sin\theta) + (\cos\theta - i\sin\theta)\\ &= (\cos\theta + i\sin\theta) + 1/(\cos\theta + i\sin\theta)\\ &= \sqrt[5]{\cos(5\theta) + \sqrt{\cos^2(5\theta)-1}} + 1/\sqrt[5]{\cos(5\theta) + \sqrt{\cos^2(5\theta)-1}}.\end{aligned}$$

Finally, by setting $x = 2\cos\theta$ and $a = \cos(5\theta)$, we observe that the equation

$$x^5 - 5x^3 + 5x - 2a = 0$$

has a root of the form

$$x = \sqrt[5]{a + \sqrt{a^2-1}} + \frac{1}{\sqrt[5]{a + \sqrt{a^2-1}}}.$$

However, as with Cardano's formula, it is tricky to interpret this formula. You will investigate the details in Exercise 1.C. ///

Despite progress with various special polynomials, by the late 1700s it began to seem that a general polynomial of degree greater than four is **not** solvable by radicals. In (1770) Joseph-Louis Lagrange summarized the state of knowledge on this problem. Lagrange's work was deep and technical, and he despaired that the subject had perhaps become too difficult to merit further investigation:

> *I begin to notice how my inner resistance increases little by little, and I cannot say whether I will still be doing geometry ten years from now. It also seems to me that the mine has maybe already become too deep and unless one finds new veins it might have to be abandoned.*
>
> *Physics and chemistry now offer a much more glowing richness and much easier exploitation. Also, the general taste has turned entirely*

> *in this direction, and it is not impossible that the place of Geometry in the Academies will someday become what the role of the Chairs of Arabic at the universities is now.*[3]

Nevertheless, mathematicians persisted in the same vein for a few more decades. The next breakthrough was made by the young Carl Friedrich Gauss in his first major work, the *Disquisitiones Arithmeticae* (1798), written when he was just 21. In the final Chapter 7 of this work he sketched out a complete solution to the so-called "cyclotomic equations".

Gauss' Cyclotomy Theorem. Recall that the equation $x^n - 1 = 0$ has n distinct complex roots, given explicitly by

$$1 = \omega^0, \omega, \omega^2, \ldots, \omega^{n-1},$$

where $\omega = e^{2\pi i/n} = \cos(2\pi/n) + i\sin(2\pi/n)$.[4] Thus, from the factorization

$$x^n - 1 = (x-1)(x^{n-1} + x^{n-2} + \cdots + x + 1),$$

we conclude that the *cyclotomic equation*

$$x^{n-1} + x^{n-2} + \cdots + x + 1 = 0.$$

has $n-1$ distinct complex roots. Namely: $\omega, \omega^2, \ldots, \omega^{n-1}$. Gauss gave an explicit algorithm (though not an explicit formula) to express each of these roots in terms of rational numbers $\mathbb{Q}$, field operations $+, -, \times, \div$ and radicals $\sqrt{}, \sqrt[3]{}, \ldots, \sqrt[n-1]{}$ of order less than n.[5] However, he only provided a complete proof in the case that n is prime, and it seems that he lost interest in the subject after this. For more information see Olaf Neumann (2010). We will discuss the details of Gauss' theorem at the end of next semester. ///

In the other direction, some mathematicians made progress in showing that the general quintic is **not** solvable by radicals. Paolo Ruffini claimed to have a proof of this result in 1799 but it was flawed. Finally, in (1826) the young Norwegian mathematician Niels Henrik Abel built on Gauss' work to provide the first rigorous proof of the following theorem.

The Abel-Ruffini Theorem. If $n \geq 5$ then 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

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

[3] From a 1781 letter of Lagrange to d'Alembert, quoted in *Mathematical Expeditions* (1999) by Laubenbacher and Pengelley. The word "geometer" was the generic 18th century term for a mathematician.

[4] Indeed, it follows easily from de Moivre's formula that every power of ω is a root of the equation $x^n - 1 = 0$. We will give a rigorous proof next semester that a polynomial of degree n can have **no more** than n distinct complex roots.

[5] It was traditional to insist on radicals of order no greater than the degree of the equation, since otherwise we have the completely trivial expression $\sqrt[n]{1}$ that includes all of the roots.

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

I should clarify an important point here. The Fundamental Theorem of Algebra guarantees that any polynomial of degree n with complex coefficients possesses a full set of complex roots. That is, given any complex numbers $a_0, \ldots, a_n \in \mathbb{C}$, there exist some complex numbers $r_1, \ldots, r_n \in \mathbb{C}$ (possibly not distinct), such that

$$a_0 + a_1 x + \cdots + a_n x^n = (x - r_1)(x - r_2) \cdots (x - r_n).$$

This fact was generally believed throughout the 1700s but it is quite difficult to prove. The first satisfactory proofs only appeared in the early 1800s. (We will see my favorite proof, due to Laplace, in Week ?) Thus the **existence** of roots is not in question. It is the **nature** of the roots that is important. Abel and Ruffini proved that the roots (which always exist) can not in general be expressed algebraically in terms of the coefficients.

As with the work of Lagrange and Gauss, Abel's work was deep and technical. Furthermore, it did not provide a complete understanding of the problem because it did not explain the distinction between solvable and unsolvable polynomials. Abel intended to investigate this issue further:

> *I am working at this time on the theory of equations, my favorite topic, and it seems to me that I have finally found the means of solving the general problem, which is* **to determine the form of all the algebraic equations which can be solved algebraically**.[6]

But then Abel died of tuberculosis in 1829, at the age of 26. Meanwhile, a young Frenchman named Évariste Galois was working in obscurity on the same problem.[7] Galois had some brilliant and visionary ideas but he also had a volatile personality. He died in a duel in 1832, at the age of 20, before he could gain any recognition for his ideas.

This brings us to the end of the pre-modern era. The two greatest algebraists of the age (Abel and Galois) were dead—one of whose work was well known and one of whose work was still unknown. However, Galois' work was not lost forever. Under the instigation of Galois' family and friends, the eminent mathematician Joseph Liouville undertook the task of editing and publishing Galois' work in (1846). Due to political controversies[8] Liouville felt the need to defend this decision:

[6] From a letter of Abel to Berndt Holmboe, October 1826. Quoted in Kiernan, *The Development of Galois Theory from Lagrange to Artin* (1971, page 71).

[7] It is worth mentioning that Galois was not aware of Abel's work. Instead, he was inspired by the earlier work of Lagrange and Gauss.

[8] These controversies included Galois' actions after the revolution of 1830 and antagonistic letters toward prominent mathematicians. The short life of Galois was extremely colorful. As an entry point I recommend the online article *Radical Solutions* (2020) by Marisa Brook and J. A. Macfarlane.

> *... yielding to pleas of some friends of Évariste, I have devoted myself, so to speak, under the gaze of his brother, to a careful study of all the printed or handwritten papers which he has left behind. As for us, who have neither known nor even seen this ill-fated young man, we limit ourselves to our role as geometer; the observations we can allow ourselves, publishing these works under the instigation of his family, concerns only the mathematics.*[9]

The publication of Galois' work set off a chain reaction that within 100 years would completely redefine the word "algebra". The transformation began in France with the work of Augustin-Louis Cauchy and Camille Jordan. But then the center of activity shifted to Germany, where Richard Dedekind and Leopold Kronecker were the main innovators. Finally, the new style of algebra came to maturity in the 1920s in the work of Emil Artin and Emmy Noether, which was immortalized in the epoch-making textbook *Moderne Algebra* (1930) by Bartel van der Waerden. This new "modern" version of algebra is the subject of our course. The subject is quite deep and it will take two full semesters to tell the whole story.

1.2 Lagrange's Solution of the Quadratic

The essence of Galois' work is that it shifts the focus of algebra from "numbers" and "polynomial equations" to "symmetries" and "relationships among symmetries". I will begin to 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

$$x^2 - e_1 x + e_2 = 0.$$

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

$$x^2 - e_1 x + e_2 = (x - r_1)(x - r_2).$$

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

$$x^2 - e_1 x + e_2 = (x - r_1)(x - r_2)$$

[9] Liouville (1846, page 382), quoted in Lützen (1990, page 561).

$$x^2 - e_1 x + e_2 = x^2 - (r_1 + r_2)x + r_1 r_2.$$

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

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

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

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

Naively, we would hope to "invert the system", to obtain an equivalent system of equations of the form

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

But this is **impossible**. Indeed, since each of $e_1(r_1, r_2)$ and $e_2(r_1, r_2)$ is symmetric under permuting $r_1 \leftrightarrow r_2$, then any function of $e_1(r_1, r_2)$ and $e_2(r_1, r_2)$ is also symmetric. To be specific, consider any function f of e_1 and e_2, so that f is also a function of r_1 and r_2. Then we have

$$f(r_1, r_2) = f(e_1(r_1, r_2), e_2(r_1, r_2)) = f(e_1(r_2, r_1), e_2(r_2, r_1)) = f(r_2, r_1).$$

On the other hand, the simple functions $f(r_1, r_2) = r_1$ and $g(r_1, r_2) = r_2$ are certainly **not symmetric**:

$$\begin{aligned} f(r_1, r_2) = r_1 \neq r_2 = f(r_2, r_1) \\ g(r_1, r_2) = r_2 \neq r_1 = g(r_2, r_1). \end{aligned}$$

Thus, Lagrange's problem is to somehow "break the symmetry" of the symmetric functions $e_1(r_1, r_2) = r_1 + r_2$ and $e_2(r_1, r_2) = r_1 r_2$ to obtain the non-symmetric functions $f(r_1, r_2) = r_1$ and $g(r_1, r_2) = r_2$. To do this, Lagrange first made a change of variables. He defined two new functions[10] $s_1(r_1, r_2)$ and $s_2(r_1, r_2)$ as follows:

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

Note that this (linear) system is easily invertible:

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

[10] These functions are called "Lagrange resolvents". We will return to this story in Week 23 below.

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

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

The first one is easy:

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

But solving for s_2 in terms of e_1 and e_2 is still impossible because the function 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 $s_2(r_1, r_2) = -s_2(r_2, r_1)$.]

So here is the central issue:

> How can we convert the alternating function $s_2(r_1, r_2) = r_1 - r_2$ into a symmetric function of r_1 and r_2?

This is easy—we just square it:

$$s_2^2(r_1, r_2) := [s_2(r_1, r_2)]^2 = (r_1 - r_2)^2 = r_1^2 - 2r_1r_2 + r_2^2.$$

Since s_2^2 is a symmetric function, a general theorem[11] guarantees that we can express s_2^2 as a polynomial in the "elementary" symmetric functions e_1 and e_2. (This is why I use the letter "e" for the coefficients.) Later we will discuss a general algorithm, but for now trial-and-error works fine:

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

Finally, let $s_2 = \sqrt{e_1^2 - 4e_2}$ denote **one** of the two square roots of $e_1^2 - 4e_2$, it doesn't matter which. (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? ///

[11] See Exercise 18.D below. Jean-Pierre Tignol (2001) attributes this general theorem to Edward Waring (1770), but notes that it was considered common knowledge at the time. Indeed, Lagrange himself (1770, article 98, page 372) called the result "self-evident".

1.3 Lagrange's Solution of the Cubic

Now we extend Lagrange's method to solve cubic equations. 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 (non-linear) 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 inputs r_1, r_2, r_3. For example,

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

Our job is to "break the symmetry" in a controlled way. Lagrange's first step is to define three new functions s_1, s_2, s_3 by a system of linear equations

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

where $\omega = e^{2\pi i/3}$. This system is invertible:[12]

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

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 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 is the question:

> 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?

We should expect this to be tricky. Indeed, the Abel-Ruffini theorem says that the analogous problem in degrees 5 and above is not solvable. After a bit of trial-and-error you might find that s_2s_3 is a symmetric function and after

[12] Today this is called "discrete Fourier transform". It generalizes to any dimension.

a **lot** of trial-and-error you might find that $s_2^3 + s_3^3$ is a symmetric function. Thus each of these can be expressed as a polynomial in e_1, e_2, e_3. There is an algorithm to do this (see See Exercise 18.D), but for now I will just tell you the answers:

$$\begin{cases} s_2 s_3 & = e_1^2 - 3e_2 \\ s_2^3 + s_3^3 & = 2e_1^3 - 9e_1e_2 + 27e_3. \end{cases}$$

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

$$A \coloneqq s_2^3 + s_3^3 = 2e_1^3 - 9e_1e_2 + 27e_3 \qquad \text{and} \qquad B \coloneqq s_2 s_3 = e_1^2 - 3e_2.$$

Then we observe that

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

Thus s_2^3 and s_3^3 are the two roots of a quadratic equation with coefficients A, B, and we can apply the quadratic formula. Let the symbol $\sqrt{A^2 - 4B^3}$ denote a specific (and arbitrary) square root of $A^2 - 4B^3$. Now let us "break the symmetry" by defining

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

Finally, we "break the symmetry" one more time by choosing a specific cube root for each of s_2^3 and s_3^3:[13]

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

And now we're done. Do you recognize this as Cardano's Formula? Instead of writing out the complete formulas for r_1, r_2, r_3 in terms of e_1, e_2, e_3, I'll have you work out an example in Exercise 1.A. ///

The general quartic equation is also solvable in this manner but the details are too complicated to discuss here.[14] At this point you might agree with Lagrange that "the mine has already become too deep and might have to be abandoned". In fact, we **will** abandon this line of thought. Starting next week we will begin to develop a completely new language that will eventually allow us to see the general outlines of the theory of equations without having to deal with the messy details. The need to suppress detail in order to focus on higher-level structure was clearly recognized by Galois. He wrote the following while he while he was in prison in December 1832, just six months before his death:[15]

[13] The choice is not completely arbitrary, since s_2 and s_3 must also satisfy $s_2 s_3 = B$. You will prove in Exercise 1.B that this is always possible over the complex numbers.

[14] We will give a high-level treatment of the general quartic in Week 23.

[15] Quoted from Kiernan (1971, page 92). See also Neumann (2011, page 251).

> *Since the beginning of the century, computational procedures have become so complicated that any progress by those means has become impossible, without the elegance which modern mathematicians have brought to bear on their research, and by means of which the spirit comprehends quickly and in one step a great many computations.*

This philosophy applied to the theory of equations became known as "Galois theory". Eventually the same philosophy was applied to other areas of mathematics and the resulting subject was called "modern algebra". Today we just call it "algebra".

Exercises

1.A A Cubic Equation

Consider the equation $x^3 - 6x - 6 = 0$.

(a) Apply Cardano's formula to find one root of the equation.

(b) Apply Lagrange's method to find all three roots.

1.B Interpreting Cardano's Formula

For any complex numbers $A, B \in \mathbb{C}$ we will prove that there exist some complex numbers $u, v \in \mathbb{C}$ such that

$$u^3 + v^3 = A \quad \text{and} \quad uv = B.$$

Furthermore, if $A^2 - 4B^3 \neq 0$ then there are exactly three such pairs (u, v).

(a) Given an integer $n \geq 1$, show that every nonzero complex number $\alpha \in \mathbb{C}$ has at least one complex nth root. [Hint: If $\alpha \neq 0$ then we can write $\alpha = re^{i\theta}$ for some real numbers $r > 0$ and $0 \leq \theta < 2\pi$. Then by the Intermediate Value Theorem there exists a real number $r' > 0$ such that $(r')^n = r$. Consider the number $\alpha' = r'e^{i\theta/n}$.]

(b) If $(\alpha')^n = \alpha \neq 0$ for some $u, \alpha \in \mathbb{C}$ and $n \geq 1$, show that α has exactly n complex nth roots given by[16]

$$\alpha', \omega\alpha', \omega^2\alpha', \ldots, \omega^{n-1}\alpha', \quad \text{where } \omega = e^{2\pi i/n}.$$

(c) From (a) we know that there exists $\delta \in \mathbb{C}$ such that $\delta^2 = A^2 - 4B^3$. Show that $\alpha, \beta = (A \pm \delta)/2$ satisfy $\alpha + \beta = A$ and $\alpha\beta = B^3$.

[16] For now you can assume that a polynomial of degree n has **no more** than n distinct roots. We will prove this rigorously in Week 16. The proof is not difficult but it would distract us from the topic at hand.

(d) Continuing from (c), we want to find $u, v \in \mathbb{C}$ such that $u^3 = \alpha$, $v^3 = \beta$ and $uv = B$. From (a) we know that there exist some $\alpha', \beta' \in \mathbb{C}$ satisfying $(\alpha')^3 = \alpha$ and $(\beta')^3 = \beta$, hence $(\alpha'\beta')^3 = B^3$. If $\omega = e^{2\pi i/3}$ then from (b) we must have

$$\begin{aligned} u &\in \{\alpha', \omega\alpha', \omega^2\alpha'\}, \\ v &\in \{\beta', \omega\beta', \omega^2\beta'\}, \\ B &\in \{\alpha'\beta', \omega\alpha'\beta', \omega^2\alpha'\beta'\}. \end{aligned}$$

For each possible value of B, show that there exists at least one choice of (u, v) such that $uv = B$. Furthermore, if $A^2 - 4B^3 \neq 0$ show that there are exactly three such pairs (u, v).

1.C Interpreting de Moivre's Quintic

For a given complex number $a \in \mathbb{C}$, let $\delta \in \mathbb{C}$ be any square root of $a^2 - 1$ and let $u \in \mathbb{C}$ be any fifth root of $a + \delta$, which exist by Exercise 1.B(a).

(a) Observe that $a + \delta \neq 0$ and hence $u \neq 0$. Check that $x = u + 1/u$ is a root of de Moivre's quintic equation:

$$x^5 - 5x^3 + 5x - 2a = 0.$$

(b) Since $a + \delta \neq 0$ we know from Exercise 1.B(b) that $a + \delta$ has five distinct complex fifth roots $u_1, u_2, u_3, u_4, u_5 \in \mathbb{C}$. Now define

$$r_k = u_k + \frac{1}{u_k} \in \mathbb{C} \quad \text{for } k \in \{1, 2, 3, 4, 5\}.$$

If $|a + \delta| \neq 1$,[17] prove that the roots $r_1, r_2, r_3, r_4, r_5 \in \mathbb{C}$ are distinct, hence they give the complete solution of de Moivre's quintic.[18] [Hint: Observe that $|u_k|^5 = |a + \delta|$ for any k. If $u_k \neq u_\ell$, show that $r_k = r_\ell$ implies $u_k u_\ell = 1$, and use this to get a contradiction.]

(c) Use (b) to find all five roots of the equation $x^5 - 5x^3 + 5x - 4 = 0$.

[17] One can show that this condition is equivalent to $a \neq \pm 1$.

[18] Remark: Even when the list $r_1, r_2, r_3, r_4, r_5 \in \mathbb{C}$ contains repetition, we still have

$$x^5 - 5x^3 + 5x - 2a = (x - r_1)(x - r_2)(x - r_3)(x - r_4)(x - r_5),$$

but it is very annoying to check this directly. This is precisely the kind of calculation that Galois theory will allow us to circumvent.

Week 2

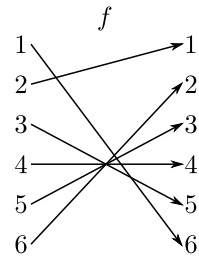
2.1 Permutations

Last week we saw Lagrange's approach to the solution of equations, and I mentioned that Galois "changed the rules of the game". Specifically, he shifted the focus from "symmetric functions" to the concept of "symmetry for its own sake". What does this mean?

Permutations.[19] 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 $\{1, 2, \ldots, n\}$. Let S_n denote the set of all permutations

$$f : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}.$$

For example, here is a typical element of S_6:



It is cumbersome to draw the full diagram every time, so we define the following two notations.

Word Notation. Given $f : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ we prefer to write f_i instead of $f(i)$. Then to specify the function f it is enough to give the list of

[19] Though the concept had appeared in Lagrange's work (1770), the first mathematician to consider permutations for their own sake was Augustin-Louis Cauchy, *Mémoire sur le nombre des valeurs* (1815). The French word for permutation is *substitution* because these were thought of as substitutions of the inputs in a multi-variable function $\mathbb{C}^n \to \mathbb{C}$. See Exercise 10.A for more details.

values $f_1, f_2, \ldots, f_n$. To save as much space as possible we will even omit the commas and write[20]

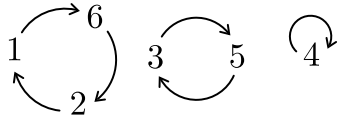
$$f = f_1 f_2 \cdots f_n.$$

For example, the permutation above is

$$f = 615432.$$

[**Exercise:** Explain why $\#S_n = n!$.]

Cycle Notation. Word notation is the most efficient 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 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:** Explain why the set S_n contains $n(n-1)/2$ transpositions.]

[20] However, this notation becomes impractical when $n \geq 10$.

2.2 Definition of Groups

The following theorem is Galois' big contribution to mathematics. I will state the result in modern language, with some details temporarily suppressed.

Galois' Solvability Theorem. Consider a positive integer $n \geq 1$ and let S_n be the set of all permutations of $\{1, 2, \ldots, n\}$. Let $id : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ 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

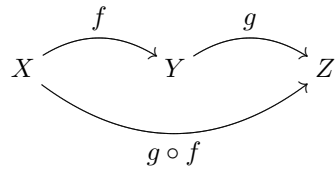
$$S_n = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_r = \{id\}$$

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

The details are quite complicated. For now I can give a brief description of how this relates to Lagrange's method. The set $G_0 = S_n$ tells us that each coefficient of a polynomial is symmetric under every permutation of the roots. The set $G_r = \{id\}$ tells us that each individual root of a polynomial is symmetric under no permutations of the roots. Each step $G_{i-1} \supseteq G_i$ correspond to "breaking the symmetry" by choosing an arbitrary root of some function. The advantage of Galois' reformulation is that it will allow us (in Week 10) to give a short proof of unsolvability for $n \geq 5$, that does not even mention "equations" or "roots". However, the full proof of Galois' theorem will have to wait until the end of the course.

Our first job is to describe the technical conditions satisfied by $G_{i-1} \supseteq G_i$.

Composition of Permutations. Let X, Y, Z be sets and let $f : X \to Y$ and $g : Y \to Z$ 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 we write functions to the left of their arguments:

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

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

Proof. The inverse permutations satisfy $f \circ f^{-1} = f^{-1} \circ f = id$ and $g \circ g^{-1} = g^{-1} \circ g = id$, where id is the identity permutation. Then since functional composition is "associative" we have

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

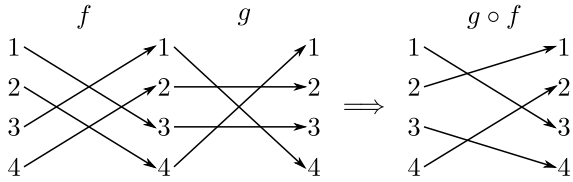
and

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

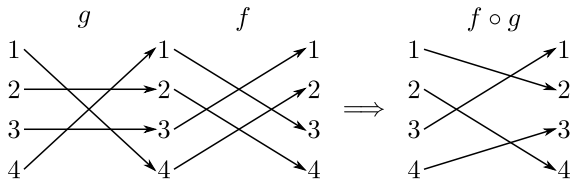
In other words, $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. □

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

Here is a picture showing that $g \circ f = 3142 = (1342)$:



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



We note from this example 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 $\circ : S_n \times S_n \to S_n$, which is associative but not necessarily commutative. Furthermore, every element $f \in S_n$ has a compositional inverse $f^{-1} \in S_n$ (indeed, permutations are invertible by definition), and there exists a special element $id \in S_n$ satisfying

$f \circ id = id \circ f = f$ for all $f \in S_n$. Galois used the word "group" to encapsulate these three properties. Here is the modern axiomatic formulation.

Definition of Groups and Subgroups.[21] Let G be a set equipped with an abstract binary operation $* : G \times G \to G$, which we will write as $(a, b) \mapsto a * b$. We say that the pair $(G, *)$ is a *group* if the following three[22] axioms hold:

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

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

(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).[23]

Now let $H \subseteq G$ be any subset. We say that H is a *subgroup* of $(G, *)$ 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.

In other words: A subgroup is a subset that is also a group with respect to the same operation $*$ and identity ε. ///

Remarks:

[21] The axiomatic definition of groups was first stated by the English mathematician Arthur Cayley, *On the theory of groups* (1854), but this work was not very influential. The group axioms were later studied by several American mathematicians, including E.V. Huntington, *Simplified definition of a group* (1902). The use of axioms for the development of group theory did not become standard until the work of German mathematicians in the early twentieth century.

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

[23] Many students are bemused that such a basic property (i.e., commutativity) has such a pretentious name. Apparently the notation goes back to an 1856 paper of Leopold Kronecker. See *A history of abstract algebra* (2018, page 140) by Jeremy Gray.

- In Exercise 2.A below you will prove that a subset $H \subseteq G$ is a subgroup if and only if for all $a, b \in H$ we have $a * b^{-1} \in H$. This is called the "one-step subgroup test".
- In Exercise 2.C 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}. Later in Exercise 3.C you will generalize this notation to define a^n for all $n \in \mathbb{Z}$.

What led Galois to the definition of "groups"? Recall from Lagrange's method that certain functions $f : \mathbb{C}^n \to \mathbb{C}$ are "invariant" or "symmetric" under certain permutations of their inputs. To be precise, let $f(x_1, \ldots, x_n)$ be any function with n inputs and let $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be any permutation. Then we can define a new function $f^\pi : \mathbb{C}^n \to \mathbb{C}$ by permuting the inputs:

$$f^\pi(x_1, \ldots, x_n) := f(x_{\pi(1)}, \ldots, x_{\pi(n)}).$$

Galois made the following observation:

> Let $f : \mathbb{C}^n \to \mathbb{C}$ be any function with n inputs and let $H \subseteq S_n$ be the sub**set** of permutations that leave f invariant:
>
> $$H = \{\pi \in S_n : f^\pi = f\}.$$
>
> Then H is a sub**group** of S_n.[24]

Here is Galois' theorem with the technical conditions filled in.

Galois' Solvability Theorem (Precise Version). To each polynomial equation $f(x) = 0$ of degree n, there corresponds a certain **group** of permutations $G \subseteq S_n$, called the *Galois group* of the equation. The equation is solvable by radicals if and only if there exists a chain of sub**groups**

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_r = \{id\},$$

satisfying the condition that **for each pair** $G_{i-1} \supseteq G_i$ **the quotient group** G_{i-1}/G_i **exists and is abelian**.[25] ///

There are still some undefined terms in this theorem. We will study quotient groups in Weeks 6 and 7, and I will give the official definition of the Galois group of an equation in Week 14. For now let me just tell you that a typical

[24] See Exercise 10.A for more details.
[25] Recall that "abelian" is just a fancy word for "commutative".

equation of degree n has Galois group equal to the full set of permutations. Assuming this, I will present a short proof in Week 10 that the general equation of degree $n \geq 5$ is not solvable by radicals.

The proof of Galois' theorem itself will take much longer. Galois' own treatment was overly concise and the first mathematicians to read it (Cauchy, Fourier and Poisson) did not fully comprehend it. Due to Galois' age they might not even have considered it worth their time. It was 15 years after Galois' death when Joseph Liouville finally worked through the details and showed that Galois' ideas were both correct and significant. As Joseph Bertrand later recalled:[26]

> *When Liouville published the Memoir, which Poisson had found obscure, fifteen years after the death of Galois, he announced a commentary which he has never given. I have heard him declare that the proof was very easy to understand. When he saw that I made a gesture of astonishment he added, "It is sufficient to devote a month or two to it, without thinking of anything else". These words explain and justify the embarrassment to which Poisson dutifully admitted and which was no doubt met by Fourier and Cauchy as well.*

The fact that Liouville's commentary never appeared suggests that even he had difficulty Galois' ideas. Inspired by the publication of Galois' work, the French mathematicians Cauchy and Jordan vastly extended the theory of permutations. This led to the modern subject of "group theory", which is our main topic for this semester.

The French school persisted in using the mathematical language of Lagrange, hence they did not clarify the philosophical shift inherent in Galois' ideas. The person most responsible for our modern language is the German mathematician Richard Dedekind. To the abstract concept of a "group", he added several more, including the abstract concept of a "field".. The subject of "field theory" will be a main topic for next semester. However, I will give a short introduction in the next section so that we can discuss some interesting examples of groups.

2.3 Basic Examples of Groups

The group concept was invented by Galois to study polynomial equations. In retrospect, however, the reason that group theory became so fundamental is the fact that it synthesizes several topics that previously were independent. In my view there are three basic examples:

- Abelian groups come from number theory.
- Groups of permutations come from Galois theory.

[26] From Bertrand's 1902 eulogy of Galois, quoted in Lützen (1990, page 130).

- Groups of matrices come from geometry.

In order to gain a good understanding we must discuss all three types of examples. This can be difficult in a first course due to the students' limited background. In this section I will briefly describe the important ideas without giving all of the details.

Groups of Permutations. These were introduced in the previous section. For every integer $n \geq 1$, the structure $(S_n, \circ, id)$ is a group, called the *symmetric group on n letters*. In Week 8 we will prove "Cayley's Theorem", which says that any finite group can be viewed as a subgroup of S_n for some n, though this is not always a useful point of view. In Exercise 2.B you will investigate an important subgroup $A_n \subseteq S_n$ called the *alternating subgroup*.

Abelian groups. This type of group emerged from number theory, i.e., the study of numbers. The most basic kind of "numbers" are the integers

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}.$$

These form an abelian group under addition, where 0 is the identity element. We will denote this structure by $\mathbb{Z}^+ = (\mathbb{Z}, +, 0)$. The set of natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$ is also closed under addition and contains the additive identity 0. However, the set $\mathbb{N}$ does not contain "additive inverses" (e.g., $-1 \notin \mathbb{N}$) so the structure $(\mathbb{N}, +, 0)$ is not a group. A structure that satisfies axioms (G1) and (G2) but not (G3) is called a *monoid*.[27] Thus $(\mathbb{N}, +, 0)$ is a monoid.

The integers are contained a in a chain of larger number systems (the rational numbers, real numbers and complex numbers), each of which is an abelian group under addition:

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

Numbers can also be multiplied, but this does not immediately give a group structure. The most glaring problem is the fact that 0 does not have a multiplicative inverse. After deleting 0 from the sets $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ 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 is harder to squeeze a multiplicative group from the integers. For example, the equation $2x = 1$ has no integer solution, so the integer 2 has no multiplicative inverse in $\mathbb{Z}$. In fact, the only integers with multiplicative inverses are ± 1. These form a finite group with two elements:

$\times$	1	-1
1	1	-1
-1	-1	1

[27] There are many different ways to weaken the group concept, with names such as "semigroups" and "quasigroups". In this class we will only discuss monoids.

The concept of a "number system" was later axiomatized via the concept of a "ring".[28] Though we will not systematically study this concept until next semester, it is good to see the definition now.

Definition of Rings. Let R be a set equipped with two binary operations $+, \times : R \times R \to R$ and two special elements $0, 1 \in R$. We call this structure a *ring* if the following axioms hold:

(R1) $(R, +, 0)$ is an abelian group.

(R2) $(R, \times, 1)$ is a monoid.

(R3) For all $a, b, c \in R$ we have $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$.

If $ab = ba$ for all $a, b \in R$ then we say that the ring R is *commutative*.[29]

Note that the monoid $(R, \times, 1)$ is never a group because $0a = 0$ for all $a \in R$. In particular, the equation $0x = 1$ has no solution, hence 0 does not have a multiplicative inverse. We can create a group by throwing away all of the non-invertible elements:

$$R^\times := \{a \in R : \text{ the multiplicative inverse } a^{-1} \text{ exists}\}.$$

The structure $(R^\times, \times, 1)$ is called the *group of units* of the ring R. A commutative ring R satisfying $R^\times = R - \{0\}$ is called a *field*. Thus the rings $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields, while the ring $\mathbb{Z}$ is not a field.

Groups of Matrices. Commutative rings are used to model numbers; non-commutative rings are used to model matrices. For any ring R (commutative or non-commutative) we can define a ring of matrices:

$$\mathrm{Mat}_n(R) = \{n \times n \text{ matrices with entries from } R\}.$$

Addition of matrices is defined componentwise, using the addition operation from R. The multiplication operation in $\mathrm{Mat}_n(R)$ is matrix multiplication, defined using the addition and multiplication operations from R. There are two ways to define matrix multiplication. First, the bad way. Consider two matrices $A, B \in \mathrm{Mat}_n(R)$ with entries $a_{ij} \in R$ and $b_{ij} \in R$, respectively. If $c_{ij} \in R$ is the ij entry of the matrix AB, then we have

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

This definition is not very illuminating. For example, it is not at all clear from this definition that matrix multiplication is associative. Here is a better way to define matrix multiplication:

[28] The word *ring* (or *Zahlring*) comes from David Hilbert's *Zahlbericht* (1897). The word "Zahlbericht" means "number report" and "Zahlring" means "number ring". Nobody knows why he chose the word "ring".

[29] For some obscure reason, the prefix "abelian" only applies to groups.

- Let A be an $n \times n$ matrix with jth column vector $\mathbf{a}_j \in R^n$. Then for any column $\mathbf{x} \in R^n$ we define the column $A\mathbf{x} \in R^n$ as a linear combination of the columns of A:

$$A\mathbf{x} = \sum_j x_j \mathbf{a}_j.$$

More visually, we have

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{n1} \\ \vdots \\ a_{nn} \end{pmatrix}$$
$$= \begin{pmatrix} x_1 a_{11} + \cdots + x_n a_{1n} \\ \vdots \\ x_1 a_{n1} + \cdots + x_n a_{nn} \end{pmatrix}.$$

One can check that the function $R^n \to R^n$ defined by $\mathbf{x} \mapsto A\mathbf{x}$ "preserves linear combinations". That is, for any vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r \in R^n$ and scalars $c_1, \ldots, c_k \in R$ we have

$$A(c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r) = c_1 A\mathbf{v}_1 + \cdots + c_r A\mathbf{v}_r.$$

Functions satisfying this property are called *linear*.

- Conversely, for any linear function $f : R^n \to R^n$ we define the $n \times n$ matrix $[f] \in \mathrm{Mat}_n(R)$ whose jth column is given by f applied to the jth standard basis vector:

$$[f] := \begin{pmatrix} f\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & f\begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} & \cdots & f\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \end{pmatrix}.$$

Using this definition one can check that

$$[f]\mathbf{x} = f(\mathbf{x}) \quad \text{for all column vectors } \mathbf{x} \in R^n.$$

This establishes a one-to-one correspondence between $n \times n$ matrices and linear functions $R^n \to R^n$.

- Finally, for any matrices $A, B \in \mathrm{Mat}_n(R)$ we have linear functions $A, B : R^n \to R^n$. We let AB denote the matrix corresponding to the composite function $A \circ B : R^n \to R^n$ (which is also linear). That is, we define the matrix AB so that

$$A(B\mathbf{x}) = (AB)\mathbf{x} \quad \text{for all column vectors } \mathbf{x} \in R^n.$$

One can check that this is equivalent to the previous definition.

The first definition is of course necessary for computations. But the second definition is more meaningful. For example, it explains why matrix multiplication is associative — because it is just the composition of linear functions.

In summary, the operations of addition and matrix multiplication, together with the $n \times n$ *zero matrix* and *identity matrix*

$$O = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \quad \text{and} \quad I = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix},$$

make the set $\mathrm{Mat}_n(R)$ into a ring. **In this class we will only consider matrices over commutative rings** R such as $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$. The group of units of $\mathrm{Mat}_n(R)$ is called the *general linear group*:

$$\begin{aligned} GL_n(R) &:= \mathrm{Mat}_n(R)^\times \\ &= \{A \in \mathrm{Mat}_n(R) : \text{the inverse matrix } A^{-1} \text{ exists}\}. \end{aligned}$$

We care about matrix groups because they are used to describe symmetry in geometry and physics. On the homework you will investigate certain subgroups of $GL_n(R)$. You may use the following facts from linear algebra:

- We say that a square matrix A is invertible if there exists a (necessarily unique) matrix B satisfying $AB = I$ and $BA = I$. Actually, it is sufficient to check that $AB = I$, since it turns out that

  $$AB = I \quad \Longrightarrow \quad BA = I.$$

  The proof of this fact is shockingly difficult and it depends on the theory of dimension for vector spaces. We will develop the proof in a sequence of future exercises.

- There exists a special function $\det : \mathrm{Mat}_n(R) \to R$ called the *determinant*. It satisfies the following properties: $\det(O) = 0$, $\det(I) = 1$, $\det(AB) = \det(A)\det(B)$ and

  $$A^{-1} \in \mathrm{Mat}_n(R) \text{ exists} \quad \Longleftrightarrow \quad \det(A) \in R^\times.$$

  If R is a field then this reduces to the statement

  $$A^{-1} \in \mathrm{Mat}_n(R) \text{ exists} \quad \Longleftrightarrow \quad \det(A) \neq 0.$$

  The theory of determinants is quite intricate, and is beyond the scope of this text.

- Given a matrix A, the *transpose matrix* A^T is defined by reversing the rows and columns. Transposition interacts with multiplication and determinants via the formulas $(AB)^T = B^T A^T$ and $\det(A^T) = \det(A)$. If A^{-1} exists, then we also have $(A^T)^{-1} = (A^{-1})^T$.

Exercises

2.A One Step Subgroup Test

Let $(G, *, \varepsilon)$ be a group and let $H \subseteq G$ be a subset. Prove that H is a subgroup if and only if for all $a, b \in G$ we have

$$a, b \in H \quad \Longrightarrow \quad a * b^{-1} \in H.$$

2.B Working With Permutations

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

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

(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 $\{1, 2, \ldots, n\}$. 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.

[**Remark:** The subgroup $A_n \subseteq S_n$ is called the *alternating subgroup* of S_n. 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. See Exercise 7.A.]

2.C 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 (G3) is called a *right inverse* for a.

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

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

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

2.D Matrix Groups

Let R be a commutative ring. Prove that each of the following sets of matrices is a subgroup of $GL_n(R)$:

$$\begin{aligned} SL_n(R) &= \{A \in \mathrm{Mat}_n(R) : \det A = 1\}, \\ O_n(R) &= \{A \in \mathrm{Mat}_n(R) : A^T A = I\}, \\ SO_n(R) &= \{A \in \mathrm{Mat}_n(R) : A^T A = I \text{ and } \det A = 1\}. \end{aligned}$$

Week 3

3.1 Intersection and Join of Subgroups

We have seen the definition of abstract groups and we have played with the main examples. It is surprising that three basic group axioms lead to an extraordinarily rich theory. The topic of "cyclic groups" will be our first glimpse of this theory. Before diving in, we discuss some generalities on subgroups.

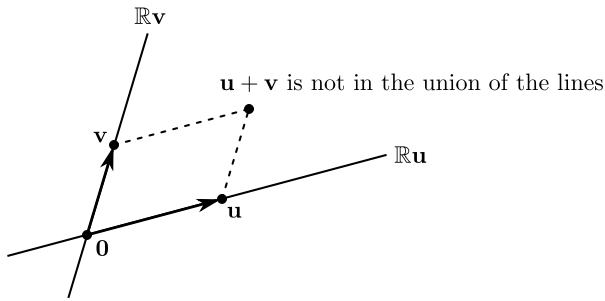
Intersection of Subgroups is a Subgroup. Let $(G, *, \varepsilon)$ be a group and let $H_i \subseteq G$ be any family of subgroups (possibly infinite or even uncountable). Then the intersection $\cap_i H_i \subseteq G$ is also a subgroup.

Proof. We will use the one step subgroup test (Exercise 2.A). 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 that $a * b^{-1}$ is contained in the intersection. □

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

$$\mathbb{R}\mathbf{u} \coloneqq \{\alpha\mathbf{u} : \alpha \in \mathbb{R}\} \qquad \text{and} \qquad \mathbb{R}\mathbf{v} \coloneqq \{\alpha\mathbf{v} : \alpha \in \mathbb{R}\}$$

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:

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

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

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

From the previous result we know that $\langle S \rangle \subseteq G$ is a subgroup. I claim that it is the smallest subgroup of G that contains the set S. We call it the *subgroup of G generated by S*.

Proof. By definition, $\langle S \rangle$ is a subgroup of G that contains S. Let $K \subseteq G$ be any other subgroup that contains S. In this case we must show that $\langle S \rangle \subseteq K$. (This is what we mean by the "smallest subgroup of G containing S".) To show this, we observe that $K \in X$ by definition, and hence[30]

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

as desired. □

Let's examine the previous example in light of this definition. Since the union of two subgroup is not necessarily a subgroup, we define the following operation.

The Join of Two Subgroups. Let G be any group and let $H, K \subseteq G$ be any two subgroups. We define their *join* as the smallest subgroup containing their union:[31]

$$H \vee K \coloneqq \langle H \cup K \rangle. \qquad ///$$

This definition merely establishes existence of the join; it is useless for computations. In special cases we can be much more explicit. In the group $(\mathbb{R}^n, +, \mathbf{0})$, I claim that the join $\mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v}$ is equal to the plane spanned by $\mathbf{u}$ and $\mathbf{v}$:

$$\mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v} = \mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v}.$$

[30] Recall that $A \cap B \subseteq A$ for any sets A and B.

[31] You may have see the symbol "$\vee$" used for "logical conjunction". In Week 5 we will discuss the notion of a "lattice", which unifies the two concepts.

Proof. Since $\mathbb{R}\mathbf{u}$ and $\mathbb{R}\mathbf{v}$ are subsets of $\mathbb{R}\mathbf{u} \cup \mathbb{R}\mathbf{v}$, which is a subset of $\mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v}$, we see that $\mathbb{R}\mathbf{u}$ and $\mathbb{R}\mathbf{v}$ are subsets of $\mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v}$. This means that the vectors $\alpha\mathbf{u}$ and $\beta\mathbf{v}$ are contained in $\mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v}$ for all $\alpha, \beta \in \mathbb{R}$. Since $\mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{v}$ is a subgroup of $\mathbb{R}^n$ this implies that $\alpha\mathbf{u} + \beta\mathbf{v} \in \mathbb{R}\mathbf{u} \vee \mathbb{R}\mathbf{u}$, 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 $\mathbb{R}^n$. Indeed, for any $\alpha_1\mathbf{u} + \beta_1\mathbf{v}$ and $\alpha_2\mathbf{u} + \beta_2\mathbf{v}$ in $\mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v}$ we have[32]

$$(\alpha_1\mathbf{u} + \beta_1\mathbf{v}) - (\alpha_2\mathbf{u} + \beta_2\mathbf{v}) = (\alpha_1 - \alpha_2)\mathbf{u} + (\beta_1 - \beta_2)\mathbf{v} \in \mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v}.$$

Then since the plane contains the set $\mathbb{R}\mathbf{u} \cup \mathbb{R}\mathbf{v}$, it must contains the subgroup **generated** by this set:

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

The key to this proof was the fact that $\mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v} \subseteq \mathbb{R}^n$ is a subgroup. In Exercise 3.B you will show that a similar construction works for all abelian groups, but fails for non-abelian groups.

3.2 Cyclic Groups

Now we examine the simplest possible case of a subgroup generated by a subset.

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

$$\langle g \rangle \coloneqq \langle \{g\} \rangle \subseteq G.$$

We call this the *cyclic subgroup generated by* g. If there exists an element $g \in G$ such that $\langle g \rangle = G$ then we say that G is a *cyclic group*. ///

Again, this abstract definition is useless for computations. To describe cyclic groups explicitly, we introduce the exponential notation.

Exponential Notation for Groups. Let $g \in G$ be an element of a group $(G, *, \varepsilon)$. There is a unique way to define a group element "g^n" for each integer $n \in \mathbb{Z}$ so that

- $g^0 = \varepsilon$,
- $g^1 = g$,
- $g^{m+n} = g^m * g^n$ for all integers $m, n \in \mathbb{Z}$.

[32] This is the one step subgroup test expressed in additive language.

It follows from these properties that the inverse of g is g^{-1}, which agrees with our previous notation.

Proof. Informally, we just define

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

A formal proof would use induction.[33] □

The notations g^{-1} and g^n are based on the intuition that $*$ is multiplication (or functional composition). 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 "$n \cdot g$". Then the defining properties become

- $0 \cdot g = 0$,
- $1 \cdot g = g$,
- $(m + n) \cdot g = m \cdot g + n \cdot g$ for all integers $m, n \in \mathbb{Z}$.

Sometimes we will write $n \cdot g$ as ng when no confusion can result.

Based on this notation we can give a more explicit definition of cyclic groups.

Alternate Definition of Cyclic Groups. Let $g \in G$ be an element of a group $(G, *, \varepsilon)$ and let $\langle g \rangle$ be the smallest subgroup of G containing the element g. Then $\langle g \rangle$ is just the set of powers of g:

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

Proof. Let $P = \{g^n : n \in \mathbb{Z}\}$ be the set of powers of g. We first note that P is a subgroup of G. Indeed, for any powers g^k and g^ℓ we have[34]

$$(g^k) * (g^\ell)^{-1} = g^{k-\ell} \in \{g^n : n \in \mathbb{Z}\}.$$

Then since P contains $g = g^1$, we must have $\langle g \rangle \subseteq P$. On the other hand, since $\langle g \rangle$ is a group and since $g \in \langle g \rangle$ one can prove by induction that $g^n \in \langle g \rangle$ for all $n \in \mathbb{Z}$. Hence $P \subseteq \langle g \rangle$. □

[33] This is similar to the way that addition and multiplication of natural numbers are defined by induction using the Peano axioms.

[34] The general identity $(g^m)^n = g^{mn}$ can be proved by induction.

As a preview of things to come, we will summarize these results in more abstract language. For each element $g \in G$ of a group $(G, *, \varepsilon)$ there exists a special function $\varphi_g : \mathbb{Z} \to G$ defined by

$$\begin{array}{rccc} \varphi_g : \mathbb{Z} & \to & G \\ n & \mapsto & g^n, \end{array}$$

which satisfies the property

$$\varphi_g(m+n) = \varphi_g(m) * \varphi_g(n) \text{ for all } m, n \in \mathbb{Z}.$$

In this next section we will see that φ_g is an example of a *group homomorphism*. This function need not be injective or surjective. It is surjective if and only if every element of G has the form g^n for some $n \in \mathbb{Z}$, i.e., if and only if G is a cyclic group and g is a generator.

What about injectivity? If φ_g is injective then since $\mathbb{Z}$ is infinite the group G must also be infinite. If G is a **finite** group then the function φ_g **cannot** be injective, which means that there exist some integers $k \neq \ell$ satisfying $\varphi_g(k) = \varphi_g(\ell)$. In more mundane terms, the infinite list of elements

$$\ldots, g^{-2}, g^{-1}, \varepsilon, g, g^2, \ldots$$

must contain some repetition. Suppose that $g^k = g^\ell$ for some integers $k < \ell$. Since the group element g^k has inverse g^{-k} we see that

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

where $\ell - k \geq 1$ is a positive integer. If m is the **smallest positive integer** satisfying $g^m = \varepsilon$ then you will prove in Exercise 3.C that the group $\langle g \rangle$ has exactly m elements:

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

The positive integer m is called the *order of the group element* g. If we don't want to waste a letter of the alphabet then we can just write

$$\#\langle g \rangle = \text{order of the group element } g.$$

The oldest non-trivial theorem of group theory says that in a finite group the order of an element always divides the order of the group. It is difficult to attribute this result to any one person since it developed slowly over centuries. I attribute this result to Euler, Fermat and Lagrange, even though they all died before groups were invented.

The Euler-Fermat-Lagrange Theorem. Let $(G, *, \varepsilon)$ be a finite group. Then for any element $g \in G$ we have

$$\#\langle g \rangle \,\Big|\, \#G.$$

In Exercise 3.E you will prove this theorem when G is a finite **abelian** group, using a method due to Euler. In Week 6 we will prove the non-abelian case as an easy corollary to Lagrange's Theorem. In Exercise 6.C you will examine Euler's and Fermat's original versions of the theorem. In Exercise 10.A you will see why Lagrange's name is involved.

This family of theorems is a perfect advertisement for the abstract approach to algebra, since the original proofs were quite intricate, but the modern abstract proof (via Lagrange's Theorem) is almost trivial.

3.3 Isomorphism of Groups

Last time we gave the abstract definition of cyclic groups. Note that any cyclic group $G = \langle g \rangle$ is abelian. Indeed, for any two powers g^m and g^n we have

$$g^m * g^n = g^{m+n} = g^{n+m} = g^n * g^m.$$

Thus any non-abelian group (e.g. the symmetric group S_3) cannot be cyclic. On the other hand, the Euler-Lagrange-Fermat Theorem allows us to prove that any group with a prime number of elements **must** be cyclic.

Groups of Prime Order are Cyclic. Let p be a prime and consider any group G with $\#G = p$. Then there exists an element $g \in G$ such that $G = \langle g \rangle$.

Proof. Let $(G, *, \varepsilon)$ be a group with $\#G = p$ for some prime $p \geq 2$. Since $\#G \neq 1$ we know that $G \neq \{\varepsilon\}$, hence there exists some non-identity element $g \in G$. Consider the cyclic subgroup $\langle g \rangle \subseteq G$. By Euler-Fermat-Lagrange we know that the size of $\langle g \rangle$ divides the size of G. Since $\#G$ prime and $\#\langle g \rangle \neq 1$ this implies that $\#\langle g \rangle = \#G$, hence $\langle g \rangle = G$. □

In fact, the proof demonstrates that **any** non-identity element of G is a generator, hence the number of generators is $p - 1$. For a general cyclic group $G = \langle g \rangle$ of size n we will see later that the number of generators is $\phi(n)$, where ϕ is Euler's *totient function*:

$$\phi(n) = \#\{1 \leq k \leq n : k \text{ is coprime to } n\}.$$

For example, we have $\phi(10) = 4$ because among the numbers $1, 2, \ldots, 10$ only the four numbers $1, 3, 7, 9$ are coprime to 10. We will prove later that if g is a generator of a cyclic group G of size 10 then g^3, g^7 and g^9 are also generators.

The theory of cyclic groups was developed long before the invention of groups, because of two basic examples:

- The group $\mathbb{Z}/n\mathbb{Z}$ of integers modulo n.
- The group Ω_n of complex nth roots of unity.

We will now describe these groups and the relationship between them.

Modular Arithmetic.

Roots of Unity.

3.4 Cyclic and Dihedral Groups

Last time we 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 for each integer $n \in \mathbb{Z}$ we define

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

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:

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

Since every integer is a multiple of 1 (also of -1) we conclude that the additive 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,

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

which is defined by $|a + ib| \coloneqq \sqrt{a^2 + b^2}$. 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 $\mathbb{C}^\times$, which we call $U(1)$:

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

Since these numbers form a circle in the complex plane, we sometimes 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 $\omega \in U(1)$ consists of the integer powers of this element, and hence is a **countable set**. However, the set of all points on the unit circle is **uncountable**. Let's examine the cyclic subgroups of $U(1)$.

Recall that every unit length complex number has the form

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

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

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

This is the *group of n-th roots of unity*. It is important to note that this group does not have a unique generator. In Exercise 3.A you will show (indirectly) that the group of 12-th roots of unity has four possible generators:

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

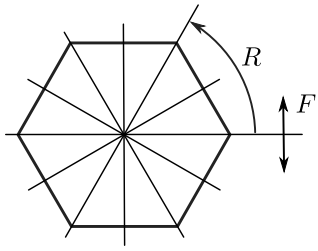
which are called the *primitive 12-th roots of unity*.

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:

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

In this case we say that ω is a "topological generator" of $U(1)$. ///

Example: Symmetries of a Regular Polygon. Consider a regular hexagon. In the following discussion and exercises 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 $f : \mathbb{R}^2 \to \mathbb{R}^2$ 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 faces: front and back). If we let R denote rotation

(counterclockwise) by $2\pi/6$ and if we let F denote any reflection symmetry of the hexagon then one can represent R and F as elements of the orthogonal group $O_2(\mathbb{R})$. Furthermore, one can show that D_{12} is the subgroup of $O_2(\mathbb{R})$ generated by these two elements:

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

Finally, one can show that 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 $f : \mathbb{R}^2 \to \mathbb{R}^2$ that preserves distance and sends the origin to itself is a linear function.
- Any linear function $f : \mathbb{R}^2 \to \mathbb{R}^2$ has the form $f(\mathbf{x}) = A\mathbf{x}$ for some matrix $A \in \mathrm{Mat}_2(\mathbb{R})$.
- If the linear function preserves distance then the corresponding matrix A satisfies $A^T A = I$.
- Any such matrix represents a rotation or a reflection.
- Finally, rotations and reflections are related by the identity $RF = FR^{-1}$.

I will prove some of these assertions next week and you will prove the rest in Exercises 3.D, 4.B and 4.C.

Exercises

3.A Powers of a Cycle

Consider the "standard 12-cycle" in cycle notation:

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

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

3.B 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*:

$$\begin{aligned} H \vee K &\coloneqq \langle H \cup K \rangle \\ &\coloneqq \text{ the intersection of all subgroups that contain } H \cup K. \end{aligned}$$

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

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

Prove that this is a subgroup.

(b) If $(G, +, 0)$ is abelian, use part (a) 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:

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

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

3.C 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

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

(a) Check that $g^m * g^n = g^{m+n}$ for all $m, n \in \mathbb{Z}$.

(b) Use this to prove that $\langle g \rangle \coloneqq \{g^n : n \in \mathbb{Z}\}$ 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 positive integer** satisfying $g^m = \varepsilon$, 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 then we will say that g has *infinite order*.

3.D Matrices of Finite and Infinite Order

Consider the matrices

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

(a) Show that J is invertible and has infinite order.

(b) Next week we will show that the function $\mathbb{R}^2 \to \mathbb{R}^2$ defined by $\mathbf{x} \mapsto R_\theta \mathbf{x}$ rotates each vector counterclockwise around the origin by angle θ. Use this fact to prove that $R_\alpha R_\beta = R_{\alpha+\beta}$ for all angles $\alpha, \beta \in \mathbb{R}$. [Hint: Matrix multiplication is the same as composition of functions.]

(c) For any integer $n \geq 1$, use part (b) to show that the matrix $R_{2\pi/n}$ has order n.

(d) For which angles θ does the matrix R_θ have infinite order?

3.E The Euler-Fermat-Lagrange Theorem, I

Let $(G, *, \varepsilon)$ be a group and let $g \in G$ be any element. Define the function $\mu_g : G \to G$ by $\mu_g(a) \coloneqq g * a$.

(a) Prove that $\mu_g : G \to G$ is a bijection.

(b) Recall that $\langle g \rangle \subseteq G$ is the subgroup consisting of powers of $g \in G$. If G is a **finite abelian** group, prove that

$$\#\langle g \rangle \text{ divides } \#G \text{ for all } g \in G.$$

That is, "the order of each element divides the order of the group". [Hint: Suppose that $G = \{a_1, a_2, \ldots, a_n\}$. Use part (a) to show that $\prod_i a_i = \prod_i f_g(a_i)$. Rearrange and then cancel.]

(c) Use part (b) to show that any abelian group of prime order is cyclic.

[**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". In Exercise **??** you will see what this result has to do with Euler and Fermat.]

Week 4

4.1 Euclidean Space and Orthogonal Matrices

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:

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

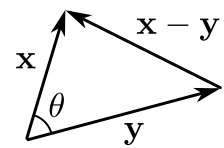
The *length* $\|\mathbf{x}\|$ of a vector $\mathbf{x}$ (i.e., the distance between the points $\mathbf{x}$ and $\mathbf{0}$) is defined by the extended Pythagorean theorem:

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

This allows us to compute the *distance* $\|\mathbf{x} - \mathbf{y}\|$ between any two points $\mathbf{x}, \mathbf{y}$:

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

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



By expanding the the expression $\|\mathbf{x}-\mathbf{y}\|^2 = \langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y}\rangle$ we obtain

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

On the other hand, applying the law of cosines to the same triangle gives

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

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

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

In particular, we find that two nonzero vectors $\mathbf{x}, \mathbf{y}$ are perpendicular precisely when their inner product is zero:

$$\mathbf{x} \perp \mathbf{y} \quad \Longleftrightarrow \quad \cos\theta = 0 \quad \Longleftrightarrow \quad \langle \mathbf{x}, \mathbf{y}\rangle = 0.$$

Thus **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 structure $(\mathbb{R}^n, +, \mathbf{0}, \langle -, -\rangle)$ as n-dimensional *Euclidean space*.[35] ///

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. In Exercise 2.D you considered the "orthogonal group" of $n \times n$ orthogonal matrices:

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

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 by A:

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

[35] Strictly speaking, this structure also includes the binary operation of *scalar multiplication*, $\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$. In other words, Euclidean space consists of the *real vector space* $(\mathbb{R}^n, +, \cdot, \mathbf{0})$ together with the standard inner product. See Exercise 7.C for an axiomatic definition.

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

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

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 $f : \mathbb{R}^n \to \mathbb{R}^n$ be any function that fixes the origin and preserves distances between points. In other words, suppose that

- $f(\mathbf{0}) = \mathbf{0}$,
- $\|f(\mathbf{x}) - f(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Then there exists an orthogonal matrix $A \in O_n(\mathbb{R})$ such that $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Proof. See Exercise 4.B. □

Example: Symmetries of a Regular Polygon (Continued). Now let's return to our discussion of a regular n-sided polygon in the Euclidean plane. By a *symmetry* of the polygon I mean any function $f : \mathbb{R}^2 \to \mathbb{R}^2$ 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 $A^TA = I$. What are the possibilities for this matrix? Suppose that

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

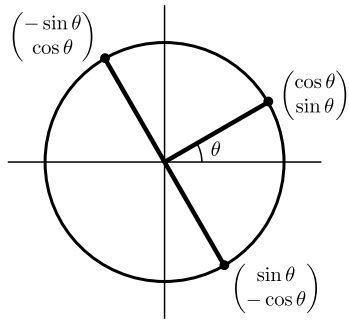
Then the equation $A^TA = I$ tells us that

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

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

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

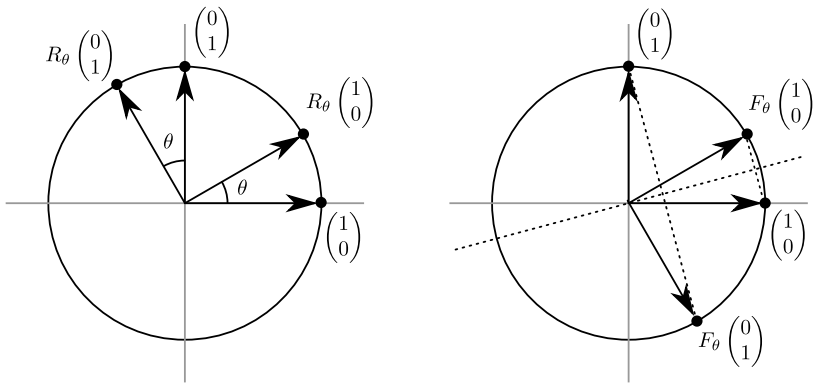
And then since (b, d) is a unit vector perpendicular to (a, c), there are only two possibilities. Namely, we must have $(b, d) = (-\sin\theta, \cos\theta)$ or $(b, d) = (\sin\theta, -\cos\theta)$. Here is a picture:



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

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

You proved in Exercise 3.D that $\mathbf{x} \mapsto R_\theta\mathbf{x}$ is the function that rotates each vector counterclockwise by angle θ. Since rotation is a linear function it was sufficient to prove that R_θ rotates each of the standard basis vectors. The following picture also 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:[36]



[**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

[36] You will formalize this argument in Exercise 4.C.

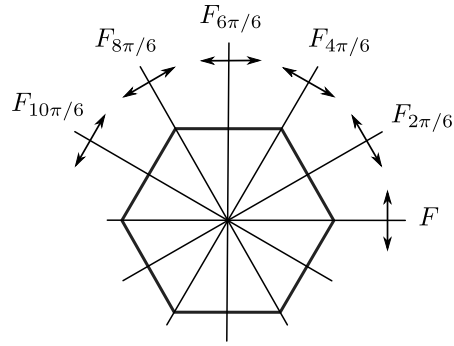
reflections. It turns out that the group consists of n rotations and n reflections, where the rotations form a cyclic subgroup generated by $R \coloneqq R_{2\pi/n}$:

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

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

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

Here is the picture when $n = 6$:



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

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

In Exercise 4.C you will find a more efficient way to work with this group. Namely, if we define $R \coloneqq R_{2\pi/n}$ and $F \coloneqq F_0$ then you will show that the dihedral group is generated by these two elements as follows:

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

4.2 Isomorphism of Groups

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. We express this fact by saying that the group groups are "isomorphic". The word was coined by Camille Jordan in his *Traité des substitutions et des équations algébriques* (1870, page 56), the first textbook-length treatment of group theory.[37]

[37] Today we use *homomorphism* and *isomorphism*, where Jordan used *isomorphism* and *holoedric isomorphism*. The terminology for different kinds of functions (such as "injective" and "surjective") took a long time to stabilize. Van der Waerden lamented the lack of a stable terminology for functions between groups in *Moderne Algebra* (1930, page 32). Even today there is a lack of stable terminology for "group actions". See Week 10.

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

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

We say that G and H are *isomorphic (as groups)*, and write $G \cong H$, if there exists a function $\varphi : G \to H$ satisfying three properties:

- the function $\varphi : G \to H$ is invertible,
- $\varphi : G \to H$ is a group homomorphism,
- $\varphi^{-1} : H \to G$ is a group homomorphism.

In this case say that the function $\varphi : G \to H$ is a *(group) isomorphism.*[38] It defines a one-to-one pairing between the elements of G and H which "preserves the group structure". ///

Remarks:

- The three conditions of isomorphism listed above are not independent. You will show in 4.A that the third condition follows automatically from the first two.
- We write $G \cong H$ to mean that there exists **at least one** group isomorphism $\varphi : G \to H$. But this isomorphism need not be unique. If $G \cong H$ then we will see in Week 8 that the set of all isomorphisms $G \to H$ is itself a group.

Example: Cyclic Groups. Let $(G, *, \varepsilon) = \langle g \rangle$ be a cyclic group. Recall from Exercise 3.C that if $\#G < \infty$ then there exists a **smallest positive integer** m such that $g^m = \varepsilon$, and it follows from this that

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

In this case I claim that

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

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 $k - \ell$ mod m we obtain

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

[38] The words homomorphism and isomorphism are used in many different contexts. Next week we will discuss "poset homomorphism" and "poset isomorphism".

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

$$g^k = g^\ell \quad \Longleftrightarrow \quad k = \ell.$$

Proof. Clearly $k = \ell$ implies $g^k = g^\ell$. Conversely, suppose for contradiction that there exist integers $k \neq \ell$ with $g^k = g^\ell$. Without loss of generality we may assume that $k < \ell$. Then we have

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

where $\ell - k$ is a positive integer. If m is the **smallest** positive integer such that $g^m = \varepsilon$ then from Exercise 3.C we have $\#G = m$, which contradicts the fact that G is infinite. □

With these facts in hand we can prove the following theorem.

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

$$G \cong H \quad \Longleftrightarrow \quad \#G = \#H.$$

In other words, there is exactly one cyclic group of each 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 $\varphi : G \to H$ and there is an obvious candidate:

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

There are four things to check:

> *Well-Defined.*[39] 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
>
> $$g^k = g^\ell \quad \Rightarrow \quad k - \ell \in m\mathbb{Z} \quad \Rightarrow \quad h^k = h^\ell \quad \Rightarrow \quad \varphi(g^k) = \varphi(g^\ell).$$

[39] Whenever a function is defined on a set of equivalence classes, one must check that the image of a class does not depend on the choice of representative used to define the image. This will be formalized when we discuss "quotient groups".

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,

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

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.[40] □

It follows from this theorem that every infinite cyclic group is isomorphic to $\mathbb{Z}^+$. In Weeks 6 and 7 we will see that every finite cyclic group of size n is isomorphic to the "quotient group" $\mathbb{Z}/n\mathbb{Z}$.[41]

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

EXPLAIN BETTER? I THINK I WILL JUST DELETE THE ABSTRACT CHARACTERIZATION OF ADJOINT OPERATORS.

Definition of Unitary Matrices. If $A \in \mathrm{Mat}_n(\mathbb{R})$ 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 characterized the fact that

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

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

$$\langle \mathbf{x}, \mathbf{y} \rangle \coloneqq \mathbf{x}^*\mathbf{y} = \sum_i x_i^* y_i,$$

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

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

[40] Indeed, in this case the function is defined on the set of integers, not on a set of equivalence classes of integers.

[41] Preview: If G is a cyclic group of size n then there exists a surjective group homomorphism $\varphi : \mathbb{Z}^+ \to G$ with kernel $n\mathbb{Z}$.

Based on this, we define the *(special) orthogonal* and *(special) unitary* groups as follows:[42]

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

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 quantum physics. These four groups are distinct in general but for small values of n there can be "accidental isomorphisms".

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

[**Remark:** We call this "Euler's isomorphism" because it is essentially equivalent to Euler's formula:

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

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

Proof. Let $(\alpha) \in \mathrm{Mat}_1(\mathbb{C})$ 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 element of $O(2)$ 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 $\varphi : U(1) \to SO(2)$ by

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

There are four things to check:

[42] In Exercise 2.D we used the notations $O_n(\mathbb{R})$ and $SO_n(\mathbb{R})$ instead of $O(n)$ and $SO(n)$. There is no difference.

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

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

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

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

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 in Exercise 3.D. □

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

Exercises

4.A Homomorphism and Isomorphism

Consider two groups $(G, *, \varepsilon_G)$ and $(H, \bullet, \varepsilon_H)$. A function $\varphi : G \to H$ is called a *(group) homomorphism* when it satisfies the following condition:

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

(a) If $\varphi : G \to H$ is a homomorphism, prove that $\varphi(\varepsilon_G) = \varepsilon_H$.

(b) If $\varphi : G \to H$ is a homomorphism, prove that

$$\varphi(a^{-1}) = \varphi(a)^{-1} \text{ for all } a \in G.$$

(c) Let $\varphi : G \to H$ be a homomorphism and suppose the inverse function $\varphi^{-1} : H \to G$ exists. Prove that the φ^{-1} is also a homomorphism. It follows that invertible homomorphisms are the same as isomorphisms. [Hint: Given elements $a, b \in H$, apply the function φ to the group element $\varphi^{-1}(a) * \varphi^{-1}(b) \in G$.]

4.B 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:

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

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 $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ 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 $f : \mathbb{R}^n \to \mathbb{R}^n$ is any function that preserves distance and sends the origin to itself then it preserves the inner product. It follows that the function is linear, and hence we have $f(\mathbf{x}) = A\mathbf{x}$ for some $n \times n$ matrix A satisfying $A^T A = I$.

(a) Assume that the function $f : \mathbb{R}^n \to \mathbb{R}^n$ 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

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

[Hint: First show that $\langle f(\mathbf{x}), f(\mathbf{x}) \rangle = \langle \mathbf{x}, \mathbf{x} \rangle$ for all $\mathbf{x} \in \mathbb{R}^n$.]

(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 $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$ 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.]

4.C Rotation and Reflection

We have seen that every element of $O(2)$ has the form

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

(a) Verify that $R_\theta \in SO(2)$ and that $F_\theta \in O(2) - SO(2)$.

(b) You proved in Exercise 3.D 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 $R \coloneqq R_{2\pi/n}$ and $F \coloneqq F_0$. Prove that the subgroup $\langle R, F\rangle \subseteq O(2)$ generated by the subset $\{R, F\}$ has $2n$ elements and is given by

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

[Hint: It follows from (c) that $RF = FR^{-1}$.]

4.D Two Groups with Eight Elements

There are two different non-abelian groups with eight elements, called the *dihedral group* D_8 and the *quaternion group* Q_8. We will use multiplicative notation to describe them.

(a) The dihedral group has elements

$$D_8 = \{1, r, r^2, r^3, r^4, f, rf, r^2f, r^3f\},$$

subject to relations $r^4 = f^2 = rfrf = 1$. Write out the full group table.

(b) The quaternion group has elements

$$Q_8 = \{1, i, j, k, e, ei, ej, ek\},$$

subject to the relations $i^2 = j^2 = k^2 = ijk = e$, $e^2 = 1$ and $ae = ea$ for all $a \in Q_8$/ Write out the full group table. [If you want you can write e as "-1" and write the elements ei, ej, ek as $-i, -j, -k$, respectively.]

(c) Prove that D_8 and Q_8 are **not isomorphic**. [Hint: Isomorphic groups must have the same number of elements of each order.]

[**Remark:** In Week 12 we will see that there three different non-isomorphic groups with eight elements:

$$\mathbb{Z}/8/\mathbb{Z}, \quad \mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}, \quad \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}.$$

Thus there 5 non-isomorphic groups of size 8. In general it is hopeless to compute the number of groups of size n. On one extreme, if p is prime then there is only one group of size p — the cyclic group. On the other extreme, Higman[43] showed that there are at least $p^{2k^2(k-6)/27}$ non-isomorphic groups of size p^k. In fact, "most" groups have size 2^k: computations have shown that among all groups of size ≤ 2000, more than 99% have size $1024 = 2^{10}$.]

[43] Higman, *Enumerating p-groups. I. Inequalities.* (1960)

Week 5

5.1 Posets and Lattices

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) and "lattice". SAY MORE ABOUT THE CORRESPONDENCE THEOREM

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:[44]

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

In the special case of two elements $\{a, b\} \subseteq P$ we say that $\ell \in P$ is the *least upper bound* if it satisfies the following two properties:

- We have $a \leq \ell$ and $b \leq \ell$.

[44] Some authors call this a *complete lattice* and use the term *lattice* to indicate that every **finite** subset has a greatest lower bound and least upper bound. The distinction is not important in this course.

- if $a \leq c$ and $b \leq c$ then we must have $\ell \leq c$.

If ℓ' is another least upper bound of $\{a, b\}$ then the second property implies that $\ell \leq \ell'$ and $\ell' \leq \ell$, hence we conclude from axiom (P2) that $\ell = \ell'$. In this case we will write $\ell = a \vee b$ and we will say that ℓ the *join* of a and b. Dually, $g \in P$ is the *greatest lower bound* of a, b if it satisfies:

- The element g is a lower bound of a and b:

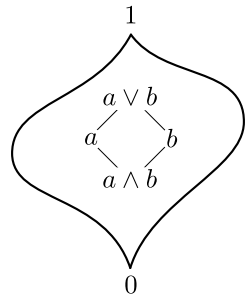
$$\text{we have } g \leq a \text{ and } g \leq b.$$

- The element g is greater than every other lower bound:

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

In this case we will write $g = a \wedge b$ and we will say that g is the *meet* of a and b. We can define the join and meet of any finite set by induction.

We will use the symbols 1 and 0 to denote the least upper bound and the greatest lower bound of the entire set P, so that $a \leq 1$ and $0 \leq a$ for all $a \in P$. [**Exercise:** By convention we will also view 0 and 1 as the least upper bound and the greatest lower bound, respectively, of the empty set $\emptyset \subseteq P$.] 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 $\subseteq$
join $\vee$	union $\cup$
meet $\wedge$	intersection $\cap$
bottom 0	empty set $\emptyset$
top 1	the universe U

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

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

The join operation was defined in Week 3 as follows:

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

Now here is an example with a different flavor.

Example: The Lattice of Divisors. For any integer $n \geq 0$ let $\mathrm{Div}(n) \subseteq \mathbb{N}$ be the set of non-negative divisors of n. Thus we have $\mathrm{Div}(0) = \mathbb{N}$ and for any $n \geq 1$ the set $\mathrm{Div}(n)$ is finite. I claim that $\mathrm{Div}(n)$ is a lattice with the following structure:

partial order $\leq$	reverse divisibility
join $\vee$	greatest common divisor
meet $\wedge$	least common multiple
bottom 0	the integer n
top 1	the integer 1

Indeed, it is immediate that divisibility (or reverse divisibility) satisfies the three axioms of partial order. You will prove in Exercise 5.A that the greatest common divisor $d = \gcd(a, b)$ satisfies the universal property of the join under reverse divisibility (equivalently, the meet under divisibility). That is:

- We have $d|a$ and $d|b$.
- If $c|a$ and $c|b$ then we must have $c|d$.

Dually, one can show that the least common multiple $m = \mathrm{lcm}(a, b)$ satisfies the following properties:

- We have $a|m$ and $b|m$.
- If $a|c$ and $b|c$ then we must have $m|c$.

Remarks:

- The concept of a lattice was introduced by Ernst Schröder in his *Vorlesungen über die Algebra der Logic* (Lectures on the algebra of logic, 1890). In this context the elements of the lattice are logical statements and the meet $\wedge$ and join $\vee$ operations correspond to the logical operators "AND" and "OR", and the special elements "0" and "1" correspond to "false" and "true".

- Dedekind later connected the lattice concept to divisors and subgroups in his paper *Über die Theorie der ganzen algebraischen Zahlen* (On the theory of algebraic integers, 1894). Essentially, this was an abstract approach to the concepts of gcd and lcm. It is interesting to note that Dedekind's term for a lattice was *Dualgruppe*.
- These ideas were not taken up by Dedekind's contemporaries, and did not appear in van der Waerden's *Morderne Algebra* (1930). Instead, the ideas went dormant for 30 years and reemerged in the work of the American mathematicians Garrett Birkhoff and Oystein Ore. Eventually the lattice concept was absorbed into the language of category theory.[45]

5.2 The Lattice of Subgroups of $\mathbb{Z}^+$

Last time we defined lattices of subgroups and lattices of divisors. The following theorem shows how these two concepts are related.[46]

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:

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

The proof has three steps.

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

$$m\mathbb{Z} = \{mk : k \in \mathbb{Z}\} \quad \text{for some } m \in \mathbb{N}.$$

I claim that **every** subgroup has this form.[47]

Proof. Let $H \subseteq \mathbb{Z}^+$ be a 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

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

[45] See Leo Corry's *Modern algebra and the rise of mathematical structures* (2004), especially Section 2.3.

[46] Indeed, Heinrich Weber's *Lehrbuch der Algebra* (1895) uses the same word *Teiler* for both subgroups and divisors. The current German word for subgroup is *Untergruppe*, but *Normalteiler* is still used for normal subgroups. See Week 7 for the definition.

[47] In Week 17 we will re-interpret this result in terms of ring theory, by saying that the ring $(\mathbb{Z}, +, \times, 0, 1)$ is a "principal ideal domain".

We will show that $r = 0$. To see this, first observe that since n and m are in H we also have $r = n - qm \in H$.[48] 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. □

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

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

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

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

Proof. Let $a\mathbb{Z} \subseteq b\mathbb{Z}$. Then since $a \in a\mathbb{Z} \Rightarrow a \in b\mathbb{Z}$ we must have $a = bk$ for some $k \in \mathbb{Z}$, hence $b|a$. Conversely, let $b|a$ so that $a = bk$ for some $k \in \mathbb{Z}$. Then for any $a\ell \in a\mathbb{Z}$ we have

$$a\ell = (bk)\ell = b(k\ell) \in b\mathbb{Z}$$

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

Step 3. Now consider the function $f : \mathbb{N} \to \mathscr{L}(\mathbb{Z}^+)$ defined by $f(m) \coloneqq m\mathbb{Z}$. 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} 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 then we have $a = b = 0$ as desired. Otherwise, both a and b are positive and we have

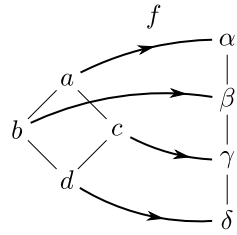
$$\begin{aligned} a &= b\ell \\ a &= ak\ell \\ a(1 - k\ell) &= 0 \\ (1 - k\ell) &= 0 \\ 1 &= k\ell. \end{aligned}$$

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. □

Remarks:

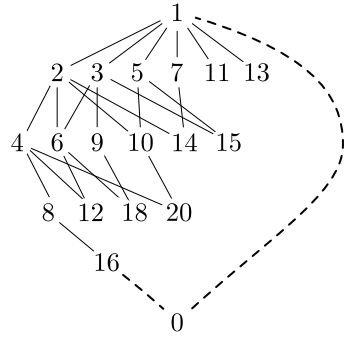
[48] We note that $qm \in H$ for any $q \in \mathbb{Z}$ because H is closed under repeated addition or subtraction.

- 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 group homomorphisms, an invertible poset homomorphism is not necessarily an isomorphism. For example, the following function of posets is invertible and it preserves order:



  However the inverse f^{-1} does **not** preserve order because γ is below β but $c = f^{-1}(\gamma)$ is not below $b = f^{-1}(\beta)$.

This completes the proof that $\mathscr{L}(Z^+)$ is isomorphic to $\mathbb{N}$ under reverse divisibility. Here is a picture of this lattice.[49] Notice that 0 is divisible by every integer, and every integer is divisible by 1.



Since any isomorphism of posets preserves meets and joins, it follows from the theorem that the meet and join operations in $\mathscr{L}(\mathbb{Z}^+)$ correspond to the least common multiple and the greatest common divisor.

Corollary (Meet and Join of Subgroups of $\mathbb{Z}^+$). For all $a, b \in \mathbb{Z}$ we have

$$a\mathbb{Z} \wedge b\mathbb{Z} = a\mathbb{Z} \cap b\mathbb{Z} = \operatorname{lcm}(a, b)\mathbb{Z},$$
$$a\mathbb{Z} \vee b\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z} = \gcd(a, b)\mathbb{Z}. \qquad \square$$

[49] An interesting property of this lattice is that it contains infinite decreasing chains, such as $1 \geq 2 \geq 4 \geq 8 \geq \cdots$. However, the well-ordering principle tells us that there is no infinite **increasing** chain. In modern terms we say that the ring $\mathbb{Z}$ is *Noetherian* but not *Artinian*.

Here are some special cases:

- For any $a \in \mathbb{N}$ we have $a\mathbb{Z} \cap 0\mathbb{Z} = 0\mathbb{Z}$ and $a\mathbb{Z} + 0\mathbb{Z} = a\mathbb{Z}$, hence
$$\text{lcm}(a, 0) = 0 \quad \text{and} \quad \gcd(a, 0) = a.$$
- For any $a \in \mathbb{N}$ we have $a\mathbb{Z} \cap 1\mathbb{Z} = a\mathbb{Z}$ and $a\mathbb{Z} + 1\mathbb{Z} = 1\mathbb{Z}$, hence
$$\text{lcm}(a, 1) = a \quad \text{and} \quad \gcd(a, 1) = 1.$$
- Since $0\mathbb{Z} \cap 0\mathbb{Z} = 0\mathbb{Z}$ and $0\mathbb{Z} + 0\mathbb{Z} = 0\mathbb{Z}$, the corollary also says that
$$\text{lcm}(0, 0) = 0 \quad \text{and} \quad \gcd(0, 0) = 0.$$
You are free to disagree with these last identities if you want.

5.3 Galois Connections

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

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

Fundamental Theorem of Cyclic Groups. For any $n \in \mathbb{N}$ recall that $\text{Div}(n) \subseteq \mathbb{N}$ is 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 $f : \text{Div}(0) \to \mathscr{L}\langle g \rangle$ defined by $f(k) \coloneqq \langle g^k \rangle$ is a poset isomorphism:

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

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

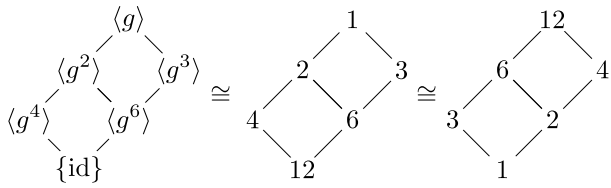
$$(\text{Div}(n), \text{reverse divisibility}) \cong (\mathscr{L}\langle g \rangle, \subseteq).$$

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

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

[**Remark:** This last isomorphism is false when $n = 0$, since there exist infinite decreasing chains in $\text{Div}(0)$ but no infinite increasing chains. See the footnote on page 60.] ///

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 more abstract proof that illustrates what's really going on. The key idea is the concept of a "Galois connection" between posets.

Definition of Galois Connections.[50] Let $(P, \leq)$ and $(Q, \leq)$ be posets and consider any functions $f : P \rightleftarrows Q : g$. The pair f, g is called a *Galois connection* if it satisfies

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

For example, if $f : P \to Q$ 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:

- For all $p_1, p_2 \in P$ and $q_1, q_2 \in Q$ we have

  $$p_1 \leq p_2 \Rightarrow f(p_1) \leq f(p_2) \quad \text{and} \quad q_1 \leq q_2 \Rightarrow g(q_1) \leq g(q_2).$$

  We say that $f : P \to Q$ and $g : Q \to P$ are *poset homomorphisms.*

- For all $p \in P$ and $q \in Q$ we have

  $$p \leq g(f(p)) \quad \text{and} \quad f(g(q)) \leq q.$$

  We say that $g \circ f : P \to P$ is an *increasing function* function and $f \circ g : Q \to Q$ is a *decreasing function.*

- $f \circ g \circ f = f$ and $g \circ f \circ g = g$.

[50] This definition was introduced by Oystein Ore in his *Galois Connexions* (1944) but you won't find it in any standard algebra textbook. I learned this concept from George Bergman's *Invitation to General Algebra* (1998) and it changed the way I teach the subject. It is a convenient way to organize messy families of results such at the Correspondence Theorem for Groups. It is also the prototype for the mid-twentieth-century concept of "adjoint functors" and serves as an excellent entry point into category theory.

Furthermore, if we define the subposets

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

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

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

Proof. See Exercise 5.C. □

[**Remark:** Specific examples of this theorem are often called "correspondence theorems". We will see one of these in the next section.]

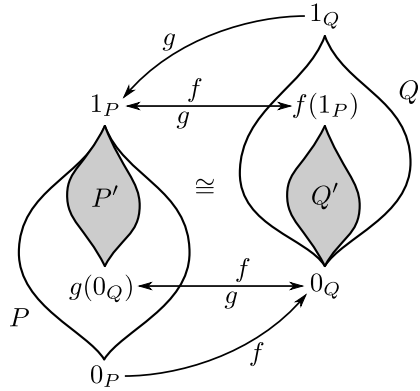
It is easier to understand Galois connections via pictures. Since I can't draw general posets, let me assume for convenience that $(P, \leq, \vee, \wedge, 0_P, 1_P)$ and $(Q, \leq, \vee, \wedge, 0_Q, 1_Q)$ are lattices. In this case I claim that $f(0_P) = 0_Q$ and $g(1_Q) = 1_P$.

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

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

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. □

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 (in Week 14). For now, an example.

Example: Image and Preimage. Let $(G, *, \varepsilon_G)$ and $(H, \bullet, \varepsilon_H)$ be groups and let $\varphi : G \to H$ 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:

$$\varphi[S] \coloneqq \{\varphi(g) : g \in S\} \subseteq H,$$
$$\varphi^{-1}[T] \coloneqq \{g \in G : \varphi(g) \in T\} \subseteq G.$$

Remarks:

- I use square brackets to distinguish between the function $\varphi : G \to H$ that sends elements to elements and the function $\varphi : 2^G \to 2^H$ that sends subsets to subsets.
- The **preimage** function $\varphi^{-1} : 2^H \to 2^G$ always exists, but the **inverse** function $\varphi^{-1} : H \to G$ need not exist. The inverse function exists if and only if for all $h \in H$ the preimage $\varphi^{-1}[\{h\}] \subseteq G$ consists of one element, which we may call $\varphi^{-1}(h) \in G$. SEE EXERCISE ??

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 to any **function** $\varphi : G \to H$. Now let us assume that φ is a **group homomorphism**. In this case you will show in Exercise 5.D that

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

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

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

In other words, we obtain an inclusion-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 $\mathscr{L}(G)'$ consists of subgroups that "contain the kernel of φ", and $\mathscr{L}(H)'$ consists of subgroups of H that "are contained in the image of φ". Next time I will define these notions and I will prove the theorem.

5.4 The Correspondence Theorem for Groups

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

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

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

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

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

From general properties (see Exercise 5.D) we know that the sub**sets** $\ker \varphi \subseteq G$ and $\operatorname{im} \varphi \subseteq H$ are sub**groups**. ///

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

The Correspondence Theorem for Groups. Let $\varphi : (G, *, \varepsilon_G) \to (H, \bullet, \varepsilon_H)$ be any group homomorphism and define the following posets:

$$\mathscr{L}(G, \ker \varphi) \coloneqq \{\text{subgroups } K \subseteq G : \ker \varphi \subseteq K\}$$
$$\mathscr{L}(\operatorname{im} \varphi) \coloneqq \{\text{subgroups } L \subseteq H : L \subseteq \operatorname{im} \varphi\}.$$

I claim that the image and preimage $\varphi : \mathscr{L}(G) \rightleftarrows \mathscr{L}(H) : \varphi^{-1}$ restrict to an isomorphism of posets:

$$\varphi : \mathscr{L}(G, \ker \varphi) \xleftrightarrow{\sim} \mathscr{L}(\operatorname{im} \varphi) : \varphi^{-1}. \qquad ///$$

Proof. Since $\varphi : \mathscr{L}(G) \rightleftarrows \mathscr{L}(H) : \varphi^{-1}$ is a Galois connection we automatically obtain a poset isomorphism $\varphi : \mathscr{L}(G)' \longleftrightarrow \mathscr{L}(H)' : \varphi^{-1}$ between certain subposets $\mathscr{L}(G)' \subseteq \mathscr{L}(G)$ and $\mathscr{L}(H)' \subseteq \mathscr{L}(H)$. In Exercise 5.C you will show that these subposets satisfy

$$\mathscr{L}(G)' = \{K \subseteq G : K = \varphi^{-1}[\varphi[K]]\},$$
$$\mathscr{L}(H)' = \{L \subseteq H : L = \varphi[\varphi^{-1}[L]]\}.$$

So far these are purely poset-theoretic facts that apply to any function between sets. To complete the proof of the theorem we will use the fact that φ is a **group homomorphism** to prove the following identities:

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

There are two steps in the proof. The first step makes heavy use of Exercises 5.C and 5.D.

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

$$\varphi[\varphi^{-1}[L]] = L \wedge \operatorname{im}\varphi,$$
$$\varphi^{-1}[\varphi[K]] = K \vee \ker\varphi.$$

For the first equality, note that $\varphi^{-1}[L] \subseteq G$ implies $\varphi[\varphi^{-1}[L]] \subseteq \varphi[G] = \operatorname{im}\varphi$ because $\varphi[-]$ preserves order, and that $\varphi[\varphi^{-1}[L]] \subseteq L$ because $\varphi \circ \varphi^{-1}[-]$ is decreasing. Therefore we have $\varphi[\varphi^{-1}[L]] \subseteq L \cap \operatorname{im}\varphi = L \wedge \operatorname{im}\varphi$. For the converse, consider any element $h \in L \wedge \operatorname{im}\varphi = L \cap \operatorname{im}\varphi$. Since $h \in \operatorname{im}\varphi$ 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 $L \wedge \operatorname{im}\varphi \subseteq \varphi[\varphi^{-1}[L]]$.

For the second equality, note that $\{\varepsilon_H\} \subseteq \varphi[K]$ implies $\ker\varphi = \varphi^{-1}[\{\varepsilon_H\}] \subseteq \varphi^{-1}[\varphi[K]]$ 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

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

and hence $k^{-1} * g \in \ker\varphi$. Finally, since $K \vee \ker\varphi$ is a subgroup of G containing the elements $k \in K$ and $k^{-1} * g \in \ker\varphi$ we conclude that

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

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

[**Remark:** The equation $\varphi[\varphi^{-1}[L]] = L \wedge \operatorname{im}\varphi$ did not use any group theory. For the equation $\varphi^{-1}[\varphi[K]] = K \vee \ker\varphi$ we used the fact that the product set $K * \ker\varphi := \{k * \ell : k \in K, \ell \in \ker\varphi\}$ is contained in the join $K \vee \ker\varphi$. We will see in Week 9 that in fact $K * \ker\varphi = K \vee \ker\varphi$. More generally, you will show in Exercise 9.B that the product set $KN \subseteq G$ equals the join $K \vee N \subseteq G$ whenever $N \trianglelefteq G$ is a "normal subgroup".]

Step 2. In Step 1 we proved that

$$\mathscr{L}(G)' = \{K \subseteq G : K = K \vee \ker\varphi\},$$
$$\mathscr{L}(H)' = \{L \subseteq H : L = L \wedge \operatorname{im}\varphi\}.$$

Now it only remains to show that

$$K = K \vee \ker\varphi \quad \Longleftrightarrow \quad \ker\varphi \subseteq K,$$

$$L = L \wedge \operatorname{im}\varphi \quad \iff \quad L \subseteq \operatorname{im}\varphi.$$

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

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

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 $a = a \wedge b$. □

This completes the proof of the Correspondence Theorem for Groups. Finally, we obtain 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 $\varphi : \mathbb{Z}^+ \to \langle g \rangle$ defined by $\varphi(k) \coloneqq g^k$. Note that we have $\operatorname{im}\varphi = \langle g \rangle$ 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

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

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

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

and the explicit isomorphism $\operatorname{Div}(n) \to \mathscr{L}\langle g \rangle$ is given by the image function:

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

Exercises

5.A Bézout's Identity

Consider $a, b \in \mathbb{Z}$ with $ab \neq 0$ and let $\operatorname{Div}(a, b) = \{k \geq 1 : k|a \text{ and } k|b\}$ be the set of common divisors. We define $\gcd(a, b)$ as the greatest element of this set.

(a) *Bézout's Identity.* Consider the set of "positive linear combinations":

$$S \coloneqq \{ax + by : x, y \in \mathbb{Z} \text{ and } ax + by > 0\}.$$

This set is non-empty because $ab \neq 0$, hence by well-ordering there exists a least element $d \in S$. Prove that $d = \gcd(a, b)$. [Hint: Divide a by d to obtain $a = dq + r$ with $0 \leq r < d$. Show that $r \neq 0$ leads to a contradiction, therefore we must have $r = 0$ and hence $d|a$.]

(b) Use part (a) to prove that $d = \gcd(a, b)$ satisfies the universal property of the meet under divisibility:

$$\text{If } c|a \text{ and } c|b \text{ then we must have } c|d.$$

It follows from this that $\gcd(a, b)$ can also be characterized as the smallest positive element of the set $a\mathbb{Z} + b\mathbb{Z}$.

[**Remark:** I follow Wikipedia in calling this result "Bézout's Identity". According to Jean-Pierre Tignol (2001, page 87) the result was actually proved by Bachet de Méziriac in 1624. Etienne Bézout proved an analogous result for polynomials in his *Théorie générale des équations algébriques* (1779).]

5.B Order of a Power

Let G be a group and let $g \in G$ be an element of order n (see Exercise 3.C).

(a) For all $k \in \mathbb{Z}$, prove that $\langle g^k \rangle = \langle g^d \rangle$ where $d = \gcd(n, k)$. [Hint: From Exercise 5.A we can write $d = nx + ky$ for some $x, y \in \mathbb{Z}$.]

(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 $n/\gcd(n, k)$.

[**Remark:** This verifies the observation that we made in Exercise 3.A.]

5.C Galois Connections

Let $(P, \leq)$ and $(Q, \leq)$ be posets and let $f : P \to Q$ and $g : Q \to P$ 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

$$p \leq g(f(p)) \qquad \text{and} \qquad f(g(q)) \leq q.$$

(b) For all $p_1, p_2 \in P$ and $q_1, q_2 \in Q$ prove that

$$p_1 \leq p_2 \Rightarrow f(p_1) \leq f(p_2) \qquad \text{and} \qquad q_1 \leq q_2 \Rightarrow g(q_1) \leq g(q_2).$$

(c) For all $p \in P$ and $q \in Q$ prove that

$$f(p) = f(g(f(p))) \qquad \text{and} \qquad g(q) = g(f(g(q))).$$

(d) Define the "images" $P' \coloneqq g[Q] \coloneqq \{g(q) : q \in Q\}$ and $Q' \coloneqq f[P] \coloneqq \{f(p) : p \in P\}$. 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:

$$f : P' \longleftrightarrow Q' : g.$$

5.D Image and Preimage

Let $(G, *, \varepsilon_G)$ and $(H, \bullet, \varepsilon_H)$ be groups and let $\varphi : G \to H$ be any group homomorphism. For every subset $S \subseteq G$ we define the *image set*

$$\varphi[S] \coloneqq \{\varphi(g) : g \in S\} \subseteq H,$$

and for every subset $T \subseteq H$ we define the *preimage set*

$$\varphi^{-1}[T] \coloneqq \{g \in G : \varphi(g) \in T\} \subseteq G.$$

(a) The preimage function $\varphi^{-1} : 2^H \to 2^G$ always exists, however the inverse function $\varphi^{-1} : H \to G$ might not. Prove that the inverse function exists if and only if $\#\varphi^{-1}[\{h\}] = 1$ for all $h \in H$.

(b) If $S \subseteq G$ is a subgroup prove that $\varphi[S] \subseteq H$ is a subgroup.

(c) If $T \subseteq H$ is a subgroup prove that $\varphi^{-1}[T] \subseteq G$ is a subgroup.

(d) Now you have two functions $\varphi : \mathscr{L}(G) \leftrightarrows \mathscr{L}(H) : \varphi^{-1}$ between the subgroup lattices. Prove that this is a Galois connection.

Week 6

6.1 Equivalence Modulo a Subgroup

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 Relations. 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 hold:

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

$$a \sim a.$$

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

$$a \sim b \quad \Longleftrightarrow \quad b \sim a.$$

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

$$(a \sim b \text{ and } b \sim c) \quad \Longrightarrow \quad a \sim c.$$

[**Remark:** The symbol "$=$" always denotes our favorite equivalence relation on a given set.] For each element $a \in S$ we define the *equivalence class*:

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

Then we use the notation

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

You will verify in Exercise 6.A that for all $a, b \in S$ the following three conditions are equivalent:

- $a \sim b$,

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

We say that the equivalence classes form a *partition* of the set S and we express this fact by writing S as a *disjoint union* of classes:

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

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

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

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 result below.

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

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

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 "equivalence mod 0" is the same as "equality":

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

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:

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

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

$$\begin{array}{rcccl} a \sim_H b & \iff & a^{-1} * b \in H & \iff & b^{-1} * a \in H, \\ a \,{}_H\!\sim b & \iff & a * b^{-1} \in H & \iff & b * a^{-1} \in H. \end{array}$$

I claim that each of $\sim_H$ and ${}_H\!\sim$ 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 ${}_H\!\sim$ 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 confirms that the axioms are good.]

The equivalence 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 equivalence 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 prefer the following notations:

$$\begin{aligned} G/H &\coloneqq \text{ the set of left cosets of } H, \\ H\backslash G &\coloneqq \text{ the set of right cosets of } H. \end{aligned} \qquad ///$$

Now before we go any further let me explain the meaning of the notation "G/H". The following basic result has been known as "Lagrange's Theorem" for over 100 years. However, in the words of R. D. Carmichael:

> *In this case we have attributed to Lagrange a theorem which he probably never knew or conjectured, on the ground (it would seem) that he knew a certain special case of it.*[51]

You will examine this "certain special case" in Exercise 10.A. Here is the general modern statement.

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

$$\#(H\backslash G) = \#(G/H) = \#G/\#H,$$

hence the number of elements of H divides the number of elements of G. [**Remark:** This confirms that the fractional notation is good.] ///

Proof. For each element $a \in G$ note that the surjective function $H \to aH$ defined by $h \mapsto a * h$ is invertible with inverse $g \mapsto a^{-1} * g$. Similarly, the surjective function $H \to Ha$ 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 $\#H$:

$$\#G = \#(G/H) \cdot \#H = \#(H\backslash G) \cdot \#H. \qquad \square$$

[51] The quotation is from Carmichael's review in the Bulletin of the American Mathematical Society (1921) of a book by G. H. Hardy on *Some famous problems of the theory of numbers and in particular Waring's problem.*

In Exercise **??** you proved the following theorem for finite abelian groups. Now we can use Lagrange's Theorem to prove it for non-abelian groups.

Corollary (The Euler-Fermat-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 $\#\langle g \rangle = d$. 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

$$g^{\#G} = g^{dk} = (g^d)^k = \varepsilon^k = \varepsilon. \qquad \square$$

I still haven't told you what this has to do with Fermat and Euler. Be patient.[52]

Historical Remarks:

- Gauss introduced the concept of equivalence modulo an integer in his *Disquisitiones Arithmeticae* (1801). He called this relation "congruence", and he used the symbol "$\equiv$" to provide a clear distinction from "$=$".
- Gauss' notion of congruence was successively generalized by various mathematicians. The notion of congruence modulo a subgroup was explicitly defined by Camille Jordan in (1873). Dedekind wrote about the concept in the 1850s, but this work was only published after his death in (1932).
- The abstract concept of equivalence developed slowly over the following years. It appeared in its modern form in van der Waerden's *Moderne Algebra* (1930, page 13).

6.2 Quotients of Abelian Groups

That was the theory. Now let's see some examples of cosets.

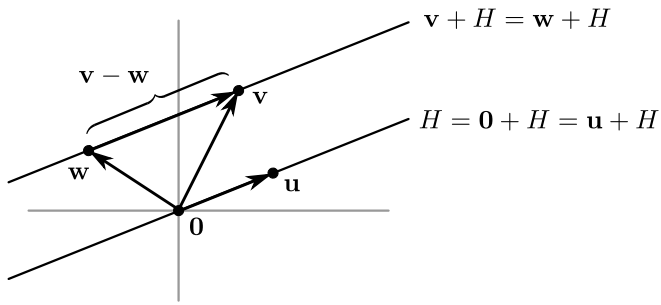
Example: Parallel Lines. Let $G = (\mathbb{R}^2, +, \mathbf{0})$ be the additive group of points in the plane, and for any nonzero vector $\mathbf{0} \neq \mathbf{u} \in \mathbb{R}^2$ let $H = \mathbb{R}\mathbf{u} := \{\alpha\mathbf{u} : \alpha \in \mathbb{R}\}$ be the line through the origin in the 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:

[52] Or skip ahead to Exercise **??**.



Indeed, for any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ we have by definition that

$$\mathbf{v} + H = \mathbf{w} + H \quad \Longleftrightarrow \quad \mathbf{v} - \mathbf{w} \in H \quad \Longleftrightarrow \quad \mathbf{v} - \mathbf{w} \text{ is parallel to } H.$$

In this case the bijection $\tau_{\mathbf{v}} : H \to \mathbf{v} + H$ defined by $\tau_{\mathbf{v}}(\mathbf{x}) = \mathbf{v} + \mathbf{x}$ extends to an isometry of the plane, which we call *translation by* $\mathbf{v}$. Indeed, for any two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ we observe that the distance between $\tau_{\mathbf{v}}(\mathbf{x})$ and $\tau_{\mathbf{v}}(\mathbf{y})$ equals the distance between $\mathbf{x}$ and $\mathbf{y}$:

$$\|\tau_{\mathbf{v}}(\mathbf{x}) - \tau_{\mathbf{v}}(\mathbf{y})\| = \|(\mathbf{v} + \mathbf{x}) - (\mathbf{v} + \mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|.$$

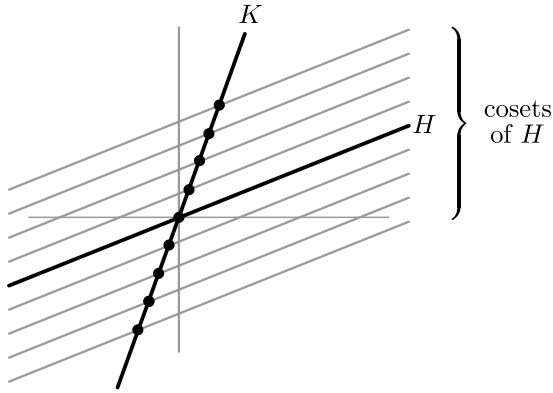
This explains why all of the cosets "look the same". However, note that only one of the cosets (namely, H itself) is a subgroup of G because only one of the parallel lines contains the origin $\mathbf{0} \in \mathbb{R}^2$. In summary, we have

$$G/H = \mathbb{R}^2/\mathbb{R}\mathbf{u} = \text{the set of all lines parallel to } \mathbb{R}\mathbf{u}.$$

Now let $K = \mathbb{R}\mathbf{v} \subseteq \mathbb{R}$ be any other line through the origin, i.e., with $\mathbf{v} \notin H$. Then each coset $\mathbf{w} + H$ intersects K in a unique point so we obtain a bijection $G/H \leftrightarrow K$ defined as follows:

$$\text{the line } \mathbf{w} + H \quad \longleftrightarrow \quad \text{the point of intersection } (\mathbf{w} + H) \cap K.$$

Following the old Euclidean terminology we will call such a bijection a *transversal* of the cosets. Here is a picture:



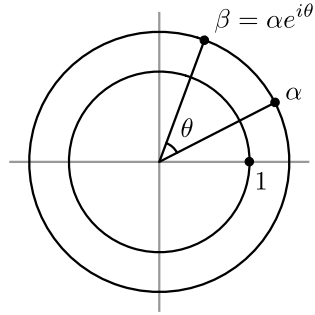
Of course the choice of the line K was arbitrary. In linear algebra it is common to let $K = H^\perp$ be the line (more generally, the complementary subspace) that is orthogonal to H. Then we obtain a bijection

$$G/H \longleftrightarrow H^\perp.$$

But note that $(H^\perp, +, 0)$ is a group. Does this mean that G/H is a group? See below. ///

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** $0 \in \mathbb{C}$:



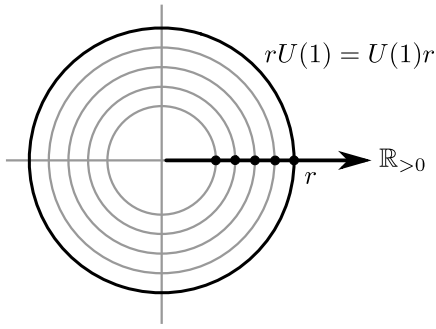
Indeed, we observe that two numbers $\alpha, \beta \in \mathbb{C}^\times$ are in the same coset if and only if they differ by a rotation:

$$\alpha U(1) = \beta U(1) \quad \Longleftrightarrow \quad \alpha^{-1}\beta \in U(1) \quad \Longleftrightarrow \quad \beta = \alpha e^{i\theta} \text{ for some } \theta \in \mathbb{R}.$$

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:

$$\text{the circle } rU(1) = \{re^{i\theta} : \theta \in \mathbb{R}\} \quad \longleftrightarrow \quad \text{the positive real number } r \in \mathbb{R}_{>0}.$$

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 this mean that $\mathbb{C}^\times/U(1)$ is a group? See below. ///

Example: Modular Arithmetic. Let $\langle g\rangle$ be a cyclic group of order $n \geq 1$ and consider the group homomorphism $\varphi : \mathbb{Z}^+ \to \langle g\rangle$ defined by

$$\varphi(k) \coloneqq g^k.$$

By convention the preimage of a single element is called a *fiber*. Note that the fibers of φ are precisely the cosets of $n\mathbb{Z}$:

$$\begin{aligned}
\varphi^{-1}[\{g^k\}] &= \{\ell \in \mathbb{Z} : \varphi(\ell) = g^k\} \\
&= \{\ell \in \mathbb{Z} : g^\ell = g^k\} \\
&= \{\ell \in \mathbb{Z} : \ell - k \in n\mathbb{Z}\} \\
&= \{\ell \in \mathbb{Z} : \ell - k = nm \text{ for some } m \in \mathbb{Z}\} \\
&= \{\ell \in \mathbb{Z} : \ell = k + nm \text{ for some } m \in \mathbb{Z}\} \\
&= \{k + nm : m \in \mathbb{Z}\} \\
&= k + n\mathbb{Z}.
\end{aligned}$$

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 this 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:

$$(k + n\mathbb{Z}) * (\ell + n\mathbb{Z}) \coloneqq (k + \ell) + n\mathbb{Z}.$$

But now the symbol "$*$" looks silly, so let's replace it by "$+$":

$$(k + n\mathbb{Z}) + (\ell + n\mathbb{Z}) \coloneqq (k + \ell) + n\mathbb{Z}.$$

[**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 $(\mathbb{Z}/n\mathbb{Z}, +, 0 + n\mathbb{Z})$ 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.

Theorem/Definition (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

$$a + H = \{a + h : h \in H\} = \{h + a : h \in H\} = H + a.$$

I claim that we can define "addition of cosets" so that for all $a, b \in G$ the following equation makes sense:

$$(a + H) + (b + H) = (a + b) + H.$$

Proof. 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 prove 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.

So let us assume that $a + H = a' + H$ and $b + H = b' + H$, which means that $a - a' \in H$ and $b - b' \in H$. In this case we need to show the cosets $(a + b) + H$ and $(a' + b') + H$ are equal, or, equivalently, that $(a + b) - (a' + b') \in H$. This follows immediately from the fact that H is closed under addition:

$$(a + b) - (a' + b') = (a - a') + (b - b') \in H. \qquad \square$$

[**Remark:** I used the fact that G is abelian when I switched $b - a'$ with $-a' + b$. In fact, the analogous result for non-abelian groups is generally **false**. We will discuss this next week.] Having checked that "addition of cosets" is well-defined, I claim that this operation defines a group structure on the set of cosets G/H.

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

$$\begin{aligned}(a+H)+[(b+H)+(c+H)] &= (a+H)+([b+c]+H)\\ &= (a+[b+c])+H\\ &= ([a+b]+c)+H\\ &= ([a+b]+H)+(c+H)\\ &= [(a+H)+(b+H)]+(c+H). \qquad \square\end{aligned}$$

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.

Exercises

6.A Equivalence = Partition

Let S be a set. A *partition of* S consists of a set of subsets $\{A_1,\dots,A_k\} \subseteq 2^S$ satisfying the following two properties:

- $S = A_1 \cup A_2 \cup \cdots \cup A_k$,
- $A_i \cap A_j = \emptyset$ for all $i \neq j$.

In this case we write $S = \coprod_i A_i$ and we say that S is the *disjoint union* of the subsets A_i. The sets A_i are called the *classes* of the partition.

(a) Given a partition $S = \coprod_i A_i$ we can define a relation $\sim \subseteq S \times S$ by setting

$$a \sim b \quad \iff \quad a \text{ and } b \text{ are in the same class } A_i.$$

Prove that this relation satisfies the three axioms of equivalence.

(b) Conversely, let $\sim \subseteq S \times S$ be any equivalence relation and let $[a]_\sim \subseteq S$ denote the $\sim$-equivalence class of the element $a \in S$. For all $a, b \in S$ prove that the following conditions are equivalent:

(1) $a \sim b$,

(2) $[a]_\sim = [b]_\sim$,

(3) $[a]_\sim \cap [b]_\sim \neq \emptyset$.

Conclude that the set $S/\sim$ of equivalence classes forms a partition of S.

6.B 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.

[**Remark:** We will say much more about this concept in Week 15.]

6.C The Euler-Fermat-Lagrange Theorem, 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 you proved in Exercise 5.A that there exist integers $x, y \in \mathbb{Z}$ with $ax + by = 1$. This is 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

$$a^{\phi(n)} = 1 \mod n.$$

(c) *Fermat's Little Theorem.* If $p \in \mathbb{Z}$ is prime and $p \nmid a$ prove that

$$a^{p-1} = 1 \mod p.$$

[**Remark:** Fermat stated this result (without proof) in a letter to Frénicle in 1640. Leibniz (between 1676 and 1680) and Euler (1731) later gave proofs by induction. Around 1750 Euler obtained a new proof by multiplying together all of the elements of $(\mathbb{Z}/p\mathbb{Z})^\times$, which allowed him to generalize the result to $(\mathbb{Z}/n\mathbb{Z})^\times$.[53] This is the proof that you gave in Exercise **??**.]

[53] See André Weil's *Number Theory: An approach through history from Hammurapi to Legendre* (1984, pages 57–57).

6.D The Chinese Remainder Theorem

In this problem I will use the shorthand notation $[a]_n \coloneqq a + n\mathbb{Z}$. Now fix some $m, n \in \mathbb{Z}$ with $\gcd(m, n) = 1$ and consider the following function from the set $\mathbb{Z}/mn\mathbb{Z}$ to the Cartesian product of sets $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$:

$$\begin{array}{rccc} \varphi: & \mathbb{Z}/mn\mathbb{Z} & \to & \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \\ & [a]_{mn} & \mapsto & ([a]_m, [a]_n). \end{array}$$

(a) Prove that φ is **well-defined**. That is, for all $a, a' \in \mathbb{Z}$ prove that

$$[a]_{mn} = [a']_{mn} \quad \text{implies} \quad [a]_m = [a']_m \text{ and } [a]_n = [a']_n.$$

(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 to conclude that the function φ is **injective**.

(c) Compare cardinalities to show that φ is **surjective**.

(d) Find an explicit formula for the inverse. [Big Hint: Given $([a]_m, [b]_n)$ we need to find some explicit $c \in \mathbb{Z}$ such that $[a]_m = [c]_m$ and $[b]_n = [c]_n$. Try $c \coloneqq any + bmx$.]

(e) Prove that φ restricts to a bijection between groups of units:

$$\varphi : (\mathbb{Z}/mn\mathbb{Z})^\times \longleftrightarrow (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times.$$

[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 function satisfies $\phi(mn) = \phi(m)\phi(n)$ for all $\gcd(m, n) = 1$.

[**Remark:** For distinct primes p, q this result says that we have $\phi(pq) = \phi(p)\phi(q) = (p-1)(q-1)$. 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

$$\begin{aligned} a^{(p-1)(q-1)} &= 1 \mod pq, \\ a^{(p-1)(q-1)k} &= 1 \mod pq, \\ a^{(p-1)(q-1)k+1} &= a \mod pq. \end{aligned}$$

In fact, one can show that the third identity still holds when $\gcd(a, pq) \neq 1$. This result is the foundation of the RSA Cryptosystem.]

Week 7

7.1 Normal Subgroups

Let me recall the final proof from last week 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:

$$(aH) * (bH) \coloneqq (a * b)H \quad \text{for all } a, b \in G.$$

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 prove that $a_1H = a_2H$ and $b_1H = b_2H$ imply $(a_1 * b_1)H = (a_2 * b_2)H$. If G is **abelian** then we have the following proof.

Proof. 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 $(a_1 * b_1)^{-1} * (a_2 * b_2) \in H$. Then we have

$$\begin{aligned}(a_1 * b_1)^{-1} * (a_2 * b_2) &= b_1^{-1} * \big[(a_1^{-1} * a_2) * b_2\big] \\ &= b_1^{-1} * \big[b_2 * (a_1^{-1} * a_2)\big] \\ &= (b_1^{-1} * b_2) * (a_1^{-1} * a_2) \in H\end{aligned}$$

because H is closed under the operation "$*$". □

If G is **non-abelian** then it might be the case that

$$(a_1^{-1} * a_2) * b_2 \neq b_2 * (a_1^{-1} * a_2),$$

which seems to break the proof. But all is not lost. If we can always find some element $h \in H$ such that

$$(a_1^{-1} * a_2) * b_2 = b_2 * h$$

then the operation "$*$" on cosets is still well-defined because

$$(a_1 * b_1)^{-1} * (a_2 * b_2) = b_1^{-1} * \big[(a_1^{-1} * a_2) * b_2\big]$$

$$= b_1^{-1} * [b_2 * h]$$
$$= (b_1^{-1} * b_2) * h \in H.$$

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 considered by Galois all the way back in 1830. If $H \subseteq S_n$ is a subgroup of permutations satisfying the above condition then Galois described the set of cosets S_n/H as a "proper decomposition". Later authors used the adjectives "distinguished" and "invariant". The modern term "normal subgroup" apparently comes from Heinrich Weber's massive *Lehrbuch der Algebra* (1895–1896).[54]

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

$$g * h * g^{-1} \in H.$$

We say that 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

$$H \trianglelefteq G. \qquad ///$$

[**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)⇒(N2)⇒(N3)⇒(N1).

[54] See Gray, *A history of abstract algebra* (2018, page 241).

(N1)⇒(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 (Exercise 6.A) this implies that $Hg = Ha$. We conclude that

$$gH = Ha = Hg.$$

(N2)⇒(N3): Assume that $gH = Hg$ for all $g \in G$. Then for all $g \in G$ and $h \in H$ we have $g * h \in gH$ and hence $g * h \in Hg$. 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)⇒(N1): Assume for all $g \in G$ and $h \in H$ that we have $g * h * g^{-1} \in H$. In this case we will show that left and right equivalence mod H are the same relation on G, and hence the equivalence classes are the same. In other words, for all $a, b \in G$ we want to prove that

$$a^{-1} * b \in H \quad \Longleftrightarrow \quad b * a^{-1} \in H.$$

For one direction, assume that $a^{-1} * b = h \in H$. Then we have

$$b * a^{-1} = a * (a^{-1} * b) * a^{-1} = a * h * a^{-1} \in H.$$

For the other direction, assume that $b * a^{-1} = h' \in H$. Then we have

$$a^{-1} * b = a^{-1} * (b * a^{-1}) * (a^{-1})^{-1} = a^{-1} * h' * (a^{-1})^{-1} \in H. \qquad \square$$

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:

$$R = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix} \qquad \text{and} \qquad F = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

One can check directly from the matrices that $FR = R^2F = R^{-1}F$. To see this geometrically we should think of R^{-1} and 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 **counter**clockwise 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. [**Remark:** The same counting argument shows that $H \trianglelefteq G$ whenever $\#G/\#H = 2$.]

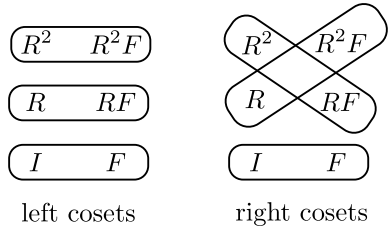
Now for the non-example. Consider the cyclic subgroup $\langle F\rangle = \{I, F\} \subseteq D_6$. Since $\#D_6/\#\langle F\rangle = 3$ it follows again from Lagrange's Theorem that there are three left cosets and three right cosets. Explicitly, the left cosets are

$$\begin{aligned}D_6/\langle F\rangle &= \{\langle F\rangle, R\langle F\rangle, R^2\langle F\rangle\}\\ &= \{\{I, F\}, \{R, RF\}, \{R^2, R^2F\}\}\end{aligned}$$

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 $FR = R^2F$ and $FR^2 = RF$. 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, using $R = (123)$ and $F = (12)$.] ///

7.2 Quotient Groups in General

The definition of normal subgroups might seemed a bit arbitrary. Today I'll show you a modern characterization that is more natural. But first let me remind you of two important concepts:

- Let $(G, *, \varepsilon_G)$ and $(H, \bullet, \varepsilon_H)$ be abstract groups and let $\varphi : G \to H$ be a function. We say that φ is a *group homomorphism* if
$$\varphi(a * b) = \varphi(a) \bullet \varphi(b) \quad \text{for all } a, b \in G.$$
- The *kernel* of the homomorphism is the subgroup of G defined by
$$\ker \varphi = \varphi^{-1}[\{\varepsilon\}] = \{a \in G : \varphi(a) = \varepsilon\} \subseteq G.$$

Historical Remarks:

- The concept of "homomorphism" is more modern than the concept of "isomorphism". The distinction was understood by Camille Jordan in his *Traité* (1870) but the term "homomorphism" apparently originated in lectures of Felix Klein and first appeared in print with Fricke and Klein's *Vorlesungen über die Theorie der automorphen Functionen* (1897).
- The close relationship between homomorphisms and normal subgroups emerged slowly in the 1800s and was strongly emphasized by Emmy Noether in the 1920s.[55] The word "kernel" was imported by Lev Pontryagin (1931) from linear algebra into group theory.
- The use of the arrow notation "$\varphi : G \to H$" is surprisingly recent. The first use of this notation is sometimes credited to Witold Hurewicz (1941). Apparently the arrow notation emerged from a synthesis of the concepts of group homomorphisms and continuous maps in topology. Arrow-theoretic thinking (also called *category theory*) exploded in the mid twentieth-century, to the extent that today the concept of "homomorphism" is more fundamental than the concept of "group".[56] ///

Here is the modern point of view on normal subgroups.

Theorem (Definition of Quotient Groups). Let $(G, *, \varepsilon)$ be a group and let $H \subseteq G$ be any subgroup. Then the following are equivalent:

[55] Her definitive statement appears in *Hypercomplexe Größen und Darstellungstheorien* (1929, page 648). According to van der Waerden, *On the sources of my book Moderne Algebra* (1975), Emmy Noether's motto was: *Es steht alles schon bei Dedekind.* (All of this is already found in the work of Dedekind.) The motto is accurate in this case, since Dedekind realized as early as the 1850s that the cosets of a normal subgroup themselves form a group (see Gray's *History of Abstract Algebra*, page 142).

[56] For more information on terminology see the MathOverflow answers by Francois Ziegler (2017, 2019). See Julia Nicholson (1993) for more information on the development of the quotient group concept.

(N) $H \trianglelefteq G$ is normal,

(N4) H is the kernel of a group homomorphism $\varphi : G \to G'$. ///

Proof. Let $(G', *', \varepsilon')$ be any group and let $\varphi : G \to G'$ be any group homomorphism. To show that $\ker\varphi \subseteq G$ is normal, consider any elements $g \in G$ and $h \in \ker\varphi$. Then 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 $\varphi : G \to G'$ such that $\ker\varphi = H$. The idea is to let G' be the set of left (or right) cosets of H:

$$G' = G/H \ \ (= H\backslash G).$$

Since $H \trianglelefteq G$ is normal I claim that the following operation on cosets is well-defined:

$$(aH) * (bH) \coloneqq (a * b)H.$$

To see this, suppose that we have $a_1H = a_2H$ and $b_1H = b_2H$, so that $a_1^{-1} * a_2 \in H$ and $b_1^{-1} * b_2 \in H$. Since H is normal we have $(a_1^{-1} * a_2) * b_2 = b_2 * h$ for some element $h \in H$. It follows that

$$\begin{aligned}(a_1 * b_1)^{-1} * (a_2 * b_2) &= b_1^{-1} * \big[(a_1^{-1} * a_2) * b_2\big] \\ &= b_1^{-1} * [b_2 * h] \\ &= (b_1^{-1} * b_2) * h \in H,\end{aligned}$$

and hence $(a_1 * b_1)H = (a_2 * b_2)H$. One can easily check 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 $\varphi : G \to G/H$ such that $\ker\varphi = H$, and the choice is completely obvious: for all $g \in G$ we define

$$\varphi(g) \coloneqq gH.$$

This function is a group homomorphism **by definition** and the kernel is

$$\ker\varphi = \{g \in G : gH = H\} = H. \qquad \square$$

Remarks:

- If I were teaching this course for graduate students I would probably take (N4) 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 $\varphi : G \to G/H$ is the *canonical surjection*.
- If $\varphi : G \to G'$ is **any** surjective group 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, consisting of matrices with determinant 1. For example:

$$SL_n(\mathbb{F}) \subseteq GL_n(\mathbb{F}),$$
$$SO(n) \subseteq O(n),$$
$$SU(n) \subseteq U(n).$$

I claim that each of these subgroups is normal.

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 $\mathbb{F}$:

$$\det : G \to \mathbb{F}^\times = (\mathbb{F} - \{0\}, \times, 1).$$

It follows that the kernel of the determinant is a normal subgroup. By definition we call this the "special" subgroup:

$$SG \coloneqq \ker(\det) = \{A \in G : \det(A) = 1\} \trianglelefteq G. \qquad \square$$

In Exercise 7.A you will use the "same proof" to show that the group of alternating permutations is a normal subgroup of the symmetric group:

$$A_n \trianglelefteq S_n.$$

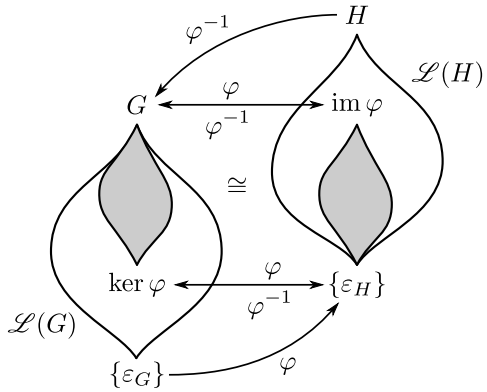
7.3 The First Isomorphism Theorem

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 $\varphi : G \to H$ is a homomorphism of groups then we have seen that there is an isomorphism (called a "Galois Correspondence")

$$\varphi : \mathscr{L}(G, \ker\varphi) \xleftrightarrow{\sim} \mathscr{L}(\operatorname{im}\varphi) : \varphi^{-1}$$

between the the lattice $\mathscr{L}(\operatorname{im}\varphi)$ of subgroups of the image and the lattice $\mathscr{L}(G, \ker\varphi)$ 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 $\varphi : G \to H$ 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:

$$\varphi^{-1}(h) \coloneqq \varphi^{-1}[\{h\}] \subseteq G.$$

Warning: This notation does not imply the existence of the inverse function. Indeed, the inverse function exists if and only if each fiber contains a single element:

$$\varphi^{-1} : H \to G \text{ exists} \quad \iff \quad \#\varphi^{-1}(h) = 1 \text{ for all } h \in H.$$

In this sense the "fiber function" $\varphi^{-1} : H \to 2^G$ from elements of H to subsets of G is a generalization of the concept of "inverse". ///

The fibers of a general function can be strange but the fibers of a group homomorphism are particularly nice.

Lemma (Nonempty Fibers = Cosets of the Kernel). Let $(G, *, \varepsilon_G)$ and $(H, \bullet, \varepsilon_H)$ be groups and let $\varphi : G \to H$ be a group homomorphism. Note that for each element $h \in H - \operatorname{im}\varphi$ we have $\varphi^{-1}(h) = \emptyset$ by definition. I claim that for each element $\varphi(g) \in \operatorname{im}\varphi$ we have[57]

$$\varphi^{-1}(\varphi(g)) = g(\ker\varphi) = (\ker\varphi)g.$$

Then since each coset of the kernel has the same size $k = \#\ker\varphi$, it follows that φ is a k-to-1 map. In particular, we see that

$$\varphi \text{ is injective} \iff \#\ker\varphi = 1 \iff \ker\varphi = \{\varepsilon_G\}.$$

///

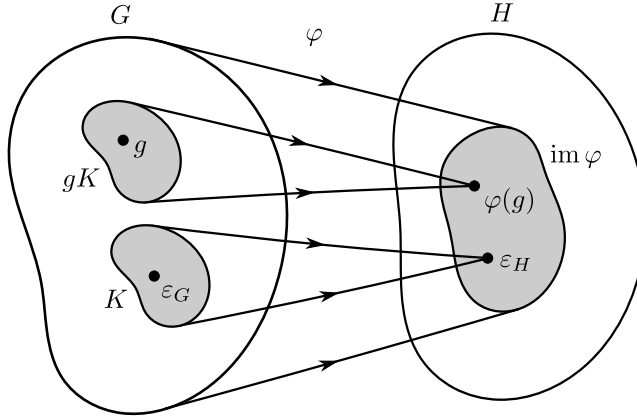
Proof. Fix an element $g \in G$. Then for all elements $a \in G$ we have

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

□

Here is the picture.

The Fiber Picture of a Group Homomorphism. Let $\varphi : G \to H$ be a group homomorphism with kernel $K = \ker\varphi$. Instead of thinking of the lattice of subgroups, I will visualize G and H as sets of points. For each element of the image $h = \varphi(g) \in \operatorname{im}\varphi$, the fiber over h equals the coset gK. Hence each fiber has the same size:



[57] Recall that kernels are normal subgroups, so there is no difference between left and right cosets.

///

But the image of φ 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. Dedekind wrote about this theorem in the 1850s, but this work was not published during his lifetime. The result later became synonymous with Emmy Noether, who emphasized this point of view in the 1920s.[58]

The First Isomorphism Theorem. Let $\varphi : G \to H$ be a group homomorphism. Then the fiber function $\varphi^{-1} : H \to 2^G$ restricts to a group isomorphism $\operatorname{im}\varphi \cong G/\ker\varphi$:

$$\varphi^{-1} : \operatorname{im}\varphi \xrightarrow{\sim} G/\ker\varphi.$$

Proof. If $h \in \operatorname{im}\varphi$ 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 $h_1, h_2 \in \operatorname{im}\varphi$ there exist $g_1, g_2 \in G$ such that $h_1 = \varphi(g_1)$ and $h_2 = \varphi(g_2)$. Then we have

$$\begin{aligned}
\varphi^{-1}(h_1) = \varphi^{-1}(h_2) &\iff g_1(\ker\varphi) = g_2(\ker\varphi) \\
&\iff g_1^{-1} * g_2 \in \ker\varphi \\
&\iff \varphi(g_1^{-1} * g_2) = \varepsilon_H \\
&\iff \varphi(g_1) = \varphi(g_2) \\
&\iff h_1 = h_2.
\end{aligned}$$

- *Surjective.* This is true by definition.
- *Homomorphism.* For all $\varphi(a), \varphi(b) \in \operatorname{im}\varphi$, the lemma says that

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

□

[58] On page 647 of her *Hypercomplexe Größen* (1929), Emmy Noether refers to three *Isomorphiesätze* (Isomorphism Theorems), which she called *Homomorphiesatz* (Homomorphism Theorem), *Erster Isomorphiesatz* (First Isomorphism Theorem) and *Zweiter Isomorphiesatz* (Second Isomorphism Theorem). Somewhere along the way the numbering system got scrambled. Following Wikipedia, I choose to call them the First, Third and Second Isomorphism Theorems, respectively. (See Exercise 7.B.) Dedekind's writing on this topic appears in his collected works (1932), which were edited by Noether.

Now here is a summary of everything we know about group homomorphisms.

Summary: From any group homomorphism $\varphi : G \to H$ we obtain:

(1) an isomorphism of posets $\varphi : \mathscr{L}(G, \ker\varphi) \xrightarrow{\sim} \mathscr{L}(\operatorname{im}\varphi)$,

(2) an isomorphism of groups $\varphi^{-1} : \operatorname{im}\varphi \xrightarrow{\sim} G/\ker\varphi$, and hence

(3) an isomorphism of posets $\mathscr{L}(\operatorname{im}\varphi) \xrightarrow{\sim} \mathscr{L}(G/\ker\varphi)$.

To be specific, we know from (1) that each subgroup of $\operatorname{im}\varphi$ 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 of lattices from $\mathscr{L}(G, \ker\varphi)$ to $\mathscr{L}(G/\ker\varphi)$:

$$\begin{array}{ccccc} \mathscr{L}(G, \ker\varphi) & \xrightarrow{\sim} & \mathscr{L}(\operatorname{im}\varphi) & \xrightarrow{\sim} & \mathscr{L}(G/\ker\varphi) \\ K & \mapsto & \varphi[K] & \mapsto & K/\ker\varphi. \end{array}$$

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. ///

There are a couple more decorations that one can add to this picture, called the Second and Third Isomorphism Theorems, but I will save those for Exercise 7.B. 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 homomomorphism from the additive integers:

$$\begin{array}{rccc} \varphi : & \mathbb{Z} & \to & G \\ & k & \mapsto & g^k. \end{array}$$

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 $n\mathbb{Z}$ for some unique integer $n \geq 0$. Thus we obtain an isomorphism of groups

$$\langle g \rangle = \operatorname{im}\varphi \cong \mathbb{Z}/\ker\varphi = \mathbb{Z}/n\mathbb{Z},$$

and two isomorphisms of lattices:

$$\begin{array}{ccccc} \mathscr{L}(\mathbb{Z}, n\mathbb{Z}) & \xrightarrow{\sim} & \mathscr{L}\langle g \rangle & \xrightarrow{\sim} & \mathscr{L}(\mathbb{Z}/n\mathbb{Z}) \\ d\mathbb{Z} & \mapsto & \langle g^d \rangle & \mapsto & d\mathbb{Z}/n\mathbb{Z}. \end{array}$$

The elements of the leftmost lattice (and hence all three lattices) are in bijection with the set of divisors $\operatorname{Div}(n) = \{d \geq 0 : d|n\}$, which is finite for $n \geq 1$ and infinite for $n = 0$.

Remark: From the group isomorphism $\langle g\rangle \cong \mathbb{Z}/n\mathbb{Z}$ we obtain a bijection between each pair of corresponding subgroups. If $n \geq 1$ then it follows[59] for all $d \in \mathrm{Div}(n)$ that

$$\#(d\mathbb{Z}/n\mathbb{Z}) = \#\langle g^d\rangle = n/d.$$

Lagrange's Theorem does not help in this case because $\#d\mathbb{Z} = \#n\mathbb{Z} = \infty$. This is an example of two infinite groups having a finite quotient.

Exercises

7.A Permutation Matrices

Let S_n be the group of permutations of $\{1, 2, \ldots, n\}$, and for each permutation $f \in S_n$ let $[f] \in \mathrm{Mat}_n(\mathbb{R})$ be the matrix whose i,j-entry is 1 if $f(j) = i$ and 0 if $f(j) \neq i$.

(a) If $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$ is the standard basis, prove that we have $[f]\mathbf{e}_j = \mathbf{e}_{f(j)}$ for each index j.

(b) Use (a) to prove that $f \mapsto [f]$ defines a group homomorphism $S_n \to O(n)$.

(c) Let $\det : O(n) \to \{\pm 1\}$ be the determinant. Use (b) to prove that $\varphi(f) \coloneqq \det[f]$ is a group homomorphism $\varphi : S_n \to \{\pm 1\}$.

(d) Show that the $\ker\varphi$ is the alternating subgroup $A_n \subseteq S_n$ which was defined in Exercise 2.B. [Hint: If t is a transposition then $\varphi(t) = -1$.]

(e) Use the First Isomorphism Theorem and Lagrange's Theorem to conclude that

$$\#A_n = n!\,/\,2.$$

7.B Second and Third Isomorphism Theorems

(a) Let $H, K \subseteq G$ be subgroups with $K \trianglelefteq G$ normal. In this case prove that the product set $HK \coloneqq \{h * k : h \in H, k \in K\} \subseteq G$ is a subgroup (but not necessarily normal).

(b) Continuing from (a), prove that $K \trianglelefteq HK$ is a normal subgroup and the map $h \mapsto hK$ defines a surjective group homomorphism $H \to (HK)/K$ with kernel $H \cap K$. It follows that

$$\frac{H}{H \cap K} \cong \frac{HK}{K}.$$

(c) 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

[59] Recall from Exercise 5.B that the element $g^k \in G$ has order $n/\gcd(n,k)$, thus for all $d|n$ we have $\#\langle g^d\rangle = n/d$.

group homomorphism $G/N \to G/K$ with kernel K/N. It follows that

$$\frac{G/N}{K/N} \cong \frac{G}{K}.$$

[**Remark:** 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 in Exercise 10.B using the Orbit-Stabilizer Theorem.]

7.C Dimension of a Vector Space, Part I

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 $\mathbb{F} \times V \to V$ 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 Lemma.*[60] For all vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m \in V$ we define their *span* as the set

$$\mathbb{F}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \coloneqq \{a_1\mathbf{u}_1 + \cdots + a_m\mathbf{u}_m : a_1, \ldots, a_m \in \mathbb{F}\} \subseteq V$$

and we say that $\mathbf{u}_1, \ldots, \mathbf{u}_m$ is a *spanning set* when $\mathbb{F}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} = V$. We say that $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ is an *independent set* if for all $b_1, \ldots, b_n \in \mathbb{F}$ we have

$$(b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n = \mathbf{0}) \Rightarrow (b_1 = \cdots = b_n = 0).$$

If $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are spanning and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 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 $\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ is spanning. Now show by induction that $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is a spanning set and use this to obtain a contradiction.]

[60] Ernst Steinitz wrote a significant paper on the *Algebraische Theorie der Körper* (1910), which directly inspired the Göttingen school of abstract algebra, and was immortalized in van der Waerden's *Moderne Algebra* (1930). Surprisingly, the "Steinitz Exchange Lemma" does not appear in the 1910 paper, but in a 1913 paper on conditionally convergent series. Steinitz regarded this idea as common knowledge; still, he was the first to write it down.

(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.

[**Remark:** This is the prototype for the concept of "dimension" in any area of mathematics. As you see, it is a subtle concept. You will develop a different characterization of vector space dimension in Exercise 9.D.]

Week 8

8.1 Historical Interlude

So far we have focused on the "external structure" of groups, which deals with the collection of all abstract groups and the collections of homomorphisms between them. The foundational results in this direction are the Correspondence Theorem and the three Isomorphism Theorems. The external point of view is quite modern. One could make the case that it goes back to the work of Galois, but it is more accurate to say that it emerged from the work of Dedekind in the 1850s, whose ideas were later standardized by German mathematicians in the early twentieth century. Today the external point of view is called "category theory".[61]

For the remainder of this semester we will turn to the "internal structure" of groups. This is an older point of view that is concerned with intricate details and specific examples. One could also call this the "French school" of group theory (though it was practiced by Germans and Norwegians as well). The internal structure of groups was first seriously studied by Augustin-Louis Cauchy between 1812 and 1815. The purpose of this work was to explain the ideas of Lagrange (1770) purely in terms of the structure of permutations. Cauchy's first major result (1815) was the statement and proof of Lagrange's Theorem[62] (see Exercise 10.A below). Later in (1844) Cauchy returned to the subject to prove the following partial converse to Lagrange's Theorem.

Cauchy's Theorem. Let G be a finite group. (For Cauchy this was always a subgroup of the symmeteric group.) If p is any prime dividing the size of G then there exists an element $g \in G$ of order p, hence also a subgroup $\langle g \rangle \subseteq G$ of size p. ///

At the time this was the deepest theorem of group theory. It was later gen-

[61] Again, I will not define categories in this course. Please feel free to look up the definition on Wikipedia. For historical perspective please see Leo Corry's *Modern Algebra and the Rise of Mathematical Structures* (2004).

[62] This was the only paper cited by Abel in his (1826) proof of the unsolvability of the general quintic.

eralized by the Norwegian mathematician Ludwig Sylow (1872). We will not prove Cauchy's nor Sylow's theorems in this course (since they are not necessary for Galois Theory) but we will discuss some preliminary results and specific examples in Week 11.

The French school of group theory was systematized in the influential textbooks of Joseph Serret (1949) and Camille Jordan (1870). This tradition also directly inspired the work of Sophus Lie (a Norwegian) and Felix Klein (a German) on continuous transformation groups. In the twentieth century this developed into the modern theory of Lie groups, which is central to mathematical physics. In the following weeks we will give a selective treatment of these ideas, focusing mainly on the parts that are relevant to Galois theory.

MENTION FROBENIUS (Hawkins, page 335)

CONJUGACY CLASSES (Frobenius exploited orbit-stabilizer type ideas by thinking of lists with repeated elements)

ABSTRACT FORMULATION AND PROOF OF SYLOW THEOREMS BY CLASS EQUATION AND DOUBLE COSETS

8.2 Automorphisms and Group Actions

The goal of modern abstract algebra is 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 poset, a group/ring/field, or any kind of mathematical structure. By an *automorphism of* X we mean any invertible function $f : X \to X$ such that f and f^{-1} "preserve the structure of X". We denote the set of automorphisms by

$$\mathrm{Aut}(X) = \{\text{invertible } f : X \to X \text{ such that } f, f^{-1} \text{ preserve structure}\}.$$

It follows directly from the definition that $(\mathrm{Aut}(X), \circ, id)$ is a group under composition, with identity given by the identity function $id : X \to X$. ///

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

$$\mathrm{Aut}(X) = \mathrm{Perm}(X) = S_X.$$

[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 $X = \{1, 2, \ldots, n\}$, in which case we have

$$S_X = S_{\{1,2,\ldots,n\}} = S_n.$$

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 axiomatic definition of a group was given by Arthur Cayley in 1854: *On the Theory of Groups, as depending on the Symbolic Equation* $\theta^n = 1$, however this level of abstraction was not widely accepted until the twentieth century.[63] ///

Example: Matrices. Let V be a vector space over a field $\mathbb{F}$. (See Exercise 7.C for the definition.) Homomorphisms of vector spaces are called *linear functions* and the group of automorphisms of V is called the *general linear group* of V:

$$\mathrm{Aut}(\text{ vector space } V) = GL(V).$$

Now suppose that V has dimension n. Given a basis $\mathscr{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subseteq V$ we can represent each vector $\mathbf{x} \in V$ as an $n \times 1$ column by defining

$$[\mathbf{x}]_{\mathscr{U}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n \quad \Longleftrightarrow \quad \mathbf{x} = \sum_i x_i \mathbf{u}_i.$$

Then for each linear function $f : V \to V$ we define the $n \times n$ matrix $[f]_{\mathscr{U}} \in \mathrm{Mat}_n(\mathbb{F})$ whose j-th column is $[f(\mathbf{u}_j)]_{\mathscr{U}}$ and it follows from linearity that

$$[f(\mathbf{x})]_{\mathscr{U}} = [f]_{\mathscr{U}}[\mathbf{x}]_{\mathscr{U}} \quad \text{for all } \mathbf{x} \in V.$$

In summary, the basis $\mathscr{U}$ gives us an identification of the group $GL(V)$ of linear automorphisms with the group of $n \times n$ invertible matrices over $\mathbb{F}$:

$$\mathrm{Aut}(\text{ vector space } V \text{ with a fixed basis } \mathscr{U}) = GL_n(\mathbb{F})$$

However, we observe that there is no canonical choice of basis. If $\mathscr{U} \subseteq V$ and $\mathscr{V} \subseteq V$ are two bases for the vector space V and if $f : V \to V$ is a linear function then I claim that

$$C[f]_{\mathscr{U}}C^{-1} = [f]_{\mathscr{V}},$$

where $C \in \mathrm{Mat}_n(\mathbb{F})$ is the (unique, invertible) matrix satisfying $C[\mathbf{x}]_{\mathscr{U}} = [\mathbf{x}]_{\mathscr{V}}$ for all $x \in V$.

Proof. For all $\mathbf{x} \in V$ we have

$$(C[f]_{\mathscr{U}}C^{-1})[\mathbf{x}]_{\mathscr{V}} = C[f]_{\mathscr{U}}(C^{-1}[\mathbf{x}]_{\mathscr{V}})$$

[63] See Part III, Chapter 4 of Hans Wussing's *The Genesis of the Abstract Group Concept* (1984).

$$\begin{aligned}
&= C[f]_{\mathscr{U}}[\mathbf{x}]_{\mathscr{U}} \\
&= C([f]_{\mathscr{U}}[\mathbf{x}]_{\mathscr{U}}) \\
&= C[f(\mathbf{x})]_{\mathscr{U}} \\
&= [f(\mathbf{x})]_{\mathscr{V}} \\
&= [f]_{\mathscr{V}}[\mathbf{x}]_{\mathscr{V}}.
\end{aligned}$$

Then by substituting $\mathbf{x} = \mathbf{v}_j$ we see that the j-th columns of the matrices $C[f]_{\mathscr{U}}C^{-1}$ and $[f]_{\mathscr{V}}$ are equal for all j. □

In other words, we have shown that conjugate elements of the group $GL_n(\mathbb{F})$ represent the same linear function with respect to different bases. In Exercise 11.A 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[64]

$$\langle -, - \rangle : V \times V \to \mathbb{R}.$$

The group of automorphisms of Euclidean space is called the *orthogonal group*:

$$\text{Aut}(\text{Euclidean space } V) = O(V).$$

If $\mathscr{U} \subseteq V$ is an *orthonormal basis* (consisting of orthogonal unit vectors) then for each automorphism $f : V \to V$ one can show that $[f]_{\mathscr{U}} \in \text{Mat}_n(\mathbb{R})$ is an orthogonal matrix. (In fact you already showed this in Exercise 4.B.) Such a basis gives us an identification of the group $O(V)$ with the group $O(n)$ of $n \times n$ orthogonal matrices:

$$\text{Aut}(\text{ Euclidean space } V \text{ with a fixed orthonormal basis } \mathscr{U}) = O(n).$$

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[65] $\langle -, - \rangle : V \times V \to \mathbb{C}$ then all of the same remarks apply for the unitary groups $U(V)$ and $U(n)$. ///

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

$$S_n, \quad GL_n, \quad O(n), \quad U(n)$$

[64] "Symmetric" means that $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ for all $\mathbf{x}, \mathbf{y} \in V$ and "positive-definite" means that $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ for all $\mathbf{x} \in V$, with $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

[65] A "sesquilinear form" satisfies $\langle \mathbf{x} + \alpha\mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \alpha^* \langle \mathbf{y}, \mathbf{z} \rangle$ and $\langle \mathbf{x}, \mathbf{y} + \alpha\mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \alpha \langle \mathbf{x}, \mathbf{z} \rangle$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $\alpha \in \mathbb{C}$, where $\alpha^* \in \mathbb{C}$ is the complex conjugate of α.

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 $G \times X \to X$ 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:

$$\text{for all } g \in G \text{ the function } x \mapsto g(x) \text{ is in } \mathrm{Aut}(X).$$

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:

$$\varphi : G \to \mathrm{Aut}(X).$$

Proof. Given any function $(g, x) \mapsto g(x)$ satisfying (A1) and (A2) we will define $\varphi_g(x) \coloneqq g(x)$. By axiom (A2) the function φ_g is in $\mathrm{Aut}(X)$. 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 $\varphi : G \to \mathrm{Aut}(X)$ sending $g \in G$ to $\varphi_g : X \to X$ is a group homomorphism. Conversely, suppose we have a group homomorphism $\varphi : G \to \mathrm{Aut}(X)$ denoted by $\varphi \mapsto \varphi_g$. Now define a function $G \times X \to X$ by $(g, x) \mapsto \varphi_g(x)$ and observe that this function satisfies (A1) and (A2). □

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 $\varphi : G \to \mathrm{Aut}(X)$.

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", which 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 $\mathrm{Aut}(X) = \mathrm{Perm}(X)$, 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 $\varphi : G \to \mathrm{Aut}(X)$ then we can write
$$G \overset{\varphi}{\curvearrowright} X.$$
- **Jargon:** If V is a vector space, then an action $\varphi : G \to GL(V)$ 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*.

8.3 Translation and Conjugation

You won't appreciate the definition of group action until you understand some examples. In this lecture we will discuss two important examples of a group acting on itself, called "translation" and "conjugation".

Example: Translation. For all $g \in G$ we define $\tau_g : G \to G$ by
$$\tau_g(a) \coloneqq g * a \quad \text{for all } a \in G.$$
For each $g \in G$ I claim that the function τ_g is invertible, with $\tau_g^{-1} = \tau_{g^{-1}}$.

Proof. 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 $\tau : G \to \mathrm{Perm}(G)$ sending $g \mapsto \tau_g$. 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*.

Example: Conjugation. For all $g \in G$ we define $\kappa_g : G \to G$ by

$$\kappa_g(a) \coloneqq g * a * g^{-1} \quad \text{for all } a \in G.$$

For each $g \in G$ I claim that the function κ_g 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 $\kappa : G \to \mathrm{Perm}(G)$ sending $g \mapsto \kappa_g$. 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 $\mathrm{Aut}(G) \subseteq \mathrm{Perm}(G)$ of automorphisms, i.e., the subgroup of permutations that preserve the group structure.

Proof. For all $g, a, b \in G$ we have

$$\begin{aligned} \kappa_g(a) * \kappa_g(b) &= (g * a * g^{-1}) * (g * b * g^{-1}) \\ &= g * a * (g^{-1} * g) * b * g^{-1} \\ &= g * a * \varepsilon * b * g^{-1} \\ &= g * (a * b) * g^{-1} \\ &= \kappa_g(a * b), \end{aligned}$$

and hence $\kappa_g \in \mathrm{Aut}(G)$. □

Thus we obtain a group homomorphism $\kappa : G \to \mathrm{Aut}(G)$, 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 because $\tau_g(a * b) = g * a * b$ is not in general equal to $\tau_g(a) * \tau_g(b) = g * a * g * b$. In other words, the image of the homomorphism $\tau : G \to \mathrm{Perm}(G)$ is not contained in the subgroup $\mathrm{Aut}(G) \subseteq \mathrm{Perm}(G)$.
- 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**

$$\varphi : G \to \mathrm{Aut}(X).$$

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 $\tau : G \to \mathrm{Perm}(G)$? 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 $\tau_g : G \to G$ 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 $\mathrm{Perm}(G)$:

$$G = G/\ker \tau \cong \mathrm{im}\, \tau \subseteq \mathrm{Perm}(G).$$

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 groups is **not more general** than the concept of permutation groups. [**Remark:** However, the subgroup $\mathrm{im}\, \tau \subseteq \mathrm{Perm}(G)$ is certainly not equal to the full permutation group, because

$$\#\mathrm{im}\, \tau = \#G < (\#G)! = \#\mathrm{Perm}(G).]$$

Application: Definition of the Center and Inner Automorphisms. If we apply the First Isomorphism Theorem to the conjugation homomorphism $\kappa : G \to \mathrm{Aut}(G)$ then we obtain

$$G/\ker \kappa \cong \mathrm{im}\, \kappa \subseteq \mathrm{Aut}(G).$$

We have a special name for the kernel. It is called the *center* of G:[66]

$$\begin{aligned} Z(G) &:= \ker \kappa \\ &= \{g \in G : \kappa_g = id\} \end{aligned}$$

[66] Z is for *Zentrum*.

$$\begin{aligned}
&= \{g \in G : \kappa_g(a) = a \text{ for all } a \in G\} \\
&= \{g \in G : g * a * g^{-1} = a \text{ for all } a \in G\} \\
&= \{g \in G : g * a = a * g \text{ for all } a \in G\}.
\end{aligned}$$

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

$$\mathrm{Inn}(G) \coloneqq \operatorname{im} \kappa \subseteq \mathrm{Aut}(G).$$

It follows from the First Isomorphism Theorem that

$$\mathrm{Inn}(G) \cong G/Z(G).$$

I have nothing interesting to say about this right now.

Exercises

8.A Automorphisms of a Cyclic Group

For all integers $n \in \mathbb{Z}$ prove that

$$\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z}) \cong (\mathbb{Z}/n\mathbb{Z})^\times.$$

[Hint: Let $[k]_n \in \mathbb{Z}/n\mathbb{Z}$ denote the equivalence class of k mod n. Show that any automorphism $\varphi : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ has the form $\varphi_a([k]_n) \coloneqq [ak]_n$ for some integer $a \in \mathbb{Z}$ satisfying $\gcd(a, n) = 1$.]

8.B An Application of Conjugation

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.

8.C Why Does $AB = I$ Imply $BA = I$?

Given a field $\mathbb{F}$ and a positive integer n we define

$$\mathbb{M} \coloneqq \mathrm{Mat}_n(\mathbb{F}) = \text{the set of } n \times n \text{ matrices with entries in } \mathbb{F}.$$

I claim that this set is a *vector space of dimension* n^2 over the field $\mathbb{F}$. (This is because each matrix has n^2 independent entries. Never mind the details.) Now consider any two matrices $A, B \in \mathbb{M}$ such that $AB = I$.

(a) Show that the set $B\mathbb{M} \coloneqq \{BM : M \in \mathbb{M}\}$ 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.[67] 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 easiest proof I know. The same ideas can be used to show 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, matrix algebras) over a field.]

[67] You will prove this in Exercise 9.D.

Week 9

9.1 The Direct Product of Groups

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 the group-theoretic side of the theorem that the general quintic equation is not solvable. (We will just assume for now that Galois' Solvability Theorem from Week 2 is true. You will have to wait until the end of next semester if you want to see a proof of that.)

Theorem (Definition of External Direct Product). Let $(H, *, \varepsilon_H)$ and $(K, \bullet, \varepsilon_k)$ be abstract groups and consider the Cartesian product set

$$H \times K \coloneqq \{(h, k) : h \in H, k \in K\}.$$

I claim that the obvious definition

$$(h_1, k_1) \boxdot (h_2, k_2) \coloneqq (h_1 * h_2, k_1 \bullet k_2)$$

makes $H \times K$ into a group with identity element $(\varepsilon_H, \varepsilon_K)$. Furthermore, I claim that the sets $\tilde{H} \coloneqq \{(h, \varepsilon_K) : h \in H\}$ and $\tilde{K} \coloneqq \{(\varepsilon_H, k) : k \in K\}$ are subgroups of $H \times K$ with the following properties:

- We have group isomorphisms $\tilde{H} \cong H$ and $\tilde{K} \cong K$.
- The intersection is trivial: $\tilde{H} \cap \tilde{K} = \{(\varepsilon_H, \varepsilon_K)\}$.
- For all $\tilde{h} \in \tilde{H}$ and $\tilde{k} \in \tilde{K}$ we have $\tilde{h} \boxdot \tilde{k} = \tilde{k} \boxdot \tilde{h}$.

///

Proof. First we prove that $(H \times K, \boxdot, (\varepsilon_H, \varepsilon_K))$ is a group:

Associative. The operation $\boxdot$ inherits associativity from $*$ and $\bullet$.

Identity. For all $(h, k) \in H \times K$ we observe that

$$(h, k) \boxdot (\varepsilon_H, \varepsilon_K) = (\varepsilon_H, \varepsilon_K) \boxdot (h, k) = (h, k).$$

Inverse. For all $(h,k) \in H \times K$ we have

$$(h,k) \boxdot (h^{-1},k^{-1}) = (h^{-1},k^{-1}) \boxdot (h,k) = (\varepsilon_H, \varepsilon_K).$$

Now consider the subsets $\tilde{H}, \tilde{K} \subset H \times K$. We clearly have $\tilde{H} \cap \tilde{K} = \{(\varepsilon_H, \varepsilon_K)\}$. To see that $\tilde{H} \subseteq H \times K$ is a subgroup we observe that for any two elements (h_1, ε_K) and (h_2, ε_k) in $\tilde{H}$ we have

$$(h_1,\varepsilon_K) \boxdot (h_2,\varepsilon_K)^{-1} = (h_1,\varepsilon_K) \boxdot (h_2^{-1},\varepsilon_K) = (h_1 * h_2^{-1}, \varepsilon_K) \in \tilde{H}.$$

Then the map $H \to \tilde{H}$ defined by $h \mapsto (h, \varepsilon_K)$ is a bijective group homomorphism. A similar argument shows that $\tilde{K} \subseteq H \times K$ is a subgroup with $\tilde{K} \cong K$. Finally, we consider any two elements $\tilde{h} = (h, \varepsilon_K) \in \tilde{H}$ and $\tilde{k} = (\varepsilon_H, k) \in \tilde{K}$ and we observe that these elements commute:

$$\begin{aligned}
\tilde{h} \boxdot \tilde{k} &= (h,\varepsilon_K) \boxdot (\varepsilon_H, k) \\
&= (h * \varepsilon_H, \varepsilon_K \bullet k) \\
&= (h,k) \\
&= (\varepsilon_H * h, k \bullet \varepsilon_K) \\
&= (\varepsilon_H, k) \boxdot (h, \varepsilon_K) \\
&= \tilde{k} \boxdot \tilde{h}.
\end{aligned}$$

□

We call this construction the "external" direct product in order to emphasize the special case when H and K are subgroups of a common group G. In this case there is an obvious function from $H \times K$ into G.

Theorem (Definition of Internal Direct Product). Let $(G, *, \varepsilon)$ be a group and let $H, K \subseteq G$ be any two subgroups. We define the following "multiplication function"[68] from the external direct product:

$$\begin{array}{rccc}
\mu: & H \times K & \to & G \\
& (h,k) & \mapsto & h * k.
\end{array}$$

And we define the *product set* as the image of this function:[69]

$$HK \coloneqq \operatorname{im} \mu = \{h * k : h \in H, k \in K\}.$$

Then I claim that:

(1) The function μ is injective if and only if $H \cap K = \{\varepsilon\}$.

[68] In general this function is not a group homomorphism.

[69] Because μ is not necessarily a group homomorphism, the product set $HK \subseteq G$ is not necessarily a subgroup. You will see an example of this in Exercise **??**.

(2) The function μ is a group homomorphism if and only if for all $h \in H$ and $k \in K$ we have $h * k = k * h$.

If both of these properties hold then since μ is a group homomorphism we find that $HK = \operatorname{im}\mu \subseteq G$ is a sub**group** and since μ is injective (equivalently, $\ker\mu$ is trivial) we obtain a group isomorphism:

$$H \times K \cong \frac{H \times K}{\ker\mu} \cong \operatorname{im}\mu = HK.$$

In this case we will write $HK = H \times K$ and we will say that HK is the *internal direct product* of the subgroups $H \subseteq HK$ and $K \subseteq HK$. In the special case that $HK = G$ we will write $G = H \times K$. ///

Proof. (1) Note that for all $g \in H \cap K$ we have

$$\mu(g, g^{-1}) = g * g^{-1} = \varepsilon = \varepsilon * \varepsilon = \mu(\varepsilon, \varepsilon).$$

If μ is injective then it follows that $(g, g^{-1}) = (\varepsilon, \varepsilon)$ and hence $g = \varepsilon$. Conversely, let $H \cap K = \{\varepsilon\}$ and suppose that

$$h_1 * k_1 = \mu(h_1, k_1) = \mu(h_2, k_2) = h_2 * k_2$$

for some $h_1, h_2 \in H$ and $k_1, k_2 \in K$. Then we have $h_2^{-1} * h_1 = k_2 * k_1^{-1} \in H \cap K$. Since $H \cap K = \{\varepsilon\}$ this implies that $h_2^{-1} * h_1 = k_2 * k_1^{-1} = \varepsilon$, hence $h_1 = h_2$ and $k_1 = k_2$. In other words, μ is injective.

(2) Assume that $h * k = k * h$ for all $h \in H$ and $k \in K$. Then multiplication defines a homomoprhism $\mu : H \times K \to G$ because

$$\begin{aligned}
\mu((h_1, k_1) * (h_2, k_2)) &= \mu(h_1 * h_2, k_1 * k_2) \\
&= (h_1 * h_2) * (k_1 * k_2) \\
&= h_1 * (h_2 * k_1) * k_2 \\
&= h_1 * (k_1 * h_2) * k_2 \\
&= (h_1 * k_1) * (h_2 * k_2) \\
&= \mu(h_1, k_1) * \mu(h_2, k_2).
\end{aligned}$$

Conversely, let μ be a homomorphism. Then for all $(h, k) \in H \times K$ we have

$$\begin{aligned}
(h * k) * (h^{-1} * k^{-1}) &= \mu(h, k) * \mu(h^{-1}, k^{-1}) \\
&= \mu(h * h^{-1}, k * k^{-1}) \\
&= \mu(\varepsilon, \varepsilon) = \varepsilon
\end{aligned}$$

and it follows that

$$\begin{aligned}
h * k * h^{-1} * k^{-1} &= \varepsilon \\
h * k &= k * h.
\end{aligned}$$

□

The idea of a direct product is that the groups H and K don't even see each other and we can really think of HK as two independent groups sitting side by side. If $G = H \times K$ then we might say that G "factors" as a product of two normal subgroups.[70] If G is abelian then every subgroup is normal and the concept of the direct product simplifies.

Simplification: The Direct Sum of Abelian Groups. Let $(G, *, \varepsilon)$ be an abelian group and let $H, K \subseteq G$ be subgroups. Then since $h * k = k * h$ for all $h \in H$ and $k \in K$ we have

$$G = H \times K \quad \Longleftrightarrow \quad \left\{ \begin{array}{ccc} HK & = & G \\ H \cap K & = & \{\varepsilon\} \end{array} \right\}.$$

When the group $(G, +, 0)$ is expressed in additive language then we prefer to use the *direct sum* notation:[71]

$$G = H \oplus K \quad \Longleftrightarrow \quad \left\{ \begin{array}{ccc} H + K & = & G \\ H \cap K & = & \{0\} \end{array} \right\}.$$

This notation 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: The Chinese Remainder Theorem. Consider any integers $m, n \in \mathbb{Z}$ with $\gcd(m, n) = 1$. In Exercise 6.D you proved that the following map from the cyclic group $\mathbb{Z}/mn\mathbb{Z}$ to the external direct sum[72] $\mathbb{Z}/m\mathbb{Z} \oplus \mathbb{Z}/n\mathbb{Z}$ is a bijection:

$$\begin{array}{rccc} \varphi: & \mathbb{Z}/mn\mathbb{Z} & \leftrightarrow & \mathbb{Z}/m\mathbb{Z} \oplus \mathbb{Z}/n\mathbb{Z} \\ & [a]_{mn} & \mapsto & ([a]_m, [a]_n). \end{array}$$

In fact, I claim that this bijection is a group isomorhpism.

Proof. For all integers $a, b \in \mathbb{Z}$ we have

$$\begin{aligned} \varphi([a]_{mn} + [b]_{mn}) &= \varphi([a + b]_{mn}) \\ &= ([a + b]_m, [a + b]_n) \\ &= ([a]_m + [b]_m, [a]_n + [b]_n) \\ &= ([a]_m, [a]_n) \boxdot ([b]_m, [b]_n) \\ &= \varphi([a]_{mn}) \boxplus \varphi([b]_{mn}). \end{aligned}$$

[70] In the next section we will discuss *semidirect products*, in which only one of the factors is normal.

[71] This 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.

[72] I will use the symbol $\boxplus$ to denote the group operation of the external direct sum.

□

It follows that $\mathbb{Z}/mn\mathbb{Z}$ can be expressed as an internal direct sum of the preimages of $\{([a]_m, [0]_n) : a \in \mathbb{Z}\}$ and $\{([0]_m, [b]_n) : b \in \mathbb{Z}\}$, which are isomorphic to $\mathbb{Z}/m\mathbb{Z}$ and $\mathbb{Z}/n\mathbb{Z}$, respectively. Specifically, we observe that the preimage of $\{([a]_m, [0]_n) : a \in \mathbb{Z}\}$ is equal to $\{[nk]_{mn} : k \in \mathbb{Z}\}$ and the preimage of $\{[0]_m, [b]_n : b \in \mathbb{Z}\}$ is equal to $\{[m\ell]_{mn} : \ell \in \ell\}$.[73] Thus we obtain the following decomposition of $\mathbb{Z}/mn\mathbb{Z}$ as an internal direct sum:

$$\mathbb{Z}/mn\mathbb{Z} = \{[nk]_{mn} : k \in \mathbb{Z}\} \oplus \{[m\ell]_{mn} : \ell \in \mathbb{Z}\}.$$

And here is the smallest non-trivial example:

$$\mathbb{Z}/6\mathbb{Z} = \{[0]_6, [3]_6\} \oplus \{[0]_6, [2]_6, [4]_6\} \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}.$$

///

Let me end this section by mentioning two important applications of the direct sum concept. In the next lecture we will discuss the more subtle case of non-abelian groups.

Application: The Fundamental Theorem of Finite Abelian Groups. An important structural result going back to Gauss tells us that, in some sense, the theory of finite abelian groups is no more complicated than the theory of cyclic groups. Here is the precise statement:

every finite abelian group is a direct sum of cyclic subgroups.

The proof is surprisingly difficult. Perhaps the first modern statement of the theorem is due to Frobenius and Stickelberger (1878), who have the following to say:

> *The theory of finite groups of commuting elements was founded, on the one hand, by Euler and Gauss, and, on the other hand, by Lagrange and Abel, the former in their number-theoretic work on residues of powers, the latter in their algebraic work on the resolution of equations. After these foundational works, Gauss and Schering developed the theory further. Gauss ... showed how to decompose a group into primary groups, whose orders are relatively prime ... Schering its decomposition into cyclic groups whose orders are such that each is divisible by those that follow.*[74]

In Exercise 11.B you will explore Gauss' decomposition of a finite abelian group into primary factors, which can be viewed as a generalization of the Chinese

[73] In Exercise 6.D you proved that the inverse function can be expressed as $\varphi^{-1}([a]_m, [b]_n) = [any + bmx]_{mn}$ for any integers $x, y \in \mathbb{Z}$ such that $mx + ny = 1$. However, we don't really need to know this to compute the preimages.

[74] Quoted in Hawkins, *The mathematics of Frobenius in context* (2013, page 307).

Remainder Theorem. Schering's decomposition of each primary factor into cyclic groups is beyond the scope of this course.[75] ///

The other application comes from the theory of vector spaces.

Application: 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 sub**group** $U \subseteq V$ is a sub**space** 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 $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in V - \{\mathbf{0}\}$ is a basis for V if and only if

$$V = \mathbb{F}\mathbf{u}_1 \oplus \mathbb{F}\mathbf{u}_2 \oplus \cdots \oplus \mathbb{F}\mathbf{u}_n.$$

Proof. We define the direct sum of multiple groups by induction. The sum condition is easy:

$$V = \mathbb{F}\mathbf{u}_1 + \mathbb{F}\mathbf{u}_2 + \cdots + \mathbb{F}\mathbf{u}_n.$$

This equation literally says that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in V$ is a spanning set. The intersection condition is trickier. By induction we will require that

$$\mathbb{F}\mathbf{u}_i \cap (\mathbb{F}\mathbf{u}_1 + \cdots + \mathbb{F}\mathbf{u}_{i-1} + \mathbb{F}\mathbf{u}_{i+1} + \cdots + \mathbb{F}\mathbf{u}_n) = \{\mathbf{0}\} \quad \text{for all } i.$$

If $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in V$ is a linearly independent set then this condition is clearly satisfied. Conversely, suppose that $b_1\mathbf{u}_1 + \cdots + b_n\mathbf{u}_n = \mathbf{0}$ for some coefficients $b_i \in \mathbb{F}$. Then for each i we have

$$b_i\mathbf{u}_i = -b_1\mathbf{u}_1 - \cdots - b_{i-1}\mathbf{u}_{i-1} - b_{i+1}\mathbf{u}_{i+1} - \cdots - b_n\mathbf{u}_n.$$

If the intersection condition holds then this implies that $b_i\mathbf{u}_i = \mathbf{0}$ and hence $b_i = 0$. We conclude that the set $\mathbf{u}_1, \ldots, \mathbf{u}_n \in V$ is linearly independent. □

9.2 Semidirect Products of Groups

Last time we discussed direct products of abelian groups (i.e., direct sums). Today we will consider the case of non-abelian groups.

Non-Example of Direct Products: Dihedral Groups. Consider the dihedral group of size $2n$:

$$D_{2n} = \langle R, F \rangle = \{I, R, \ldots, R^{n-1}, F, RF, \ldots R^{n-1}F\}.$$

[75] The theorem is best viewed as an analogue of the Jordan Normal Form of a matrix, thus it fits more naturally into the theory of modules (i.e., abstract linear algebra). I consider that topic more suitable for a gradute course.

In Exercise 4.C you showed that every element has the form $R^a F^b$ for some integers $a, b \in \mathbb{Z}$, which implies that

$$D_{2n} = \langle R \rangle \langle F \rangle = \{hk : h \in \langle R \rangle, k \in \langle F \rangle\}.$$

Furthermore, I claim that $\langle R \rangle \cap \langle F \rangle = \{I\}$. In order to prove this we only need to show that F is not a power of R. The easiest way to see this is to consider the representation where R is a rotation matrix and F is a reflection matrix, so that $\det(R) = 1$ and $\det(F) = -1$. Then since $\det(R^a) = \det(R)^a = 1$ for all $a \in \mathbb{Z}$ we conclude that F is not a power of R. It follows from the two previous remarks that the multiplication function is bijective:

$$\begin{array}{ccc} \langle R \rangle \times \langle F \rangle & \longleftrightarrow & D_{2n} \\ (R^a, F^b) & \mapsto & R^a F^b. \end{array}$$

However, if $n \geq 3$ then this is **not** a direct product because the multiplication map is not a group homomorphism. Indeed, in this case the elements of $\langle R \rangle$ and $\langle F \rangle$ do not commute:

$$RF = FR^{-1} \neq FR.$$

///

We would like to relax our concept of group products to include examples such as $D_{2n} =$ "$\langle R \rangle \times \langle F \rangle$". The relevant fact here is that the subgroup $\langle R \rangle \subseteq D_{2n}$ is normal, but if $n \geq 3$ then the subgroup $\langle F \rangle \subseteq D_{2n}$ is not normal.

Proof. To see that $\langle R \rangle \trianglelefteq D_{2n}$ is a normal subgroup 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 us assume that $g = R^b F$ for some $b \in \mathbb{Z}$. Then we have

$$\begin{aligned} gR^a g^{-1} &= (R^b F) R^a (R^b F)^{-1} \\ &= R^b F R^a F R^{-b} \\ &= R^b \cancel{FF} R^{-a} R^{-b} \\ &= R^{-a} \in \langle R \rangle. \end{aligned}$$

Now suppose that $n \geq 3$. Then since $R^2 \neq I$ we observe that

$$RFR^{-1} = R(FR^{-1}) = R(RF) = R^2 F \notin \langle F \rangle,$$

and hence $\langle F \rangle \subseteq D_{2n}$ is not a normal subgroup. □

The concept of a "semi-direct product" is meant to capture cases like this, where exactly one of the factor groups is normal.

Theorem (Definition of Internal Semidirect Product). Let $(G, *, \varepsilon)$ be a group and let $H, K \subseteq G$ be subgroups satisfying $H \cap K = \{\varepsilon\}$. In this

case we have seen that the multiplication function defines a bijection from the direct product $H \times K$ onto the product set HK:

$$\begin{array}{ccc} H \times K & \longleftrightarrow & HK \\ (h,k) & \mapsto & h*k. \end{array}$$

If the elements of H and K do not commute then we know that the set $HK \subseteq G$ is not isomorphic to the direct product $H \times K$. However, it may still be a group of some sort.

(1) If **at least one** of the subgroups $H \subseteq G$ and $K \subseteq G$ is normal then I claim that the product set $HK \subseteq G$ is a subgroup. In this case we say that HK is a *semidirect product* of H and K and we write[76]

$$\begin{aligned} HK &= H \rtimes K \quad (\text{if } H \trianglelefteq G \text{ is normal}), \\ HK &= H \ltimes K \quad (\text{if } K \trianglelefteq G \text{ is normal}). \end{aligned}$$

(2) If **both** of $H \subseteq G$ and $K \subseteq G$ are normal then I claim that $h*k = k*h$ for all $h \in H$ and $k \in K$, hence we recover the direct product **direct product**:[77]

$$HK = H \times K \quad (\text{if } H \trianglelefteq G \text{ and } K \trianglelefteq G \text{ are normal})$$

///

Proof. (1) Suppose that $K \trianglelefteq G$ is normal and consider any two elements $h_1 * k_1$ and $h_2 * k_2$ of the product set HK. Since $H, K \subseteq G$ are a subgroups we know that $h_1 * h_2^{-1} \in H$ and $k_1 * k_2^{-1} \in K$. Then since K is closed under conjugation by elements of H (in fact, by all elements of G) we know that $h_2 * (k_1 * k_2^{-1}) * h_2^{-1} = k$ for some $k \in K$. It follows that

$$\begin{aligned} (h_1 * k_1) * (h_2 * k_2)^{-1} &= (h_1 * k_1) * (k_2^{-1} * h_2^{-1}) \\ &= h_1 * h_2^{-1} * (h_2 * k_1 * k_2^{-1} * h_2^{-1}) \\ &= h_1 * h_2^{-1} * k \in HK, \end{aligned}$$

and hence $HK \subseteq G$ is a subgroup. A similar proof works in the case when $H \trianglelefteq G$ is normal.

(2) Now suppose that $H \trianglelefteq G$ and $K \trianglelefteq G$ are both normal, with $H \cap K = \{\varepsilon\}$. Then for all $h \in H$ and $k \in K$ we have $h*k \in h^{-1} \in H$ and $k*h^{-1}*k^{-1} \in K$, so that

$$h*k*h^{-1}*k^{-1} = (h*k*h^{-1})*k^{-1} = h*(k*h^{-1}*k^{-1}) \in H \cap K.$$

[76] The triangles in the symbols $\rtimes$ and $\ltimes$ are supposed to remind us of the normal subgroup symbols $\trianglelefteq$ and $\trianglerighteq$.

[77] This result suggests that $\bowtie$ might be a good notation for the direct product, but absolutely no one uses it for this purpose. In fact, the notation $G = H \bowtie K$ is sometimes used when $G = HK$ and $H \cap K = \{\varepsilon\}$, but $H \subseteq G$ and $K \subseteq G$ are **both non-normal**. This terrible situation is called a *Zappa–Szép product* and I will have no more to say about it.

It follows that $h * k * h^{-1} * k^{-1} = \varepsilon$ and hence $h * k = k * h$. □

Remark: The situation is actually a bit more complicated than this theorem lets on. You will investigate the subtleties of subgroup multiplication in Exercise 9.A

I STOPPED HERE. FINISH THE DIHEDRAL EXAMPLE: $D_{2n} = \langle R \rangle \rtimes \langle F \rangle$. WHAT ABOUT THE CONCEPT OF AN "EXTERNAL SEMIDIRECT PRODUCT"? Let $H \trianglelefteq G$ and $H \cap K = \{\varepsilon\}$. Then the factorization is unique. What is it? Let $\theta : K \to \mathrm{Aut}(H)$.

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:

$$F R^a F^{-1} = R^{-a}.$$

The geometric meaning behind this is that flipping the polygon reverses the senses of clockwise and counterclockwise. ///

Example: Dihedral Groups. Consider the dihedral group of size $2n$:

$$D_{2n} = \langle R, F \rangle = \{I, R, \ldots, R^{n-1}, F, RF, \ldots R^{n-1}F\}.$$

in Exercise 4.C 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. Since $F^2 = I$ this completes the proof.

Thus we conclude that the multiplication map is a bijection:

$$\begin{array}{ccc} \langle R \rangle \times \langle F \rangle & \longleftrightarrow & D_{2n} \\ (R^a, F^b) & \mapsto & R^a F^b. \end{array}$$

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 $\langle R \rangle \trianglelefteq D_{2n}$ is normal we need to show that $gR^ag^{-1} \in \langle R \rangle$ for all $a \in \mathbb{Z}$. This is obvious when g is a power of R, so let's assume that $g = R^bF$ for some $b \in \mathbb{Z}$. Then we have

$$gR^ag^{-1} = (R^bF)R^a(R^bF)^{-1} = R^bFR^aFR^{-b} = R^b\cancel{FF}R^{-a}R^{-b} = R^{-a} \in \langle R \rangle.$$

□

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

$$(R^aF^b)(R^cF^d) = [R^aF^b(R^c)F^{-b}][F^bF^d] = \begin{cases} (R^aR^{-c})(F^bF^d) & b \text{ odd}, \\ (R^aR^c)(F^bF^d) & b \text{ even}. \end{cases}$$

The whole structure is determined by the fact F acts on $\langle R \rangle$ by inversion:

$$FR^aF^{-1} = R^{-a}.$$

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 $(H, *, \varepsilon_H)$ and $(K, \bullet, \varepsilon_K)$ 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 $\boxdot$, so that for all $h_1, h_2 \in H$ and $k_1, k_2 \in K$ we have

$$(h_1, k_1) \boxdot (h_2, k_2) = (h_3, k_3)$$

for some unique elements h_3 and k_3. We also want to require that the sub**sets**

$$\tilde{H} = \{(h, \varepsilon_K) : h \in H\} \subseteq H \times K$$
$$\tilde{K} = \{(\varepsilon_H, k) : k \in K\} \subseteq H \times K$$

are sub**groups** 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

  $$(h_1, k_1) \boxdot (h_2, k_2) = (h_1 * h_2, k_1 \bullet k_2).$$

  It is easy to check that $G \coloneqq (H \times K, \boxdot, (\varepsilon_H, \varepsilon_K))$ is an abstract group. Furthermore, I claim that G is the internal direct product $\tilde{H}$ and $\tilde{K}$.

  Proof. Clearly we have $\tilde{H} \cap \tilde{K} = \{(\varepsilon_H, \varepsilon_K)\}$ 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:

  $$(h, \varepsilon_K) \boxdot (\varepsilon_H, k) = (h, k) = (\varepsilon_H, k) \boxdot (h, \varepsilon_K) \quad \text{for all } h \in H \text{ and } k \in K.$$

  □

  The definition of the external direct product is so obvious[78] that we just use the Cartesian product notation:

  $$H \times K = (H \times K, \boxdot, (\varepsilon_H, \varepsilon_K)).$$

- *External Semidirect Product.* Suppose that the abstract group $(K, \bullet, \varepsilon_K)$ acts on the abstract group $(H, *, \varepsilon_H)$ by automorphisms. In other words, suppose that we have a group homomorphism

  $$\theta : K \to \mathrm{Aut}(H).$$

  Then we can define the operation

  $$(h_1, k_1) \boxdot_\theta (h_2, k_2) = (h_1 * \theta_{k_1}(h_2), k_1 \bullet k_2).$$

  It is relatively easy to check that $G \coloneqq (H \times K, \boxdot_\theta, (\varepsilon_H, \varepsilon_K))$ 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 $(h, \varepsilon_K) \in \tilde{H}$ and $(\varepsilon_H, k) \in \tilde{K}$ we have

  $$(\varepsilon_H, k) \boxdot_\theta (h, \varepsilon_K) \boxdot_\theta (\varepsilon_H, k)^{-1} = (\varepsilon_H, k) \boxdot_\theta (h, \varepsilon_K) \boxdot_\theta (\varepsilon_H, k^{-1})$$

[78] Another 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 $\times$.

$$
\begin{aligned}
&= (\varepsilon_H, k) \boxdot_\theta (h * \theta_{\varepsilon_K}(\varepsilon_H), k^{-1}) \\
&= (\varepsilon_H, k) \boxdot_\theta (h * \varepsilon_H, k^{-1}) \\
&= (\varepsilon_H, k) \boxdot_\theta (h, k^{-1}) \\
&= (\varepsilon_H * \theta_k(h), k \bullet k^{-1}) \\
&= (\theta_k(h), \varepsilon_K) \in \tilde{H}.
\end{aligned}
$$

□

In summary, given any action $\theta : K \to \mathrm{Aut}(H)$ 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

$$H \rtimes_\theta K = (H \times K, \boxdot_\theta, (\varepsilon_H, \varepsilon_K)).$$

///

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.
- What if H acts on K? For any homomorphism $\theta : H \to \mathrm{Aut}(K)$ we can define the following group structure on the set $H \times K$:

  $$(h_1, k_1) \, {}_\theta\boxdot (h_2, k_2) = (h_1 * h_2, \theta_{h_2}^{-1}(k_1) \bullet k_2).$$

  If we call this group $H \, {}_\theta\ltimes K$ then it turns out that

  $$H \, {}_\theta\ltimes K \cong K \rtimes_\theta H,$$

  so there is nothing new gained by this construction.
- Let $\mathrm{triv} : K \to \mathrm{Aut}(H)$ be the "trivial action" that sends each $k \in K$ to the identity function $\mathrm{triv}_k = id : H \to H$. Then the semidirect product coincides with the direct product:

  $$H \rtimes_{\mathrm{triv}} K = H \times K.$$

- 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 because of its importance to geometry and physics. We will see an interesting example next time.

9.3 Isometries of Euclidean Space

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 proved in Exercise 8.A that every automorphism of the group $\mathbb{Z}/n\mathbb{Z}$ has the form $k \mapsto ak \bmod n$ for some $a \in \mathbb{Z}$ satisfying $\gcd(a, n) = 1$. Now suppose we have a group homomorphism

$$\begin{array}{rccc} \theta: & \mathbb{Z}/2\mathbb{Z} & \to & \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z}) \\ & 0 & \mapsto & \theta_0 \\ & 1 & \mapsto & \theta_1 \end{array}$$

By definition the automorphisms $\theta_0, \theta_1 \in \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ must satisfy

$$\theta_{a+b \bmod 2} = \theta_{a \bmod 2} \circ \theta_{b \bmod 2}.$$

This implies that $\theta_0 = id$ and $\theta_1^2 = \theta_2 = \theta_0 = id$. On the other hand we know that $\theta_1(k) = ak$ for some $a \in \mathbb{Z}$ and since $\theta_1^2 = id$ 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 $\mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$.
- The nontrivial action $\theta : \mathbb{Z}/2\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ 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 $\langle -, - \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. By an *isometry* of Euclidean space we mean any function $f : \mathbb{R}^n \to \mathbb{R}^n$ 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

$$\mathrm{Isom}(\mathbb{R}^n) = \{f : \mathbb{R}^n \to \mathbb{R}^n : f \text{ preserves distance}\}.$$

This group has two interesting subgroups:

- Let $\mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n) \subseteq \mathrm{Isom}(\mathbb{R}^n)$ denote the subset of isometries that fix the origin: $f(\mathbf{0}) = \mathbf{0}$. We saw in Exercise 4.B that this subgroup is isomorphic to the group $O(n)$ of $n \times n$ orthogonal matrices. To be specific, for each $f \in \mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ 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 is 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

  $$\tau : \mathbb{R}^n \to \mathrm{Aut}(\mathbb{R}^n)$$

  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 $\tau_{\mathbf{u}} : \mathbb{R}^n \to \mathbb{R}^n$ 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 $\mathrm{Isom}(\mathbb{R}^n)$:

  $$\tau : (\mathbb{R}^n, +, \mathbf{0}) \to \mathrm{Isom}(\mathbb{R}^n).$$

  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 $\tau_{\mathbf{u}} : \mathbb{R}^n \to \mathbb{R}^n$ 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 $\mathrm{Isom}(\mathbb{R}^n)$ 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 $T(\mathbb{R}^n)$:

  $$\mathbb{R}^n \cong \mathrm{im}\,\tau =: T(\mathbb{R}^n) \subseteq \mathrm{Isom}(\mathbb{R}^n).$$

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:

$$\mathrm{Isom}(\mathbb{R}^n) = T(\mathbb{R}^n) \rtimes \mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n).$$

Proof. There are three things to check: (1) $T(\mathbb{R}^n)$ and $\mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ meet at the identity, (2) $T(\mathbb{R}^n)$ and $\mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ join to the full group, and (3) $T(\mathbb{R}^n)$ is a normal subgroup of $\mathrm{Isom}(\mathbb{R}^n)$.

(1) To show that $T(\mathbb{R}^n) \cap \mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n) = \{id\}$, suppose that $\tau_{\mathbf{u}}$ is a translation that fixes the origin. We saw above that this implies $\mathbf{u} = \mathbf{0}$ and hence $\tau_{\mathbf{u}} = \tau_{\mathbf{0}} = id$.

(2) To show that $T(\mathbb{R}^n) \circ \mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n) = \mathrm{Isom}(\mathbb{R}^n)$ consider any isometry $f : \mathbb{R}^n \to \mathbb{R}^n$ and suppose that $f(\mathbf{0}) = \mathbf{u}$. Now define $g \coloneqq \tau_{-\mathbf{u}} \circ f$ 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 $g \in \mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ and hence

$$f = \tau_{\mathbf{u}} \circ g \in T(\mathbb{R}^n) \circ \mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n).$$

(3) To show that $T(\mathbb{R}^n) \trianglelefteq \mathrm{Isom}(\mathbb{R}^n)$ it is enough[79] to show that $T(\mathbb{R}^n)$ is closed under conjugation by elements of $\mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n)$. So consider any $f \in \mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n)$. The important fact (which was tricky to prove)[80] 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 $\mathrm{Isom}(\mathbb{R}^n)$ 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 $f \in \mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ 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 $\theta : O(n) \to \mathrm{Aut}(\mathbb{R}^n)$ 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:

$$\theta_A(\mathbf{x}) \coloneqq A\mathbf{x} \quad \text{for all } A \in O(n) \text{ and } \mathbf{x} \in \mathbb{R}^n.$$

We can then form the external semidirect product

$$\mathbb{R}^n \rtimes_\theta O(n).$$

[79] This follows from part (2).
[80] See Exercise 4.B.

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 $V \to V$. (These do not need to preserve any structure.) Inside this 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:

$$GA(V) \coloneqq T(V) \circ GL(V) = T(V) \rtimes GL(V).$$

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).[81]

Exercises

9.A Subtleties of Subgroup Multiplication

Let $(G, *, \varepsilon)$ be a group and let $H, K \subseteq G$ be subgroups satisfying $H \cap K = \{\varepsilon\}$. Let $\mu : H \times K \to G$ be the multiplication function from the direct product and let $HK = \operatorname{im} \mu = \{h * k : h \in H, k \in K\}$ be the product set.

(a) If G is finite, prove that $\#(HK) = \#H \cdot \#K$. [Warning: You may not assume that $HK \subseteq G$ is a subgroup.]

(b) For any subgroup $L \subseteq G$ we define its *normalizer*:

$$N_G(L) \coloneqq \{g \in G : g * \ell * g^{-1} \in L\}.$$

Prove that $N_G(L)$ is the smallest subgroup of G that contains L as a normal subgroup.

(c) If $H \subseteq N_G(K)$ or $K \subseteq N_G(H)$, prove that $HK \subseteq G$ is a subgroup.

(d) Find specific groups $H, K \subseteq G$ such that $HK \subseteq G$ is a subgroup, but $H \not\subseteq N_G(K)$ and $K \not\subseteq N_G(H)$. [Hint: Let $G = S_4$ with $H = \langle (1234), (12)(34) \rangle$ and $K = \langle (123) \rangle$.]

(e) Prove that $HK \subseteq G$ is a subgroup if and only if $HK = KH$.

[81] Given 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". It turns out that $GA(V)$ is the group of automorphisms of this structure.

9.B Direct Product of Subgroups

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 $\mu : H \times K \to G$ 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.

9.C Matrix Representation of Isometries

Consider the following set of matrices:

$$G = \left\{ \left(\begin{array}{ccc|c} & A & & \mathbf{u} \\ \hline 0 & \cdots & 0 & 1 \end{array} \right) : A \in O(n) \text{ and } \mathbf{u} \in \mathbb{R}^n \right\} \subseteq \mathrm{Mat}_{n+1}(\mathbb{R}).$$

(a) Prove that $G \subseteq \mathrm{Mat}_{n+1}(\mathbb{R})$ is a subgroup. [Hint: Block multiplication.]

(b) Use results from class to prove that G is isomorphic to the group $\mathrm{Isom}(\mathbb{R}^n)$ of isometries of n-dimensional Euclidean space.

[**Remark:** An *affine function* $\mathbb{R}^n \to \mathbb{R}^n$ has the form $\mathbf{x} \mapsto A\mathbf{x} + \mathbf{u}$ for some matrix $A \in \mathrm{Mat}_n(\mathbb{R})$ and some vector $\mathbf{u}$. The same trick can be used to represent affine functions as $(n+1) \times (n+1)$ matrices.]

9.D Dimension of a Vector Space, Part II

Let V be a vector space over a field $\mathbb{F}$.

(a) Let $\mathbf{u}_1, \ldots, \mathbf{u}_n \in V$ be a basis and consider the subspaces

$$V_k := \mathbb{F}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \subseteq V.$$

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

$$\{\mathbf{0}\} = V_0 \subsetneq V_1 \subsetneq \cdots \subsetneq V_n = V.$$

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

$$U = V_0 \subsetneq V_1 \subsetneq \cdots \subsetneq V_m = V.$$

[Hint: You may assume that the Correspondence Theorem and the First Isomorphism Theorem still hold after replacing the word "subgroup" with "subspace".[82]]

(d) Combine (a), (b) and (c) to prove that $\dim(V) = \dim(U) + \dim(V/U)$.

(e) *Rank-Nullity Theorem.* If $\varphi : V \to W$ is any linear function, use part (d) and the First Isomorphism Theorem to prove that

$$\dim(V) = \dim(\ker \varphi) + \dim(\operatorname{im} \varphi).$$

[**Remark:** In elementary linear algebra the subspaces $\ker \varphi \subseteq V$ and $\operatorname{im} \varphi \subseteq W$ are called the *nullspace* and the *range* of the linear function φ. The dimensions $\dim(\ker \varphi)$ and $\dim(\operatorname{im} \varphi)$ 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

$$\begin{aligned} \dim(\operatorname{im} \varphi) &= \#(\text{pivot columns in RREF of } [\varphi]) \\ \dim(\ker \varphi) &= \#(\text{non-pivot columns in RREF of } [\varphi]), \end{aligned}$$

and hence

$$\dim(\operatorname{im} \varphi) + \dim(\ker \varphi) = \#(\text{columns in } [\varphi]) = \dim(V).$$

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

$$\begin{aligned} (\varphi \text{ is injective}) &\iff \ker \varphi = \{\mathbf{0}\} \\ &\iff \dim(\ker \varphi) = 0 \\ &\iff \dim(\operatorname{im} \varphi) = \dim(V) \\ &\iff \dim(\operatorname{im} \varphi) = \dim(W) \\ &\iff \operatorname{im} \varphi = W \\ &\iff (\varphi \text{ is surjective}). \end{aligned}$$

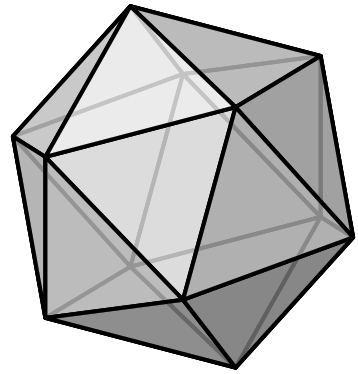
This can be used to show that that $AB = I$ implies $BA = I$ for any square matrices A, B over a field. Of course, we already had a slightly easier proof of this fact in Exercise 8.C.]

[82] For that matter, the Second and Third Isomorphism Theorems also hold.

Week 10

10.1 Unsolvability of the Symmetric Group

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:



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:

$$\{id\} \subsetneq H \trianglelefteq I \quad \Longrightarrow \quad H = I.$$

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-nontrivial 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 $\mathscr{L}(\mathbb{Z}/n\mathbb{Z})$ is isomorphic to the lattice of divisors $\mathrm{Div}(n)$. 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. □

[**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 group[83] 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 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

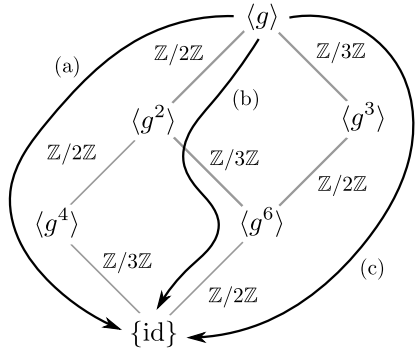
$$G_1/G_0, \quad G_2/G_1, \quad \cdots \quad G_{n-1}/G_\ell, \quad G_\ell/\{\varepsilon\} = G_\ell$$

is **unique** up to isomorphism and permutations. ///

[83] Not literally every group, but it's true for all finite groups and many infinite groups.

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. Each edge in the diagram is labeled 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

$$\mathbb{Z}/p\mathbb{Z}, \quad \mathbb{Z}/p\mathbb{Z}, \quad \cdots, \quad \mathbb{Z}/p\mathbb{Z} \quad (k \text{ times})$$

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 $p, p, \ldots, p$ (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:

$$\left\{ \begin{array}{c} G \text{ has abelian} \\ \text{composition factors} \end{array} \right\} \iff \left\{ \begin{array}{c} G \text{ has composition factors of the form} \\ \mathbb{Z}/p\mathbb{Z} \text{ for various prime numbers } p \end{array} \right\}.$$

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' Solvability 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 for all $n \geq 5$ the group S_n is **not solvable**. 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

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

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 $C \subseteq \{id\} = G_r$, 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

$$\begin{aligned}(c_1c_2c_1^{-1}c_2^{-1}G_{i+1}) &= (c_1G_{i+1})(c_2G_{i+1})(c_1G_{i+1})^{-1}(c_2G_{i+1})^{-1} \\ &= (c_1G_{i+1})(c_1G_{i+1})^{-1}(c_2G_{i+1})(c_2G_{i+1})^{-1} \\ &= (id\ G_{i+1})(id\ G_{i+1}) \\ &= id\ G_{i+1} \\ &= G_{i+1},\end{aligned}$$

which implies that $c_1c_2c_1^{-1}c_2^{-1} \in G_{i+1}$. 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 $c = c_1c_2c_1^{-1}c_2^{-1}$ 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

$$(ijk) = (jkm)(i\ell j)(jkm)^{-1}(i\ell j)^{-1}.$$

[**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 $A_5, A_6, A_7, \ldots$ 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.

10.2 The Orbit-Stabilizer Theorem

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". I 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 $\varphi : G \to \mathrm{Aut}(X)$ be a group homomorphism (i.e., an action of G on X). Then for any point $x \in X$ we define the following sets:

$$\mathrm{Orb}_\varphi(x) \coloneqq \{\varphi_g(x) : g \in G\} \subseteq X,$$
$$\mathrm{Stab}_\varphi(x) \coloneqq \{g \in G : \varphi_g(x) = x\} \subseteq G.$$

When the specific action φ is understood we will just write $\mathrm{Orb}(x)$ and $\mathrm{Stab}(x)$. We can also view orbits as the equivalence classes of the following relation:

$$x \sim_\varphi y \quad \iff \quad \exists g \in G, \varphi_g(x) = y.$$

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:

$$X = \coprod_i \mathrm{Orb}(x_i) \quad \text{for some arbitrary class representatives } x_i \in X.$$

///

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 $a, b \in \mathrm{Stab}(x)$ 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 $a * b^{-1} \in \mathrm{Stab}(x)$. □

Unfortunately the subgroup $\mathrm{Stab}(x) \subseteq G$ 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 $\varphi : G \to \mathrm{Aut}(X)$ 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:

$$\begin{array}{rccc} \Phi: & \mathrm{Orb}(x) & \longrightarrow & G/\mathrm{Stab}(x) \\ & \varphi_g(x) & \mapsto & g\,\mathrm{Stab}(x). \end{array}$$

Proof. The function Φ is well-defined and injective because

$$\begin{aligned} \varphi_a(x) = \varphi_b(x) &\iff \varphi_b^{-1}(\varphi_a(x)) = x \\ &\iff x = \varphi_{b^{-1}*a}(x) \\ &\iff b^{-1} * a \in \mathrm{Stab}(x) \\ &\iff a\,\mathrm{Stab}(x) = b\,\mathrm{Stab}(x), \end{aligned}$$

and it is surjective by definition. □

It follows that X can be identified with a disjoint union of sets of cosets:

$$X = \coprod_i \mathrm{Orb}(x_i) \quad \longleftrightarrow \quad \coprod_i G/\operatorname{Stab}(x_i).$$

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 $\Phi : \mathrm{Orb}(x) \to G/\operatorname{Stab}(x)$ "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 $\mathrm{Orb}(f) = F$. (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

$$\begin{aligned} \mathrm{Orb}(f) = F &\leftrightarrow I/\operatorname{Stab}(f) \\ \#F &= \#I/\#\operatorname{Stab}(f) \\ 20 &= \#I/3 \\ \#I &= 60. \end{aligned}$$

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 $\operatorname{Stab}(e)$ is a (cyclic) group of size 2, and for each vertex v the stabilizer $\operatorname{Stab}(v)$ is a cyclic group of size 5. (Look at the picture above.) Thus we obtain two more equations

$$\begin{aligned} \#\mathrm{Orb}(e) &= \#I/\#\operatorname{Stab}(e), & \#\mathrm{Orb}(v) &= \#I/\#\operatorname{Stab}(v), \\ \#E &= \#I/2, & \#V &= \#I/5. \end{aligned}$$

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. ///

10.3 Klein's Erlangen Program

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 $\mathbb{Z}/n\mathbb{Z}$ in his *Disquisitiones Arithmeticae* (1801). However, as you know by now, the study of finite abelian groups is only a tiny part of the subject. The main concepts of the theory were only revealed with Galois' and Cauchy's work on the (non-abelian) symmetric group S_n. Galois developed his theory on chains of subgroups just before his death in 1832, but this work was only published in 1846. During this same time, Cauchy proved some deep results on permutations, which contain the germs of important group-theoretical concepts such as the Orbit-Stabilizer Theorem and the Sylow Theorems. 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 geometries[84] 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 $\mathrm{Aut}(X)$. 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. Here is the key definition:

Definition of Transitive Actions. Let G be a group and let X be a set with structure. We say that an action $\varphi : G \to \mathrm{Aut}(X)$ is *transitive* if it has only one orbit. In other words:

$$\mathrm{Orb}_\varphi(x) = X \quad \text{for each point } x \in X.$$

Then from the Orbit-Stabilizer Theorem we obtain a bijection

$$X \quad \longleftrightarrow \quad G/\,\mathrm{Stab}_\varphi(x) \quad \text{for each point } x \in X.$$

If X is a finite set then it follows from Lagrange Theorem that $\#X$ divides $\#G$. ///

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 $f : X \to X$ that preserve the geometric structure. An essential feature of such a geometry is that any two points 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

[84] The existence of logically consistent non-Euclidean geometries was asserted independently by János Bolyai and Nikolai Lobachevky around 1830. Gauss may also have discovered non-Euclidean geometry but he published nothing on the subject.

for various $H \subseteq G$. Today we use the term *homogeneous space* instead of *non-Euclidean geometry*. After the 1870s 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 $\mathrm{Isom}(\mathbb{R}^n)$ be the group if isometries, i.e., functions $f : X \to X$ that preserve distance. Clearly the action of $\mathrm{Isom}(\mathbb{R}^n)$ on X is transitive. (Indeed, for any points $\mathbf{x}, \mathbf{y} \in X$ the translation $\tau_{\mathbf{y}-\mathbf{x}} \in \mathrm{Isom}(\mathbb{R}^n)$ sends $\mathbf{x}$ to $\mathbf{y}$.) Thus for any point $\mathbf{x} \in X$ we obtain a bijection

$$X \quad \longleftrightarrow \quad \mathrm{Isom}(\mathbb{R}^n)/\mathrm{Isom}_{\mathbf{x}}(\mathbb{R}^n),$$

where $\mathrm{Isom}_{\mathbf{x}}(\mathbb{R}^n) \coloneqq \{f \in \mathrm{Isom}(\mathbb{R}^n) : f(\mathbf{x}) = \mathbf{x}\}$ is the stabilizer of $\mathbf{x}$. In particular, we already know that $\mathrm{Isom}_{\mathbf{0}}(\mathbb{R}^n)$ is isomorphic to the orthogonal group. Hence we obtain a bijection:

$$\mathbb{R}^n \quad \longleftrightarrow \quad \mathrm{Isom}(\mathbb{R}^n)/O(n).$$

But this is not so interesting because it follows from the group isomorphism

$$\mathrm{Isom}(\mathbb{R}^n) \cong (\mathbb{R}^n, +, \mathbf{0}) \rtimes O(n)$$

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 are 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 $(n-1)$-*dimensional projective space* as 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 $\mathbf{0} \in \mathbb{R}^n$, 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 $\mathrm{Stab}(\ell) \cong O(1) \times O(n-1)$, hence we obtain a bijection

$$\mathbb{P}^{n-1}(\mathbb{R}) \quad \longleftrightarrow \quad \frac{O(n)}{O(1) \times O(n-1)}.$$

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

$$U^\perp \coloneqq \{\mathbf{v} \in \mathbb{R}^n : \langle \mathbf{u}, \mathbf{v} \rangle = 0 \text{ for all } \mathbf{u} \in U\}.$$

Then for any orthogonal matrix $A \in O(n)$ I claim that

$$A \text{ stabilizes } U \quad \Longleftrightarrow \quad A \text{ stabilizes } U^\perp.$$

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

$$\langle \mathbf{u}, A^{-1}\mathbf{v} \rangle = \langle \mathbf{u}, A^T\mathbf{v} \rangle = \mathbf{u}^T(A^T\mathbf{v}) = (A\mathbf{u})^T\mathbf{v} = \langle A\mathbf{u}, \mathbf{v} \rangle = 0.$$

If 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 $A^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ is injective, it follows from the Rank-Nullity Theorem (proved in Exercise 9.D) 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. □

[**Remark:** Note that the proof uses finite-dimensionality. The situation is more complicated for infinite-dimensional spaces.]

If $U \subseteq \mathbb{R}^n$ is a k-dimensional subspace then we can choose an orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_n \in V$ such that $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is a basis for U and $\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n$ 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 $A \in \mathrm{Stab}(U) \subseteq O(n)$ 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 $\mathrm{Stab}(U) \cong O(k) \times O(n-k)$ and we obtain a bijection

$$\left\{ \begin{array}{c} k\text{-dimensional} \\ \text{subspaces of } \mathbb{R}^n \end{array} \right\} \quad \longleftrightarrow \quad \frac{O(n)}{O(k) \times O(n-k)}.$$

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 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:

$$\left\{ \begin{array}{c} \text{subsets of size } k \text{ from} \\ \text{the set } \{1, \ldots, n\} \end{array} \right\} \quad \longleftrightarrow \quad \frac{S_n}{S_k \times S_{n-k}}.$$

  Observe 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 $\varphi : G \to \mathrm{Aut}(X)$ is *free* if each stabilizer is trivial:

$$\mathrm{Stab}_\varphi(x) = \{\varepsilon\} \quad \text{for each point } x \in X.$$

In this case, every orbit is in bijection with G:

$$\mathrm{Orb}_\varphi(x) \quad \longleftrightarrow \quad G/\{\varepsilon\} = G.$$

If G is a finite group then since X is a disjoint union of orbits, each of size $\#G$, it follows that X is finite and that $\#G$ divides $\#X$.

If in addition there is only one orbit (i.e., if the action is also transitive) then we say that the action is *regular*.[85] In this case, each point $x \in X$ defines a bijection between X and G:

$$\begin{array}{ccc} X & \longleftrightarrow & G \\ \varphi_g(x) & \longleftrightarrow & g. \end{array}$$

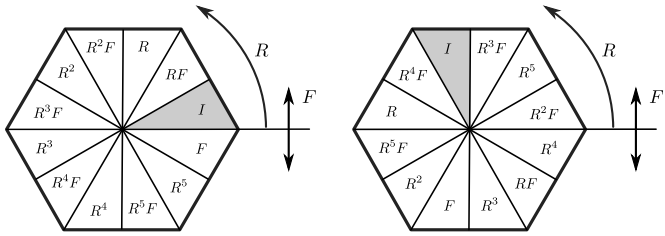
However, since there is no "best" point of X, none of these bijections is "best". ///

Example: Dihedral and Cyclic Groups. The dihedral group D_{2n} acts by symmetries on a regular n-sided polygon. This induces transitive actions of

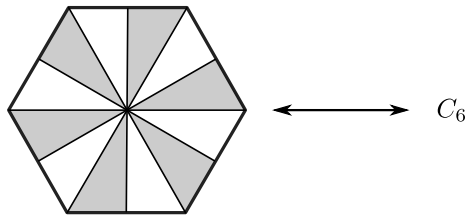
[85] I don't like this word too much. I'll probably just say *free and transitive*. The term *simply-transitive* is also common. The fanciest way to describe a regular action $G \curvearrowright X$ is to say that X is a *G-torsor*.

D_{2n} on the set of vertices and on the set of edges of the polygon. But is there some set of objects on which the group acts freely?

Answer: Divide the polygon into n isoceles 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 $n = 6$. 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?

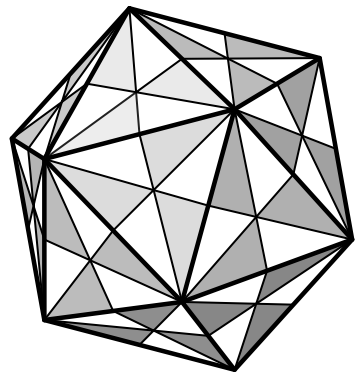
Answer: Let X be a set and let $\varphi : (\mathbb{R}^n, +, \mathbf{0}) \to \mathrm{Perm}(X)$ be a regular action. Then for any arbitrary basepoint $x \in X$ we obtain a bijection $X \leftrightarrow \mathbb{R}^n$ identifying $x \in X$ with $\mathbf{0} \in \mathbb{R}^n$. The pair (X, φ) is called *affine n-dimensional space*. Affine space is a better model for the real world because it has no origin.

///

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 we shade alternating triangles, as in the following picture:



The group I acts freely and transitively on the set of shaded triangles. If we choose an arbitrary triangle to play the role of the identity element then we obtain a bijection

$$\{\text{shaded triangles}\} \quad \longleftrightarrow \quad I.$$

And what about the 60 unshaded triangles? 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 $\mathrm{Sym}(X) \subseteq O(n)$ and an "alternating subgroup" $\mathrm{Alt}(X) = \mathrm{Sym}(X) \cap SO(n)$ consisting of "rotational symmetries".]

Exercises

10.A Lagrange's Version of Lagrange's Theorem

The original prototype of "Lagrange's Theorem" was stated by Lagrange without proof in Article 96 of his *Reflexions sur la résolution algébrique des équa-*

tions (1770). The first proof was given by Cauchy in (1815).[86] Here is the statement in modern language:

> Let X, Y be sets and let $f : X^n \to Y$ be a function with n inputs. Consider a given list of inputs: $x_1, \ldots, x_n \in X$. If the inputs are permuted in all ways then the number of different outputs is a divisor of $n!$.

Use Orbit-Stabilizer and the modern version of Lagrange's Theorem to prove this.

10.B 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 $\varphi_{(h,k)}(g) \coloneqq hgk^{-1}$.

(a) Prove that this defines a group homomorphism $\varphi : H \times K \to \mathrm{Perm}(G)$.

(b) For each $g \in G$, prove that the orbit satisfies

$$\mathrm{Orb}_\varphi(g) = HgK \coloneqq \{hgk : h \in H, k \in K\}.$$

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 $\psi : H \to \mathrm{Perm}(G/K)$ defined by $\psi_h(gK) \coloneqq (hg)K$. (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:

$$HgK = \coprod_{C \,\in\, \mathrm{Orb}_\psi(gK)} C.$$

(d) For all $g \in G$ prove that $\mathrm{Stab}_\psi(gK) = H \cap gKg^{-1}$, where $gKg^{-1} \coloneqq \{gkg^{-1} : k \in K\}$.

(e) Combine (c) and (d) with Lagrange's Theorem and Orbit-Stabilizer to conclude that

$$\#HgK = \frac{\#H \cdot \#K}{\#(H \cap gKg^{-1})}.$$

[**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.]

[86] Augustin-Louis Cauchy, *Mémoire sur les arrangements que l'on peut former avec des lettres données et sur les permutations ou substitutions à l'aide desquelles on passe d'un arrangement à un autre* (1815).

10.C Burnside's Lemma

Let $\varphi : G \to \mathrm{Perm}(X)$ be a group action, and let X/G denote the set of orbits. For each $g \in G$, let $\mathrm{Fix}_\varphi(g)$ denote the set of elements fixed by g:

$$\mathrm{Fix}_\varphi(g) \coloneqq \{x \in X : \varphi_g(x) = x\} \subseteq X.$$

(a) Count the elements of the set $\{(g, x) \in G \times X : \varphi_g(x) = x\}$ in two ways to prove that

$$\sum_{g \in G} \#\mathrm{Fix}_\varphi(g) = \sum_{x \in X} \#\,\mathrm{Stab}_\varphi(x).$$

(b) Use Orbit-Stabilizer to obtain a formula for the number of orbits:

$$\#(X/G) = \frac{1}{\#G} \sum_{g \in G} \#\mathrm{Fix}_\varphi(g).$$

(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 $\#\mathrm{Fix}(R) = k$ and $\#\mathrm{Fix}(R^2) = k^2$.]

10.D Lagrange vs. Rank-Nullity

Let $p \in \mathbb{Z}$ be prime. You showed on the previous homework that every nonzero element of the ring $\mathbb{F}_p \coloneqq \mathbb{Z}/p\mathbb{Z}$ 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.

Week 11

11.1 Conjugacy Classes

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 $\tau_g : G \to G$ by $\tau_g(a) \coloneqq ga$. Then one can show that the map $g \mapsto \tau_g$ defines a group homomorphism
$$\tau : G \to \mathrm{Perm}(G).$$
- *Conjugation.* For any $g \in G$ we define the function $\kappa_g : G \to G$ by $\kappa_g(a) \coloneqq gag^{-1}$. Then one can show that the map $g \mapsto \kappa_g$ defines a group homomorphism
$$\kappa : G \to \mathrm{Aut}(G).$$

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 $\mathrm{im}\,\tau \subseteq \mathrm{Perm}(G)$. [**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.[87]

Proof.

- *Free.* For all $a \in G$ we want to show that $\mathrm{Stab}_\tau(a) \subseteq G$ is the trivial group. So consider any elememnt $g \in \mathrm{Stab}_\tau(a)$. By definition we have $a = \tau_g(a) = ga$, and multiplying by a^{-1} on the right gives $g = \varepsilon$. We conclude that $\mathrm{Stab}_\tau(a) = \{\varepsilon\}$ 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
$$\tau_g(a) = \tau_{ba^{-1}}(a) = (ba^{-1})a = b.$$

[87] We already proved this for the abelian group $G = (\mathbb{R}^n, +, \mathbf{0})$. Now we'll show that it holds in general.

□

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 $\mathrm{Orb}_\tau(a) = G$ and $G/\operatorname{Stab}_\tau(a) = G/\{\varepsilon\} = G$. Thus we obtain a bijection from G to itself:

$$\begin{array}{ccc} G = \mathrm{Orb}_\tau(a) & \longleftrightarrow & G/\operatorname{Stab}_\tau(a) = G \\ ga & \longleftrightarrow & g. \end{array}$$

Note that we can explicitly describe the bijection $G = G/\operatorname{Stab}_\tau(a) \to \mathrm{Orb}_\tau(a) = G$ as **multiplication on the right by** a**.** It's a bit interesting that multiplication on the right 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:

$$\begin{aligned} Z(G) \coloneqq \ker \kappa &= \{g \in G : \kappa_g = id\} \\ &= \{g \in G : \kappa_g(a) = a \text{ for all } a \in G\} \\ &= \{g \in G : gag^{-1} = a \text{ for all } a \in G\} \\ &= \{g \in G : ga = ag \text{ for all } a \in G\}. \end{aligned}$$

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*]:
$$K(a) \coloneqq \mathrm{Orb}_\kappa(a) = \{gag^{-1} : g \in G\}.$$
- *Centralizers.* For all $a \in G$ we define the *centralizer* [Z is for *Zentrum* again]:
$$Z(a) \coloneqq \operatorname{Stab}_\kappa(a) = \{g \in G : gag^{-1} = a\} = \{g \in G : ga = ag\}.$$

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:

$$\begin{array}{ccc} K(a) & \longleftrightarrow & G/Z(a) \\ gag^{-1} & \longleftrightarrow & gZ(a). \end{array}$$

We will combine these facts to obtain a useful formula. First, observe for all $a \in G$ that

$$K(a) = \{a\} \quad \Longleftrightarrow \quad Z(a) = G \quad \Longleftrightarrow \quad a \in Z(G).$$

This suggests that we should collect the singleton conjugacy classes together. If $a_1, a_2, \ldots, a_k \in G$ is an arbitrary system of conjugacy class representatives, then we obtain a disjoint union:

$$\begin{aligned} G = \coprod K(a_i) &= \coprod_{K(a_i)=\{a_i\}} \{a_i\} \bigcup \coprod_{K(a_i)\neq\{a_i\}} K(a_i) \\ &= \coprod_{a_i \in Z(G)} \{a_i\} \bigcup \coprod_{K(a_i)\neq\{a_i\}} K(a_i) \\ &= Z(G) \bigcup \coprod_{K(a_i)\neq\{a_i\}} K(a_i). \end{aligned}$$

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. ///

11.2 The Sylow Theorems

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. ACTUALLY I WILL PROVE SYLOW I AND LEAVE II AND III FOR THE EXERCISES.

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 $(gZ(G))^k = g^k Z(G)$ for some $k \in \mathbb{Z}$. Then since the cosets cover G it follows that every element of G has the form $g^k z$ 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

$$\begin{array}{rcl} \mu : \langle g\rangle \times \langle h\rangle & \to & G \\ (g^k, h^\ell) & \mapsto & g^k h^\ell. \end{array}$$

Since $\langle g\rangle \cap \langle h\rangle = \{\varepsilon\}$ we know that μ is injective, hence the image $\langle g\rangle\langle h\rangle = \operatorname{im}\mu \subseteq G$ 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 decompose the group 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 easiest 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

$$P(k) \sim \frac{1}{4k\sqrt{3}} \cdot \exp\left(\pi\sqrt{\frac{2k}{3}}\right) \quad \text{as} \quad k \to \infty.$$

  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 in Exercise **??**, assuming that the famous "Sylow Theorems" are true.

The following theorems were originally proved by Ludwig Sylow (1872) in the context of permutations. The results were given new proofs and generalized to abstract groups by Frobenius (1887). One can view part (1) as a partial converse to Lagrange's Theorem.

The Sylow Theorems. 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 p^e for each $1 \leq e \leq k$. The subgroups of size p^k are called *Sylow p-subgroups.*

(2) Any p-subgroup is contained in a Sylow p-subgroup. In fact, for any subgroups $H, S \subseteq G$ where $\#S = p^k$ and $\#H = p^e \leq p^k$, there exists an element $g \in G$ such that $H \subseteq gSg^{-1}$.

(3) Let n_p be the number of Sylow p-subgroups. Then we have $n_p | m$ and $n_p = 1 \bmod p$. If $n_p = 1$ then by part (2) we conclude that the unique Sylow p-subgroup is normal.

///

I will give here an elegant proof of part (1) due to Helmut Wielandt (1959). You will supply the proofs of parts (2) and (3) in Exercise 11.C.

Proof of (1).

Counting Lemma: Let $\gcd(p, m)$ and $1 \leq e \leq k$. Then p^{k-e} is the highest power of p that divides the binomial coefficient $\binom{p^k m}{p^e}$. Indeed, consider the following expression:

$$\binom{p^k m}{p^e} = p^{k-e} m \cdot \frac{(p^k m - 1) \cdots (p^k m - \ell) \cdots (p^k m - p^e + 1)}{(p^e - 1) \cdots (p^e - \ell) \cdots 1}.$$

If $p^\alpha \| (p^e - \ell)$ for some $\ell \neq 0$ then we must have $p^\alpha \| \ell$. But then we also have $p^\alpha \| (p^e m - \ell)$, so the fraction on the right represents an integer that is not divisible by p. ///

Let $X := \{S \subseteq G : \#S = p^e\}$ be the set of all sub**sets** of size p^e. We want to prove that at least one of these subsets is a sub**group**. To do this we let G act on X by left multiplication and consider the partition into orbits

$$X = \coprod_i \mathrm{Orb}(S_i)$$

for some arbitrarily chosen subsets $S_i \subseteq G$.

I claim that there exists some i such that $p^{k-e+1} \nmid \mathrm{Orb}(S_i)$. Indeed, if we had $p^{k-e+1} \mid \mathrm{Orb}(S_i)$ for all i then from the equation $\#X = \sum_i \#\mathrm{Orb}(S_i)$ we would conclude that $p^{k-e+1} | \#X$, which contradicts the above lemma. Now consider the stabilizer subgroup $H := \mathrm{Stab}(S_i) \subseteq G$. By definition this means that the set S_i is closed under left multiplication by H. In other words, S_i is a union of left H-cosets. It follows from this that $\#H$ divides $\#S = p^e$, so $\#H$ is a power of p. Finally, we observe from the Orbit-Stabilizer Theorem that

$$\#H \cdot \mathrm{Orb}(S_i) = \#\,\mathrm{Stab}(S_i) \cdot \#\mathrm{Orb}(S_i) = \#G = p^k m.$$

Since $\#H$ is a power of p and since $p^{k-e+1} \nmid \#\mathrm{Orb}(S_i)$ this implies that $\#\mathrm{Orb}(S_i) = p^{k-e}m$ and $\#H = p^e$. Thus $H \subseteq G$ is the desired subgroup of size p^e. □

Corollary. If $\#G = p^k m$ with p prime and $\gcd(p, m)$ then G has subgroups of order p^e for all $1 \leq e \leq k$.

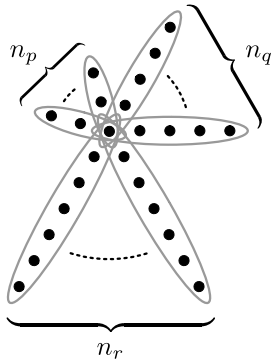
Proof. Let $H \subseteq G$ be a subgroup with $\#H = p^k$. Since $p|\#H$ we know from Cauchy's Theorem that there exists a subgroup $N \subseteq H$ with $\#N = p$. By induction the quotient H/N has subgroups of size p^e for all $1 \leq e \leq p^{k-1}$. Then the result follows from the Correspondence Theorem. □

The proof is not very instructive so I won't include it here. Instead I'll just show you a slick application.

Application of Sylow. Let $p < q < r$ be prime. No group of size pqr is 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 $n_p(p-1) + n_q(q-1) + n_r(r-1) + 1$ distinct elements of G:



Total # Dots:

$$n_p(p-1) + n_q(q-1) + n_r(r-1) + 1$$

It follows that

$$\begin{aligned} pqr = \#G &\geq n_p(p-1) + n_q(q-1) + n_r(r-1) + 1 \\ &\geq q(p-1) + r(q-1) + pq(r-1) + 1 \\ &= (\cancel{pq} - q) + (qr - q) + (pqr - \cancel{pq}) + 1 \\ &= pqr + (qr - q - r + 1) \\ &= pqr + (q-1)(r-1), \end{aligned}$$

which implies that $0 \geq (q-1)(r-1)$. But this contradicts the fact that $(q-1) > 0$ and $(r-1) > 0$. □

[**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.]

Exercises

11.A Some Examples of Conjugacy Classes

Let G be a group and for all $a, b \in G$ define the following relation:

$$a \sim b \quad \Longleftrightarrow \quad a = gbg^{-1} \text{ for some } g \in G.$$

(a) Prove that this is an equivalence relation, called *conjugacy*.

(b) Compute the conjugacy classes for the dihedral group:

$$D_{2n} = \langle R, F \rangle = \{I, R, \ldots, R^{n-1}, F, RF, \ldots, R^{n-1}F\}.$$

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?]

11.B 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 $p_1, \ldots, p_n$. The Sylow Theorems tell us that for each i there exists a unique subgroup $H_i \subseteq G$ of size $\#H = p_i^{e_i}$. 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 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.

11.C Sylow Two and Thee

11.D Euler's Rotation Theorem

Recall the definition of the special orthogonal group:

$$SO(3) = \{A \in \mathrm{Mat}_3(\mathbb{R}) : A^T A = I \text{ and } \det(A) = 1\}.$$

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.

[**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 $R_1 R_2 \in SO(3)$ is also a rotation. This fact is **not obvious** to the human visual imagination.]

Week 12

12.1 Conjugacy Classes in the Symmetric Group

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**.[88] 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 $\{G_i/G_{i+1}\}_i$ 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.

[88] In fact, the infinite group $SO(3)$ is also simple, but this is beyond the scope of the course.

It follows from this that the group S_n has a unique composition series:

$$S_n \supsetneq A_n \supsetneq \{id\}.$$

Hence the simple composition factors are $S_n/A_n \cong \mathbb{Z}/2\mathbb{Z}$ and $A_n/\{id\} \cong A_n$. ///

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 $n = 5$. ///

The following trick will help us to prove (1) that A_5 is the only non-trivial normal subgroup of S_5 and (2) that A_5 is a simple group.

Trick. Let $N \trianglelefteq G$ be a normal subgroup. Then N is a union of conjugacy classes.

Proof. Let $\kappa : G \to \mathrm{Aut}(G)$ 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

$$\mathrm{Orb}_\kappa(n) = \{gng^{-1} : g \in G\} \subseteq N.$$

It follows that

$$\bigcup_{n \in N} \mathrm{Orb}_\kappa(n) \subseteq N.$$

On the other hand, we obviously have

$$N \subseteq \bigcup_{n \in N} \mathrm{Orb}_\kappa(n)$$

because $n \in \mathrm{Orb}_\kappa(n)$ 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. The following result was first stated by Cauchy in the *Exercices d'analyse et de physique mathematique* (1844).

Theorem (Sizes of Conjugacy Classes in S_n). Let $\kappa : S_n \to \mathrm{Aut}(S_n)$ be the conjugation action and consider any permutation $f \in S_n$. Recall that the conjugacy class $K(f) \coloneqq \mathrm{Orb}_\kappa(f)$ 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 contains m_i cycles of length i, for each $i \in \{1, 2, \ldots, n\}$. Then we have

$$\#K(f) = \frac{n!}{1^{m_1} m_1!\, 2^{m_2} m_2! \cdots n^{m_n} m_n!}.$$

///

Before proving this, let's test some simple examples. Note that the identity permutation $id \in S_n$ has $m_1 = n$ cycles of length 1 and $m_i = 0$ cycles of length i for each $i \in \{2, 3, \ldots, n\}$. Thus the formula gives

$$\#K(id) = \frac{n!}{1^n n!\, 2^0 0! \cdots n^0 0!} = \frac{n!}{n!} = 1.$$

This is correct because the identity is only conjugate to itself. Next, let's count the conjugacy class of transpotitions (2-cycles), which has $m_1 = n-2$, $m_2 = 1$ and $m_i = 0$ for $i \in \{3, \ldots, n\}$. If $t \in S_n$ is any transposition then the formula gives

$$\#K(t) = \frac{n!}{1^{n-2}(n-2)!\, 2^1 1!\, 3^0 0! \cdots n^0 0!} = \frac{n!}{2(n-2)!} = \frac{n!}{2!(n-2)!} = \binom{n}{2}.$$

This is correct because each transposition $(ij) \in S_n$ corresponds to a choice of two elements $i \neq j$ from the set $\{1, 2, \ldots, n\}$.

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*:

$$Z(f) \coloneqq \mathrm{Stab}_\kappa(f) = \{g \in S_n : gfg^{-1} = f\}.$$

Now suppose that the permutation f has m_i cycles of length i for each $i \in \{1, 2, \ldots, n\}$. 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 $(j_1, j_2, \ldots, j_i)$ 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 $(g(j_1), g(j_2), \ldots, g(j_i))$ 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

$$(j_1, j_2, \ldots, j_i) = (g(j_1), g(j_2), \ldots, g(j_i)).$$

In this case there are exactly i ways to choose the values $g(j_1), g(j_2), \ldots, g(j_i) \in \{j_1, j_2, \ldots, j_i\}$ 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

- $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_5 \geq \lambda_5 \geq 0$,
- $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 5$.

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 $\{id\}$ and $\lambda = 21110$ corresponds to the conjugacy class of 2-cycles $\{(12), (13), \ldots, (45)\}$, 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 $\{id\}$. 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**. If follows that $A_5 \trianglelefteq S_5$ is the only non-trivial normal subgroup of S_5. □

12.2 The Icosahedron and A_5

Obviously, the previous 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.[89] The conjugacy classes

[89] See the second edition of Artin for a not-nice proof.

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

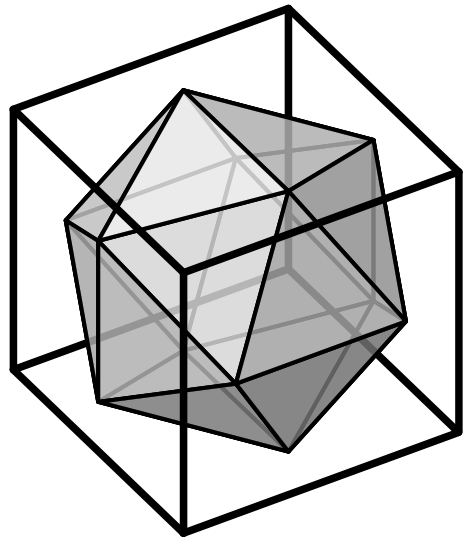
$$1 + 12 + 12 + 15 + 20 = 60 = \#I.$$

Now let's look for normal subgroups. Recall that any normal subgroup $N \trianglelefteq I$ is 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. □

Theorem. The icosahedral group $I \subseteq SO(3)$ is isomorphic to the alternating group $A_5 \subseteq S_5$.

Proof. The proof relies on the strange fact that a regular icosahedron can be inscribed in exactly 5 different cubes.[90] To see this, we observe that the 30 edges of the icosahedron divide into 15 parallel pairs. Furthermore, these 15 pairs can be divided into 5 triples of mutually orthogonal pairs. And for each triple of pairs there exists a unique cube whose 6 faces contain the 6 edges of the triple. Here is a picture:

[90] I learned this proof from Michael Artin's *Algebra*.



Now consider the set of these five cubes. Since any isometry $f \in I$ sends cubes to cubes we obtain a group homomorphism

$$\varphi : I \to \mathrm{Perm}(\{5 \text{ cubes}\}) \cong S_5.$$

Since $\ker\varphi \trianglelefteq I$ is a normal subgroup and since I is **simple**, we must have $\ker\varphi = \{id\}$ or $\ker\varphi = I$. But the second option is impossible because clearly some element of I moves the cubes. Therefore we have $\ker\varphi = \{id\}$ and hence φ is injective. It follows from the First Isomorphism Theorem that I is isomorphic to its image:

$$I \cong I/\ker\varphi \cong \mathrm{im}\,\varphi \subseteq S_5.$$

Now it only remains to show that $\mathrm{im}\,\varphi = A_5$. To do this we recall from Exercise 7.A that the determinant homomorphism $\det : S_n \to \{\pm 1\}$ has kernel $A_5 = \ker(\det)$. Consider the composition of these homomorphisms:

$$\det \circ \varphi : I \to \{\pm 1\}.$$

Again, since $\ker(\det\circ\varphi) \trianglelefteq I$ is a normal subgroup and since I is **simple**, we must have $\ker(\det\circ\varphi) = \{id\}$ or $\ker(\det\circ\varphi) = I$. This time the first option is impossible because I has more elements than $\{\pm 1\}$. It follows that $\ker(\det\circ\varphi) = I$ and hence

$$\mathrm{im}\,\varphi \subseteq \ker(\det) = A_5 \subseteq S_5.$$

Finally, since $\#\mathrm{im}\,\varphi = \#I = \#A_5 = 60$, we conclude that $\mathrm{im}\,\varphi = A_5$ as desired. □

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.

12.3 Epilogue: Finite Simple Groups

Epilogue: 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.[91]

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.[92] 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.

[91] Fine, I'll mention them. There is one more "classical" family $\mathrm{Sp}(n)$ coming from the quaternions and then five "exceptional" families called G_2, F_4, E_6, E_7, E_8.

[92] I recommend John Stillwell's *Naive Lie Theory*.

Exercises

12.A 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:

$$V = \{id, (12)(34), (13)(24), (14)(23)\} \trianglelefteq A_4.$$

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 Vierergruppe*. [Once upon a time it was surprising that not every abelian group is cyclic.]

12.B 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: $Z(S_n) = \{id\}$. [Hint: For any $id \neq g \in S_n$, 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 $\{id\}$ 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 $N = \{id, \tau\}$ for some $\tau \in S_n$ such that $\tau \neq id$ and $\tau^2 = id$. Prove that $\tau \in Z(S_n)$ and get a contradiction.

[**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.]

12.C 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 $GL_n(\mathbb{F}_p)$:

$$\left\{ \left(\begin{array}{c|c} A & C \\ \hline 0 & B \end{array} \right) : A \in GL_k(\mathbb{F}_p), B \in GL_{n-k}(\mathbb{F}_p), C \in \mathrm{Mat}_{k\times(n-k)}(\mathbb{F}_p) \right\}.$$

[Hint: Choose a basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ for $\mathbb{F}_p^n$ such that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ 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.*

[**Remark:** If we treat p as a formal variable then one can check that

$$\frac{[n]_p!}{[k]_p! \cdot [n-k]_p!} \longrightarrow \frac{n!}{k!\,(n-k)!} \qquad \text{as} \qquad p \longrightarrow 1.$$

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 this 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.]

Second Semester: Field Theory

Week 13

13.1 The Classical Problem of Algebra

Last semester I used the story of Galois Theory to motivate the study of **abstract groups**. This semester I will use the same story to motivate the study of **abstract rings** and **fields**. As before, we will find that **linear algebra** is always hiding just beneath the surface.

To begin, let me refresh your memory. The classical (pre-1830) problem of algebra was to find explicit "formulas" for the roots of a polynomial equation. To be precise, suppose that some rational numbers (called "coefficients") are given:

$$e_1, e_2, \ldots, e_n \in \mathbb{Q}.$$

Then we want to find some numbers $r_1, r_2, \ldots, r_n$ (called "roots") such that

$$x^n - e_1 x^{n-1} + e_2 x^{n-2} - \cdots + (-1)^n e_n = (x - r_1)(x - r_2) \cdots (x - r_n).$$

A priori, it is not obvious what kind of "numbers" the roots should be, or whether they exist at all. Soon we will prove a result called the Fundamental Theorem of Algebra which says that the roots always exist in the field $\mathbb{C}$ of complex numbers. Unfortunately, this theorem will not tell us how to **find** the roots.

We would really like to have some formula or algorithm for computing the roots. To state the problem explicitly, we expand the right hand side and then equate coefficients to obtain a system of n non-linear equations in n unknowns:

$$\begin{cases} e_1 &= r_1 + r_2 + \cdots + r_n \\ e_2 &= r_1 r_2 + r_1 r_3 + \cdots + r_{n-1} r_n \\ &\vdots \\ e_k &= \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} r_{i_1} r_{i_2} \cdots r_{i_k} \\ &\vdots \\ e_n &= r_1 r_2 \cdots r_n. \end{cases}$$

Our goal is to somehow "invert" this system. The best we could hope for is to

find some explicit functions $f_1, f_2, \ldots, f_n$ from $\mathbb{Q}^n$ to $\mathbb{C}$ such that

$$\begin{cases} r_1 &= f_1(e_1, e_2, \ldots, e_n) \\ r_2 &= f_2(e_1, e_2, \ldots, e_n) \\ &\vdots \\ r_n &= f_n(e_1, e_2, \ldots, e_n). \end{cases}$$

But this hope is too naive. Indeed, no such functions can exist. To see why, observe that each coefficient e_k can be thought of as a function of the roots:

$$e_k = e_k(r_1, r_2, \ldots, r_n) = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} r_{i_1} r_{i_2} \cdots r_{i_k}.$$

Furthermore, this function has the nice property of being "symmetric" under permutations of the roots. In other words, if $\sigma \in S_n$ is any permutation of the set $\{1, 2, \ldots, n\}$ then we have

$$e_k(r_1, r_2, \ldots, r_n) = e_k(r_{\sigma(1)}, r_{\sigma(2)}, \ldots, r_{\sigma(n)}).$$

The easiest way to see this is to observe that the product $(x - r_1)(x - r_2) \cdots (x - r_n)$ is symmetric under permutations of its factors. If we expand the product then each coefficient must also be a symmetric function.

[**Jargon:** The coefficients $e_1, \ldots, e_n$ are called the *elementary symmetric functions* of the roots. This explains my use of the letter "e".]

If $f : \mathbb{Q}^n \to \mathbb{C}$ is any function then we can think of the expression $f(e_1, \ldots, e_n)$ as a function of the roots, as follows:

$$f(e_1, \ldots, e_n)(r_1, \ldots, r_n) \coloneqq f\left(e_1(r_1, \ldots, r_n), \ldots, e_n(r_1, \ldots, r_n)\right).$$

Furthermore, it is clear that $f(e_1, \ldots, e_n)$ is a symmetric function of the roots. Now we see why our first hope was too naive:

> There can be no function f_k such that $r_k = f_k(e_1, e_2, \ldots, e_n)$ because $f_k(e_1, e_2, \ldots, e_n)$ is always a **symmetric** function of the roots, whereas r_k is certainly **not** a symmetric function of the roots.

The solution is to weaken the requirement that f_k is a "function". Instead, we will allow "multi-valued functions"[93] such as square roots. We often talk about "the" square root as though it were a function

$$\sqrt{-} : \mathbb{C} \to \mathbb{C}.$$

But this is **not** a function. Indeed, for any $0 \neq \alpha \in \mathbb{C}$, the expression $\sqrt{\alpha}$ represents **two distinct complex numbers** and there is no natural way

[93] Recall that a "multi-valued function" is **not** a function. This is a terrible but common notation.

to choose between them. We can use this ambiguity to solve the quadratic equation.

Example: The Quadratic Formula. For any rational coefficients $e_1, e_2 \in \mathbb{Q}$ we want to find some complex roots $r_1, r_2 \in \mathbb{C}$ such that

$$x^2 - e_1 x + e_2 = (x - r_1)(x - r_2) = x^2 - (r_1 + r_2)x + (r_1 r_2).$$

In other words, we want to find some "multi-valued functions" f_1, f_2 such that

$$\left\{ \begin{array}{lcl} e_1 &=& r_1 + r_2 \\ e_2 &=& r_1 r_2 \end{array} \right. \iff \left\{ \begin{array}{lcl} r_1 &=& f_1(e_1, e_2) \\ r_2 &=& f_2(e_1, e_2). \end{array} \right.$$

As you know, the solution is

$$\left\{ \begin{array}{lcl} f_1(e_1, e_2) &=& (e_1 + \sqrt{e_1^2 - 4e_2})/2 \\ f_2(e_1, e_2) &=& (e_1 - \sqrt{e_1^2 - 4e_2})/2. \end{array} \right.$$

The only subtlety here is that we must interpret the ambiguous expression "$\sqrt{e_1^2 - 4e_2}$" in the **same way** for both equations. In other words, we let "$\sqrt{e_1^2 - 4e_2}$" denote **one particular number** $\alpha \in \mathbb{C}$ such that $\alpha^2 = e_1^2 - 4e_2$. If $e_1^2 - 4e_2 \neq 0$ then there will be two choices and we just pick one at random.[94]

///

The process of choosing a random square root is called "breaking the symmetry". For higher degree equations we expect that we will need to break the symmetry by choosing random 3rd roots, 4th roots, etc. This leads us to the classical problem of algebra.

The Classical Problem of Algebra. Let $e_1, \ldots, e_n \in \mathbb{Q}$ be any rational numbers. By the Fundamental Theorem of Algebra there exist some unique complex numbers $r_1, \ldots, r_n \in \mathbb{C}$ such that

$$x^n - e_1 x^{n-1} + e_2 x^{n-2} - \cdots + (-1)^n e_n = (x - r_1)(x - r_2) \cdots (x - r_n).$$

Our goal is to find some way to compute these roots. Specifically, we want to find an "algebraic formula" expressing the roots in terms of the coefficients, using only the "algebraic operations"

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

///

As you know, this problem turns out to be **impossible** when $n \geq 5$. Last semester we developed the group theory necessary for the proof of impossibility.[95] This semester we will fill in the other half of the proof.

[94] If the roots are imaginary then there is a deep sense in which they are indistinguishable. Indeed, we usually define the imaginary unit i as "the" square root of -1. But if -1 has any square root then it must have two. Which one do you want to call i?

[95] Specifically, we proved that the group S_n not "solvable" when $n \geq 5$.

13.2 Definition of Fields

The Classical Problem was definitively solved by Galois in the 1820s. However, he died too soon to really explain it to anyone. Galois' work was eventually published in 1846 by Joseph Liouville. The first textbook on Galois theory was Camille Jordan's *Traité des substitutions et des équations algébraiques* (1870). At this point "Galois theory" and "group theory" were the same subject, so Jordan's work can also be viewed as the first book about groups.

However, the subject was still difficult to understand. The next major advance was made by Richard Dedekind when he defined the concept of a *field*.[96] Apparently Dedekind lectured on this material at Göttingen as early as the 1850s, but he published it in an 1894 supplement to Dirichlet's *Lectures on Number Theory*. Dedekind began §160 of the supplement with the definition of fields and later said in §164 that "the real subject of today's algebra" lies in "the detailed investigation of the relationship between different fields".[97]

Definition of Fields and Subfields/Extensions. A *field* is a set $\mathbb{F}$ together with two binary operations

$$+, \times : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$$

(called *addition* and *multiplication*) and two special elements

$$0, 1 \in \mathbb{F}$$

(called *zero* and *one*), which satisfy the following three axioms:

(F1) $(\mathbb{F}, +, 0)$ is an abelian group.

(F2) $(\mathbb{F} - \{0\}, \times, 1)$ is an abelian group. We will use juxtaposition to denote multiplication:

$$ab \coloneqq a \times b.$$

Furthermore, since multiplication is commutative we are free to use fractional notation to denote division:

$$ab^{-1} = b^{-1}a = \frac{a}{b}.$$

(F3) *Distribution.* For all $a, b, c \in \mathbb{F}$ we have

$$a(b + c) = ab + ac.$$

[96] Dedekind's name for this structure was *Körper*, short for *Zahlkörper* (body of numbers). Dedekind's rival Leopold Kronecker used the term *Rationalitätsbereich* (domain of rationality). The English term *field* was coined by E. H. Moore in 1893, possibly motivated by the word "domain". This creates a problem for English speakers: should we denote fields by the letter K or the letter F? To avoid confusion I will use the blackboard bold font (i.e., $\mathbb{K}$ or $\mathbb{F}$) to denote fields. Sadly, this is not a perfect solution because $\mathbb{Z}$ and $\mathbb{N}$ are **not** fields.

[97] Quoted from Edward T. Dean, *Dedekind's treatment of Galois theory in the Vorlesungen* (2009, page 27).

Now let $S \subseteq \mathbb{F}$ be any subset. We say that S is a *subfield* of $\mathbb{F}$ (equivalently, $\mathbb{F}$ is a *field extension* of S) if the following properties are satisfied:

- The special elements $0, 1$ are in S.
- For all $a, b \in S$ we have $a \pm b \in S$.
- For all $a, b \in S$ we have $ab \in S$.
- For all $a \in S - \{0\}$ we have $a^{-1} \in S$.

In other words: A subfield is a subset that is also a field with respect to the same operations and special elements. [**Remark:** We could shorten this definition by using the word "subgroup" in various places. Unfortunately, the subfield test cannot be reduced to one step, as the subgroup test can.] ///

The most basic examples of fields are

$$\mathbb{Q}, \mathbb{R}, \mathbb{C} \quad \text{and} \quad \mathbb{F}_p \coloneqq \mathbb{Z}/p\mathbb{Z} \text{ for } p \text{ prime.}$$

Note that the inclusions $\mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$ are field extensions. Prior to Dedekind no one felt the need to define the abstract concept of fields because it was synonymous with the concept of "numbers". However, Dedekind found that the abstract concept was helpful to simplify various ideas in number theory and Galois theory. Here is the first non-basic example.

The First Interesting Example. Let $\alpha = \sqrt{2}$ be any real number satisfying $\alpha^2 = 2$. If you want you want you can think of α as the **positive** square root of 2, but it doesn't really matter. Now consider the set

$$\mathbb{Q}(\sqrt{2}) \coloneqq \{a + b\sqrt{2} : a, b \in \mathbb{Q}\} \subseteq \mathbb{R}.$$

I claim that $\mathbb{Q}(\sqrt{2}) \subseteq \mathbb{R}$ is a subfield.

Proof. What needs to be checked?

- *Special Elements.* Note that $0 = 0 + 0\sqrt{2} \in \mathbb{Q}(\sqrt{2})$ and $1 = 1 + 0\sqrt{2} \in \mathbb{Q}(\sqrt{2})$.
- *Addition/Subtraction.* For all $a + b\sqrt{2}$ and $c + d\sqrt{2}$ in $\mathbb{Q}(\sqrt{2})$ note that

  $$(a + b\sqrt{2}) - (c + d\sqrt{2}) = (a - c) + (b - d)\sqrt{2} \in \mathbb{Q}(\sqrt{2})$$

  because $a - c \in \mathbb{Q}$ and $b - d \in \mathbb{Q}$.
- *Multiplication.* For all $a + b\sqrt{2}$ and $c + d\sqrt{2}$ in $\mathbb{Q}(\sqrt{2})$ we have

  $$(a + b\sqrt{2})(c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2} \in \mathbb{Q}(\sqrt{2})$$

  because $ac + 2bd \in \mathbb{Q}$ and $ad + bc \in \mathbb{Q}$.

- *Division.* This is the hardest step. For all nonzero elements $a + b\sqrt{2} \in \mathbb{Q}(\sqrt{2})$ we want to show that there exists some element $c+d\sqrt{2} \in \mathbb{Q}(\sqrt{2})$ such that
$$(a + b\sqrt{2})(c + d\sqrt{2}) = 1 = 1 + 0\sqrt{2}.$$
The solution is a trick called "rationalizing the denominator". First note that
$$(a + b\sqrt{2})(a - b\sqrt{2}) = (a^2 - 2b^2) + (ab - ab)\sqrt{2} = (a^2 - 2b^2) + 0\sqrt{2} \in \mathbb{Q}.$$
Now assume that $a + b\sqrt{2} \neq 0$ (i.e., assume that a and b are not both zero). We are looking for rational numbers $c, d \in \mathbb{Q}$ such that
$$\begin{aligned} c + d\sqrt{2} &= \frac{1}{a + b\sqrt{2}} \\ &= \frac{1}{a + b\sqrt{2}} \cdot \frac{a - b\sqrt{2}}{a - b\sqrt{2}} \\ &= \frac{a - b\sqrt{2}}{a^2 - 2b^2} \\ &= \left(\frac{a}{a^2 - 2b^2}\right) + \left(\frac{-b}{a^2 - 2b^2}\right)\sqrt{2}. \end{aligned}$$
If $a^2 - 2b^2 \neq 0$ then we can take
$$c = \frac{a}{a^2 - 2b^2} \in \mathbb{Q} \qquad \text{and} \qquad d = \frac{-b}{a^2 - 2b^2} \in \mathbb{Q}.$$
So assume for contradiction that $a^2 - 2b^2 = 0$. If $b = 0$ then $a^2 = 2b^2 = 0$ implies $a = 0$, which contradicts the fact that a and b are not both zero. If $b \neq 0$ then we have
$$\begin{aligned} a^2 &= 2b^2 \\ a^2/b^2 &= 2 \\ (a/b)^2 &= 2. \end{aligned}$$
Since $a/b \in \mathbb{Q}$ this contradicts the well-known fact that $\pm\sqrt{2} \notin \mathbb{Q}$.[98] □

[**Jargon:** The field $\mathbb{Q}(\sqrt{2})$ is called $\mathbb{Q}$ *adjoin* $\sqrt{2}$. We will see a generalization of this construction below.]

13.3 Adjoining a Subset to a Subfield

As with subgroups, It follows immediately from the definition that the intersection of subfields is a subfield.

[98] I'll put a proof of this on the first homework to remind you.

Intersection of Subfields is a Subfield. Let $(\mathbb{F}, +, \times, 0, 1)$ be a field and let $\mathbb{K}_i \subseteq \mathbb{F}$ be any family of subfields (possibly infinite or even uncountable). Then the intersection

$$\bigcap_i \mathbb{K}_i \subseteq \mathbb{F}$$

is also a subfield.

Proof. Since $0, 1 \in \mathbb{K}_i$ for all i we have $0, 1 \in \cap_i \mathbb{K}_i$. Now consider any two elements $a, b \in \cap_i \mathbb{K}_i$ with $a \neq 0$. By definition this means that $a, b \in \mathbb{K}_i$ for each i. But then since $\mathbb{K}_i \subseteq \mathbb{F}$ is a subfield we know that $a \pm b$, ab and a^{-1} are in $\mathbb{K}_i$. It follows that $a \pm b$, ab and a^{-1} are also in the intersection. □

However, the union of subfields is not necessarily a subfield.

Union of Subfields is Not a Subfield. Consider the subfields $\mathbb{Q}(\sqrt{2}) \subseteq \mathbb{R}$ and $\mathbb{Q}(\sqrt{3}) \subseteq \mathbb{R}$.[99] I claim that the union $\mathbb{Q}(\sqrt{2}) \cup \mathbb{Q}(\sqrt{3})$ is **not** a subfield of $\mathbb{R}$.

Proof. Suppose for contradiction that $\mathbb{Q}(\sqrt{2}) \cup \mathbb{Q}(\sqrt{3}) \subseteq \mathbb{R}$ **is** a subfield. Since a subfield is closed under addition, we must have $\sqrt{2} + \sqrt{3} \in \mathbb{Q}(\sqrt{2}) \cup \mathbb{Q}(\sqrt{3})$, which implies that $\sqrt{2} + \sqrt{3}$ is in $\mathbb{Q}(\sqrt{2})$ or in $\mathbb{Q}(\sqrt{3})$. Let's assume that $\sqrt{2} + \sqrt{3} \in \mathbb{Q}(\sqrt{2})$. Then by definition we have

$$\sqrt{2} + \sqrt{3} = a + b\sqrt{2} \quad \text{for some } a, b \in \mathbb{Q}.$$

If $b = 1$ then we obtain the contradiction that $\sqrt{3} \in \mathbb{Q}$:

$$\sqrt{2} + \sqrt{3} = a + \sqrt{2} \quad \Longrightarrow \quad \sqrt{3} = a \in \mathbb{Q}.$$

Furthermore, if $a = 0$ then we obtain the contradiction that $\sqrt{6} \in \mathbb{Q}$:

$$\begin{aligned} \sqrt{2} + \sqrt{3} &= b\sqrt{2} \\ \sqrt{3} &= (b-1)\sqrt{2} \\ \sqrt{3} \cdot \sqrt{2} &= (b-1)\sqrt{2} \cdot \sqrt{2} \\ \sqrt{6} &= 2(b-1) \in \mathbb{Q}. \end{aligned}$$

But in all other cases we obtain the contradiction that $\sqrt{2} \in \mathbb{Q}$:

$$\begin{aligned} \sqrt{2} + \sqrt{3} &= a + b\sqrt{2} \\ \sqrt{3} &= a + (b-1)\sqrt{2} \\ 3 &= \left(a + (b-1)\sqrt{2}\right)^2 \end{aligned}$$

[99] The proof that $\mathbb{Q}(\sqrt{3}) \subseteq \mathbb{R}$ is a subfield is exactly the same as the proof for $\mathbb{Q}(\sqrt{2}) \subseteq \mathbb{R}$.

$$3 = \left(a^2 + 2(b-1)^2\right) + 2a(b-1)\sqrt{2}$$
$$\sqrt{2} = \frac{3 - \left(a^2 + 2(b-1)^2\right)}{2a(b-1)} \in \mathbb{Q}.$$

In conclusion we have $\sqrt{2}+\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$. A similar proof shows that $\sqrt{2}+\sqrt{3} \notin \mathbb{Q}(\sqrt{3})$. □

Remarks:

- In the proof we needed the fact that $\sqrt{2}, \sqrt{3}, \sqrt{6} \in \mathbb{R}$ are *irrational* numbers. On the homework you will prove for all integers $D \in \mathbb{Z}$ that

$$\pm\sqrt{D} \notin \mathbb{Z} \quad \Longrightarrow \quad \pm\sqrt{D} \notin \mathbb{Q}.$$

- There is a more sophisticated way to phrase the theorem we just proved. One can view the real numbers $(\mathbb{R}, +, 0)$ as a *vector space* over $\mathbb{Q}$ in a very boring way. That is, for every "scalar" $a \in \mathbb{Q}$ and for every "vector" $b \in \mathbb{R}$ we define "scalar multiplication" via regular multiplication:

$$a(b) \coloneqq ab \in \mathbb{R}.$$

  Then the vector space axioms are easily verified.[100] So what? In this language we can rephrase the above theorem by saying that

  the real numbers $1, \sqrt{2}, \sqrt{3} \in \mathbb{R}$ are *linearly independent* over $\mathbb{Q}$.

  In fact, it is true that any list of square roots of square-free integers is linearly independent over $\mathbb{Q}$, but this is quite tricky to prove. It seems that most algebra books pass over this fact without comment.

- More generally, if $\mathbb{E} \supseteq \mathbb{F}$ is any field extension then we can view $\mathbb{E}$ as a vector space over $\mathbb{F}$ in the same boring way. For this reason it turns out that linear algebra is very useful for the study of fields. In particular, we are interested in the **dimension**:

$$[\mathbb{E}/\mathbb{F}] \coloneqq \dim_{\mathbb{F}}(\mathbb{E}) = \text{the dimension of } \mathbb{E} \text{ as a vector space over } \mathbb{F}.$$

  On the homework you will verify that $[\mathbb{Q}(\sqrt{2})/\mathbb{Q}] = 2$. With more work (for example, by using the tricky theorem about square roots stated above) one can prove that $[\mathbb{R}/\mathbb{Q}] = \infty$, which is bad. In this course we prefer to study finite-dimensional field extensions.

  ///

As with subgroups, we should replace the union of subfields with the smallest subfield that contains the union. Here is the general construction.

[100] This is a good time to remind yourself of the vector space axioms.

The Subfield Generated by a Subset. Let $(\mathbb{F}, +, \times, 0, 1)$ be a field and let $S \subseteq \mathbb{F}$ be any subset. Let $\langle S \rangle \subseteq \mathbb{F}$ denote the intersection of all subfields $\mathbb{K} \subseteq \mathbb{F}$ that contain S:

$$\langle S \rangle := \bigcap_{S \subseteq \mathbb{K} \subseteq \mathbb{F}} \mathbb{K}.$$

We know from above that $\langle S \rangle \subseteq \mathbb{F}$ is a subfield. I claim that it is the **smallest** subfield of $\mathbb{F}$ that contains S. We call it the *subfield of* $\mathbb{F}$ *generated by* S.

Proof. The intersection is contained in any field that contains S. □

Actually, the notation $\langle S \rangle$ is not standard in field theory and we will only use it temporarily. The more common notation refers to the smallest subfield $\mathbb{F}' := \langle \emptyset \rangle \subseteq \mathbb{F}$, which is called the *prime subfield*. The reason for the notation is because the prime subfield of any field satisfies

$$\mathbb{F}' \cong \mathbb{Q} \qquad \text{or} \qquad \mathbb{F}' \cong \mathbb{Z}/p\mathbb{Z} \text{ for some prime } p.$$

You will prove this on a future homework, after we develop the necessary technology.

Here is the more standard notation for a subfield generated by a set.

The Definition of Adjunction. For any field extension $\mathbb{F} \subseteq \mathbb{E}$ and for any subset $S \subseteq \mathbb{E}$ we let $\mathbb{F} \subseteq \mathbb{F}(S) \subseteq \mathbb{E}$ be the intersection of all subfields that contain $\mathbb{F} \cup S$:

$$\mathbb{F}(S) := \langle \mathbb{F} \cup S \rangle = \bigcap_{(\mathbb{F} \cup S) \subseteq \mathbb{K} \subseteq \mathbb{E}} \mathbb{K}.$$

We call this field "$\mathbb{F}$ *adjoin* S". If we omit any mention of the base field $\mathbb{F}$ then we obtain

$$\langle S \rangle = \mathbb{E}'(S),$$

where $\mathbb{E}' \subseteq \mathbb{E}$ is the prime subfield. This explains the relationship between the standard and nonstandard terminology. In the case that S is a finite set we will write

$$\mathbb{F}(\{\alpha_1, \alpha_2, \ldots, \alpha_k\}) = \mathbb{F}(\alpha_1, \alpha_2, \ldots, \alpha_k).$$

This is the smallest field between $\mathbb{E}$ and $\mathbb{F}$ that contains the elements $\alpha_1, \ldots, \alpha_k$. ///

I'll ask you to verify some formal (i.e., "trivial") properties of adjunction on the homework. For example, you will verify that $\mathbb{F}(\alpha)(\beta) = \mathbb{F}(\beta)(\alpha) = \mathbb{F}(\alpha, \beta)$ for any $\alpha, \beta \in \mathbb{E}$.

Exercises

13.A Square Roots are Irrational

Let $D \in \mathbb{N}$ be a positive integer and let $\sqrt{D} \in \mathbb{R}$ be any real square root. In this problem you will show that

$$\sqrt{D} \notin \mathbb{Z} \quad \Longrightarrow \quad \sqrt{D} \notin \mathbb{Q}.$$

(a) Consider the set $S = \{n \in \mathbb{N} : n\sqrt{D} \in \mathbb{Z}\} \subseteq \mathbb{N}$. Observe that

$$S = \emptyset \quad \Longleftrightarrow \quad \sqrt{D} \notin \mathbb{Q}.$$

(b) Assuming that $\sqrt{D} \notin \mathbb{Z}$, use Well-Ordering to prove that there exists $a \in \mathbb{Z}$ such that

$$a < \sqrt{D} < a + 1.$$

(c) Suppose in addition that $\sqrt{D} \in \mathbb{Q}$. By part (a) and Well-Ordering, this means that the set S has a smallest element, say $m \in S$. Now use part (b) to obtain a contradiction. [Hint: Consider the number $m(\sqrt{D} - a)$.]

13.B Formal Properties of Adjunction

Let $\mathbb{F} \subseteq \mathbb{E}$ be a field extension and let $S \subseteq \mathbb{E}$ be any subset. We let $\mathbb{F}(S) \subseteq \mathbb{E}$ denote the smallest subfield of $\mathbb{E}$ that contains the set $\mathbb{F} \cup S$.

(a) Prove that $\mathbb{F}(S) = \mathbb{F}(S - \mathbb{F})$.

(b) For any two subsets $S, T \subseteq \mathbb{F}$ prove that $\mathbb{F}(S)(T) = \mathbb{F}(T)(S) = \mathbb{F}(S \cup T)$.

(c) If $\mathbb{K} \subseteq \mathbb{F}$ is a subfield, prove that $\mathbb{F}(\mathbb{K}) = \mathbb{K}(\mathbb{F}) = \mathbb{F} \vee \mathbb{K}$ is the join operation in the lattice of subfields. We also call this the *compositum* of subfields:

$$\mathbb{F}\mathbb{K} \coloneqq \mathbb{F}(\mathbb{K}).$$

Week 14

14.1 Definition of Galois Groups

We saw last week that the union of two subfields is not necessarily a subfield. Instead, we will replace the union with the least upper bound in the lattice of subfields.

The Lattice of Subfields. Let $\mathbb{E} \supseteq \mathbb{F}$ be a field extension and consider the set

$$\mathscr{L}(\mathbb{E}, \mathbb{F}) = \{\text{all subfields between } \mathbb{E} \text{ and } \mathbb{F}\} = \{\mathbb{K} : \mathbb{F} \subseteq \mathbb{K} \subseteq \mathbb{E}\}.$$

This set is partially ordered by containment, with bottom element $\mathbb{F}$ and top element $\mathbb{E}$. We have seen that any two intermediate fields $\mathbb{K}, \mathbb{L} \in \mathscr{L}(\mathbb{E}, \mathbb{F})$ have a greatest lower bound ("meet") given by the intersection:

$$\mathbb{K} \wedge \mathbb{L} = \mathbb{K} \cap \mathbb{L}.$$

Furthermore, the least upper bound ("join") is the intersection of all subfields containing the set $\mathbb{K} \cup \mathbb{L}$. Equivalently, we can view this field as "$\mathbb{K}$ adjoin $\mathbb{L}$" or "$\mathbb{L}$ adjoin $\mathbb{K}$":

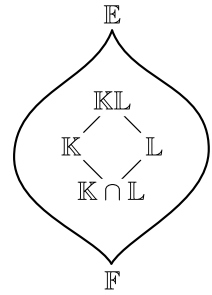
$$\mathbb{K} \vee \mathbb{L} \coloneqq \langle \mathbb{K} \cup \mathbb{L} \rangle = \mathbb{K}(\mathbb{L}) = \mathbb{L}(\mathbb{K}).$$

More commonly this operation is called the *compositum* of subfields:

$$\mathbb{KL} \coloneqq \mathbb{K} \vee \mathbb{L}.$$

With these operations, the set $\mathscr{L}(\mathbb{E}, \mathbb{F})$ is called the *lattice of intermediate fields*. Here is the picture that I have in my mind:

$\mathscr{L}(\mathbb{E}, \mathbb{F})$ =



If the base field $\mathbb{F}$ is not specified then we will write

$$\mathscr{L}(\mathbb{E}) = \mathscr{L}(\mathbb{E}, \mathbb{E}'),$$

where $\mathbb{E}' \subseteq \mathbb{E}$ is the prime subfield. This is called the *lattice of all subfields* of $\mathbb{E}$. ///

My goal for the rest of this week is to tell you Galois' Solvability Theorem in its modern form. I will also state the so-called Fundamental Theorem of Galois Theory, which is not due to Galois. It will take the rest of the semester to fill in the proofs of these theorems.

The main innovation of Galois was to associate a **group** to each polynomial equation $f(x) = 0$. If the coefficients of $f(x)$ lie in a field $\mathbb{F}$ then we will denote this group by $\mathrm{Gal}(f/\mathbb{F})$ and we will call it the *Galois group of* f *over* $\mathbb{F}$. Galois' original definition was a bit technical.

Galois' Definition of the Galois Group. The group $\mathrm{Gal}(f/\mathbb{F})$ is a certain subgroup of the group of permutations of the roots of $f(x)$. ///

It would take quite a few pages to tell you what "certain subgroup" means. Instead I will present the modern definition which is due to Dedekind. His main innovation was to translate the discussion of polynomials into the language of **fields**.

Dedekind's Definition of the Galois Group. Let $f(x)$ be a polynomial with coefficients in a field $\mathbb{F}$. There exists a certain "smallest" field extension $\mathbb{E} \supseteq \mathbb{F}$ (called the *splitting field*) in which $\mathbb{F}$ has all of its roots. For example, if $\mathbb{F} \subseteq \mathbb{C}$ then the FTA says that all the roots exist in $\mathbb{C}$, so $\mathbb{E}$ is just the intersection of all subfields that contain the roots. Then we define

$$\mathrm{Gal}(f/\mathbb{F}) \coloneqq \{\text{field automorphisms } \sigma : \mathbb{E} \to \mathbb{E} \text{ such that } \sigma(a) = a \text{ for all } a \in \mathbb{F}\}.$$

By a *field automorphism* we mean any invertible function $\sigma : \mathbb{E} \to \mathbb{E}$ that satisfies

$$\sigma(a+b) = \sigma(a) + \sigma(b) \qquad \text{and} \qquad \sigma(ab) = \sigma(a)\sigma(b)$$

for all $a, b \in \mathbb{E}$. [**Remark:** You will prove on the homework that the invertibility hypothesis is redundant. That is, you will show that any function $\sigma : \mathbb{E} \to \mathbb{E}$ that fixes $\mathbb{F}$ and preserves addition and multiplication is **necessarily** invertible.] Cleary the collection of such functions is a group under composition. Since the definition doesn't refer to the polynomial $f(x)$ we will also use the notation

$$\mathrm{Gal}(\mathbb{E}/\mathbb{F}) \coloneqq \mathrm{Gal}(f/\mathbb{F}).$$

///

And what do field automorphisms have to do with permutations of the roots? Suppose that $\alpha \in \mathbb{E}$ is a root of $f(x)$ and let $\sigma \in \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ be any element of the Galois group. I claim that $\sigma(\alpha) \in \mathbb{E}$ is also a root of $f(x)$.

Proof. Since $f(x)$ has coefficients in $\mathbb{F}$ we can write

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \qquad \text{for some } a_0, a_1, \ldots, a_n \in \mathbb{F}.$$

And since $\alpha \in \mathbb{E}$ is a root of $f(x)$ we have

$$f(\alpha) = a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n = 0.$$

Now we apply the field automorphism σ to both sides of this equation and use the fact that $\sigma(a) = a$ for all $a \in \mathbb{F}$ to obtain

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

In other words, $\sigma(\alpha) \in \mathbb{E}$ is a root of $f(x)$. □

It follows that every element $\sigma \in \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ restricts to a permutation of the roots of $f(x)$. Furthermore, it seems reasonable that since $\mathbb{E}$ is the "smallest" field containing the roots then two different field automorphisms should restrict to two different permutations. Thus we obtain an **injective group homomorphism**:

$$\mathrm{Gal}(f/\mathbb{F}) \to \{\text{permutations of the roots of } f(x)\}.$$

We will return to the details in Weeks 21 and 22.

14.2 Basic Examples

Before stating the Fundamental Theorem I want to show you a couple of basic examples.

Example: The Galois Group of $x^2 - 2$.

Consider the polynomial $x^2 - 2$ with coefficients in $\mathbb{Q}$. If $\sqrt{2} \in \mathbb{R}$ is the positive real square root of 2 then we know that this polynomial has exactly two roots:

$+\sqrt{2}$ and $-\sqrt{2}$.[101] I claim that the splitting field $\mathbb{E} \supseteq \mathbb{Q}$ is the same field that we studied above:

$$\mathbb{E} = \mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}.$$

Proof. Certainly we know that the field $\mathbb{Q}(\sqrt{2})$ contains the roots $+\sqrt{2}$ and $-\sqrt{2}$. Now let $\mathbb{K} \supseteq \mathbb{Q}$ be any field extension that contains these roots and consider any two rational numbers $a, b \in \mathbb{Q}$. Then since $\mathbb{K}$ is closed under addition and multiplication we have

$$a, b, \sqrt{2} \in \mathbb{K} \qquad \Longrightarrow \qquad a + b\sqrt{2} \in \mathbb{K},$$

and it follows that $\mathbb{Q}(\sqrt{2})$. □

Now let $\sigma \in \mathrm{Gal}(\mathbb{Q}(\sqrt{2})/\mathbb{Q})$ be an element of the Galois group. By definition this means that $\sigma : \mathbb{Q}(\sqrt{2}) \to \mathbb{Q}(\sqrt{2})$ is a field automorphism that fixes elements of $\mathbb{Q}$.[102] This means that for all $a + b\sqrt{2} \in \mathbb{Q}(\sqrt{2})$ we have

$$\sigma(a + b\sqrt{2}) = \sigma(a) + \sigma(b)\sigma(\sqrt{2}) = a + b\sigma(\sqrt{2}) \in \mathbb{Q}(\sqrt{2}),$$

hence the automorphism σ is uniquely specified by the value of $\sigma(\sqrt{2})$. What are the options? Since $\sqrt{2}$ is a root of the polynomial $x^2 - 2$ we must have

$$\begin{aligned}
(\sqrt{2})^2 - 2 &= 0 \\
\sigma\left((\sqrt{2})^2 - 2\right) &= \sigma(0) \\
\sigma(\sqrt{2})^2 - \sigma(2) &= \sigma(0) \\
\sigma(\sqrt{2})^2 - 2 &= 0,
\end{aligned}$$

and it follows that $\sigma(\sqrt{2}) \in \{\pm\sqrt{2}\}$. Therefore σ must be one of the following two functions:

$$\begin{aligned}
id\,(a + b\sqrt{2}) &= a + b\sqrt{2}, \\
\tau(a + b\sqrt{2}) &= a - b\sqrt{2}.
\end{aligned}$$

The only remaining question is whether these two functions are indeed **field automorphisms**. Well, the identity clearly is, but it needs to be checked by hand that τ preserves addition and multiplication. You will do this on the homework.

[101] Wait, why do we know this? You will prove on the next homework that a polyomial of degree n can have at most n roots in any field extension. It may have more roots in other kinds of ring extensions. For example, the polynomial $x^2 - 2$ has **uncountably many** roots in the ring of quaternions $\mathbb{H} \supseteq \mathbb{Q}$.

[102] Actually, the requirement that σ fixes $\mathbb{Q}$ is redundant here because $\mathbb{Q}$ is the prime subfield of $\mathbb{Q}(\sqrt{2})$.

In summary, we have found that the Galois group of the equation $x^2 - 2 = 0$ is the group of size 2 generated by the "conjugation automorphism" $\tau : \mathbb{Q}(\sqrt{2}) \to \mathbb{Q}(\sqrt{2})$:

$$\mathrm{Gal}((x^2-2)/\mathbb{Q}) = \mathrm{Gal}(\mathbb{Q}(\sqrt{2})/\mathbb{Q}) = \{id, \tau\} \cong \mathbb{Z}/2\mathbb{Z}.$$

On the homework you will show that essentially the same results hold for any so-called "quadratic field extension" $\mathbb{F}(\sqrt{D}) \supseteq \mathbb{F}$. Another example is the complex field over the real field, which is the splitting field of the polynomial $x^2 + 1$. In this case we have

$$\mathrm{Gal}((x^2+1)/\mathbb{R}) = \mathrm{Gal}(\mathbb{C}/\mathbb{R}) = \{id, \tau\} \cong \mathbb{Z}/2\mathbb{Z},$$

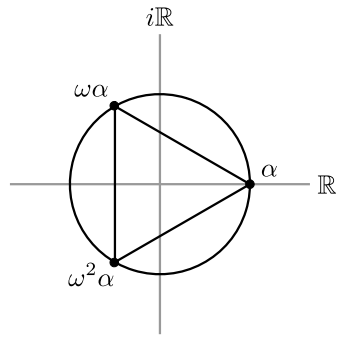
where the function $\tau : \mathbb{C} \to \mathbb{C}$ defined by

$$\tau(a + b\sqrt{-1}) \coloneqq a - b\sqrt{-1}$$

is called "complex conjugation". ///

Example: The Galois Group of $x^3 - 2$.

Consider the polynomial $x^3 - 2$ with coefficients in $\mathbb{Q}$. We know that this polynomial has one real root and two complex roots. To be specific, let $\alpha \coloneqq \sqrt[3]{2} \in \mathbb{R}$ be the real 3rd root of 2 and let $\omega \coloneqq \exp(2\pi i/3)$ be a primitive 3rd root of 1. Then the roots $\alpha, \omega\alpha, \omega^2\alpha$ are the verticies of an equilateral triangle in the complex plane:



I claim that the splitting field $\mathbb{E} \supseteq \mathbb{Q}$ is obtained by adjoining the set $\{\alpha, \omega\}$:

$$\mathbb{E} = \mathbb{Q}(\alpha, \omega) = \text{ the smallest subfield of } \mathbb{C} \text{ that contains } \mathbb{Q} \cup \{\alpha, \omega\}.$$

Proof. Since the field $\mathbb{Q}(\alpha, \omega)$ contains the elements α, ω and is closed under multiplication, it must contain the roots $\alpha, \omega\alpha, \omega^2\alpha$. Now let $\mathbb{C} \supseteq \mathbb{K} \supseteq \mathbb{Q}$ be any field that contains the roots. Then since $\mathbb{K}$ is closed under inversion we must have

$$\alpha, \omega\alpha \in \mathbb{K} \qquad \Longrightarrow \qquad \omega = (\omega\alpha)(\alpha^{-1}) \in \mathbb{K}.$$

It follows that $\mathbb{Q} \cup \{\alpha, \omega\} \subseteq \mathbb{K}$ and hence $\mathbb{Q}(\alpha, \omega) \subseteq \mathbb{K}$. □

Unlike the previous example, we do not already know a basis for the vector space $\mathbb{Q}(\alpha,\omega)/\mathbb{Q}$ and this makes it harder to compute the Galois group. So let me just tell you without proof that the set $\{1,\alpha,\alpha^2,\omega,\omega\alpha,\omega\alpha^2\}$ is a basis. In other words, every element of the splitting field $\gamma \in \mathbb{Q}(\alpha,\omega)$ can be written in the form

$$\gamma = a + b\alpha + c\alpha^2 + d\omega + e\omega\alpha + f\omega\alpha^2 \qquad \text{for some } \textbf{unique } a,b,c,d,e,f \in \mathbb{Q}.$$

If $\sigma \in \mathrm{Gal}(\mathbb{Q}(\alpha,\omega)/\mathbb{Q})$ is any element of the Galois group then we find that

$$\sigma(\gamma) = a + b\sigma(\alpha)a + c\sigma(\alpha)^2 + d\sigma(\omega) + e\sigma(\omega)\sigma(\alpha) + f\sigma(\omega)\sigma(\alpha)^2,$$

and it follows that σ is uniquely specified by the values $\sigma(\alpha)$ and $\sigma(\omega)$. What are the options? Since $\alpha^3 - 2 = 0$ and $\omega^3 - 1 = 0$ we find that

$$\sigma(\alpha)^3 - 2 = 0 \qquad \text{and} \qquad \sigma(\omega)^3 - 1 = 0.$$

Furthermore, since σ is invertible with $\sigma(1) = 1$ and $\omega \neq 1$, we know that $\sigma(\omega) \neq 1$. It follows that there are at most $6 = 3 \cdot 2$ possibilities:

$$\sigma(\alpha) \in \{\alpha, \omega\alpha, \omega^2\alpha\} \qquad \text{and} \qquad \sigma(\omega) \in \{\omega, \omega^2\}.$$

Let me claim without proof that each of these six functions is indeed a **field automorphism**. It would extremely tedious to check this by hand. Later we will have an indirect method.

Thus we obtain a Galois group of size 6. Finally, I claim that this is the group of **all permutations** of the three roots, and hence

$$\mathrm{Gal}((x^3-2)/\mathbb{Q}) = \mathrm{Gal}(\mathbb{Q}(\alpha,\omega)/\mathbb{Q}) \cong S_3.$$

Proof. The function defined by $(\sigma(\alpha),\sigma(\omega)) \coloneqq (\alpha,\omega^2)$ transposes the roots $\omega\alpha$ and $\omega^2\alpha$ and leaves α alone. (In fact this map is just complex conjugation.) Furthermore, the function defined by $(\sigma(\alpha),\sigma(\omega)) \coloneqq (\omega\alpha,\omega^2)$ transposes the roots α and $\omega\alpha$ and leaves $\omega^2\alpha$ alone. Any other permutation can be obtained by composing these two transpositions. □

I apologize that there were some gaps in the second example. Sadly it will take some time to fill them in. But I wanted to have this example available next time when we discuss the Fundamental Theorem.

14.3 Preview of the Fundamental Theorem

Now we have enough ingredients that I can state the main theorems of Galois theory. The modern definitions are really due to Dedekind, and the notation

is heavily influenced by Emil Artin's 1942 lectures at the University of Notre Dame.[103]

First, here is the Dedekind-Artin translation of the notion of "solvability".

Definition of Solvable Field Extensions. Let $\mathbb{E} \supseteq \mathbb{F}$ be a field extension. We say that this extension is *solvable* if there exists a chain of field extensions

$$\mathbb{F} = \mathbb{F}_0 \subseteq \mathbb{F}_1 \subseteq \cdots \subseteq \mathbb{F}_k \supseteq \mathbb{E}$$

satisfying the following condition:

> For all i we have $\mathbb{F}_i = \mathbb{F}_{i-1}(\alpha_i)$ for some element $\alpha_i \in \mathbb{F}_i$ such that $\alpha_i \notin \mathbb{F}_{i-1}$ but $\alpha_i^{n_i} \in \mathbb{F}_{i-1}$ for some power $n_i \geq 2$.

Essentially, this just means that every element of the field $\mathbb{E}$ can be expressed in terms of the elements of $\mathbb{F}$ using only the operations

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

Beginning with $\mathbb{F} = \mathbb{F}_0$, if we apply the operations $+, -, \times, \div$ then we will stay inside the same field. But if we adjoin a specific n_i-th root α_i of some element $\alpha_i^{n_i} \in \mathbb{F}_{i-1}$ then we may jump up into a bigger field $\mathbb{F}_i = \mathbb{F}_{i-1}(\alpha_i)$. The goal is to obtain every element of $\mathbb{E}$ after a finite number of adjunctions. If $\mathbb{E}$ is a field containing the roots of a polynomial $f(x) \in \mathbb{F}[x]$ then we also say that $f(x) = 0$ is *solvable by radicals.* ///

And here is the big theorem. This theorem is the ultimate motivation for many of the definitions in field theory and group theory. It took over 100 years to clean up all the details and still most mathematicians have never seen a full proof.

Galois' Solvability Theorem. Let $\mathbb{E} \supseteq \mathbb{F}$ be the splitting field for some polynomial $f(x)$ with coefficients in $\mathbb{F}$ and let $G = \mathrm{Gal}(\mathbb{E}/\mathbb{F}) = \mathrm{Gal}(f/\mathbb{F})$ be the Galois group. Then we have

$$\left\{ \begin{array}{c} \mathbb{E} \supseteq \mathbb{F} \text{ is a solvable} \\ \text{field extension} \end{array} \right\} \iff \left\{ \begin{array}{c} G \text{ is a solvable} \\ \text{group} \end{array} \right\}.$$

///

[103] Dedekind was the last student of Carl Friedrich Gauss at the University of Göttingen. The modern language of abstract algebra later emerged through the lectures of Emmy Noether at Göttingen and Emil Artin as Hamburg in the 1920s. Noether in particular viewed Dedekind as the spiritual father of the subject. When the German universities were decimated by the Nazis, many prominent mathematicians, including Artin and Noether, ended up in the United States.

The key idea of the proof is a certain "abstract Galois connection" between the lattice of intermediate fields $\mathscr{L}(\mathbb{E},\mathbb{F})$ and the lattice of subgroups $\mathscr{L}(G)$. Recall that $G = \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ is the group of field automorphisms $\mathbb{E} \to \mathbb{E}$ that fix elements of the subfield $\mathbb{F}$. If $\mathbb{F} \subseteq \mathbb{K} \subseteq \mathbb{E}$ is any intermediate field then it follows by definition that $\mathrm{Gal}(\mathbb{E}/\mathbb{K})$ is a subgroup of G. Indeed, since $\mathbb{F} \subseteq \mathbb{K}$ we know that any automorphism fixing $\mathbb{K}$ also fixes $\mathbb{F}$. On the other hand, let $H \subseteq G$ be any subgroup and consider the set

$$\mathrm{Fix}_{\mathbb{E}}(H) \coloneqq \{a \in \mathbb{E} : \sigma(a) = a \text{ for all } \sigma \in H\} \subseteq \mathbb{E}.$$

This set contains $\mathbb{F}$ because $\sigma(a) = a$ for all $\sigma \in G$ and because $H \subseteq G$. I claim that $\mathbb{F} \subseteq \mathrm{Fix}_{\mathbb{E}}(H) \subseteq \mathbb{E}$ is an intermediate field, called the *fixed subfield* of H.

Proof. Consider any $\sigma \in H$. Since $\sigma : \mathbb{E} \to \mathbb{E}$ is a field automorphism we must have $\sigma(0) = 0$ and $\sigma(1) = 1$, which implies that $0, 1 \in \mathrm{Fix}_{\mathbb{E}}(H)$. Then for all $a, b \in \mathrm{Fix}_{\mathbb{E}}(H)$ we have

$$\sigma(a+b) = \sigma(a) + \sigma(b) = a + b \qquad \text{and} \qquad \sigma(ab) = \sigma(a)\sigma(b) = ab,$$

which implies that $a + b, ab \in \mathrm{Fix}_{\mathbb{E}}(H)$. Finally, for all $a \in \mathbb{E} - \{0\}$ we have

$$\sigma(a^{-1}) = \sigma(a)^{-1} = a^{-1},$$

which implies that $a^{-1} \in \mathrm{Fix}_{\mathbb{E}}(H)$. □

In summary, we have a pair of functions between the lattices $\mathscr{L}(\mathbb{E},\mathbb{F})$ and $\mathscr{L}(G)$. In the language of Week 5, I claim that these two functions form an *abstract Galois connection.*[104] Actually it will be a Galois connection after we reverse the partial order on one of the posets. We will do this with the superscript "op" for "opposite":

$$\mathrm{Gal}(\mathbb{E}/-) : \mathscr{L}(\mathbb{E},\mathbb{F}) \rightleftarrows \mathscr{L}(G)^{\mathrm{op}} : \mathrm{Fix}_{\mathbb{E}}(-).$$

Proof. Consider any intermediate field $\mathbb{F} \subseteq \mathbb{K} \subseteq \mathbb{E}$ and any subgroup $H \subseteq G$. By the definition of Galois connection we need to show that

$$\mathbb{K} \subseteq \mathrm{Fix}_{\mathbb{E}}(H) \iff \mathrm{Gal}(\mathbb{E}/\mathbb{K}) \supseteq H.$$

And this is immediate from the definitions of $\mathrm{Gal}(\mathbb{E}/-)$ and $\mathrm{Fix}_{\mathbb{E}}(-)$:

$$\begin{aligned} \mathbb{K} \subseteq \mathrm{Fix}_{\mathbb{E}}(H) &\iff \forall a \in \mathbb{K}, a \in \mathrm{Fix}_{\mathbb{E}}(H) \\ &\iff \forall a \in \mathbb{K}, \forall \sigma \in H, \sigma(a) = a \end{aligned}$$

[104] This is a good time to remind yourself of the definition.

$$\begin{aligned}
&\iff \forall \sigma \in H, \forall a \in \mathbb{K}, \sigma(a) = a \\
&\iff \forall \sigma \in H, \sigma \in \mathrm{Gal}(\mathbb{E}/\mathbb{K}) \\
&\iff H \subseteq \mathrm{Gal}(\mathbb{E}/\mathbb{K}).
\end{aligned}$$

□

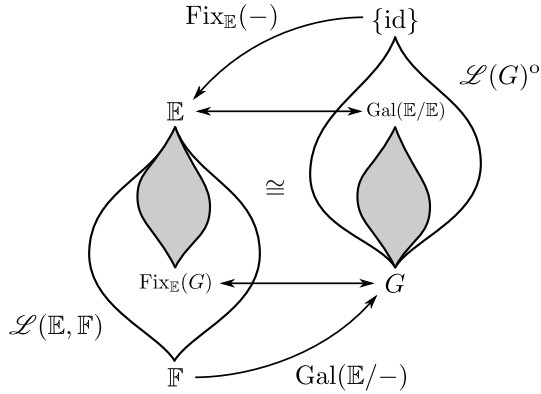
Let me remind you what we get from this. It follows for purely formal (i.e., "trivial") reasons that these functions restrict to an isomorphism between certain subposets

$$\mathrm{Gal}(\mathbb{E}/-) : \mathscr{L}(\mathbb{E},\mathbb{F})' \xleftrightarrow{\sim} (\mathscr{L}(G)')^{\mathrm{op}} : \mathrm{Fix}_{\mathbb{E}}(-),$$

where the subposets are defined by

$$\begin{aligned}
\mathscr{L}(\mathbb{E},\mathbb{F})' &:= \{\mathbb{F} \subseteq \mathbb{K} \subseteq \mathbb{E} : \mathrm{Fix}_{\mathbb{E}}(\mathrm{Gal}(\mathbb{E}/\mathbb{K})) = \mathbb{K}\}, \\
\mathscr{L}(G)' &:= \{H \subseteq G : \mathrm{Gal}(\mathbb{E}/\mathrm{Fix}_{\mathbb{E}}(H)) = H\}.
\end{aligned}$$

Here's a picture:



Actually this picture is a bit too loose because we always have $\mathrm{Gal}(\mathbb{E}/\mathbb{E}) = \{id\}$. But never mind. The Fundamental Theorem says that under certain nice conditions (when $\mathbb{E} \supseteq \mathbb{F}$ is a splitting field for some polynomial) then the correspondence is as tight as possible.

The Fundamental Theorem of Galois Theory. Let $\mathbb{E} \supseteq \mathbb{F}$ be the splitting field of some polynomial $f(x) \in \mathbb{F}[x]$.[105] and let $G = \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ be the Galois group. Then:

(1) The Galois connection $\mathrm{Gal}(\mathbb{E}/-) : \mathscr{L}(\mathbb{E},\mathbb{F}) \leftrightarrows \mathscr{L}(G)^{\mathrm{op}} : \mathrm{Fix}_{\mathbb{E}}(-)$ is actually a **bijection**. That is, for all intermediate fields $\mathbb{E} \supseteq \mathbb{K} \supseteq \mathbb{F}$ and

[105] Later we will see that we also need to restrict our attention to certain "perfect" kinds of fields $\mathbb{F}$. Luckily, every field you have ever seen is "perfect".

for all subgroups $H \subseteq G$ we have

$$\mathrm{Fix}_{\mathbb{E}}(\mathrm{Gal}(\mathbb{E}/\mathbb{K})) = \mathbb{K} \qquad \text{and} \qquad \mathrm{Gal}(\mathbb{E}/\mathrm{Fix}_{\mathbb{E}}(H)) = H.^{106}$$

(2) For any pair $\mathbb{K} = \mathrm{Fix}_{\mathbb{E}}(H)$ and $H = \mathrm{Gal}(\mathbb{E}/\mathbb{K})$ we have

$$\#\{\text{cosets of } H \text{ in } G\} = \#(G/H) = [\mathbb{K}/\mathbb{F}] = \dim(\mathbb{K} \text{ as a vector space over } \mathbb{F}).$$

(3) Furthermore, we have

$$\mathbb{K} \supseteq \mathbb{F} \text{ is a Galois field extension} \iff H \trianglelefteq G \text{ is a normal subgroup,}$$

in which case the quotient group is isomorphic to the Galois group:

$$\frac{G}{H} = \frac{\mathrm{Gal}(\mathbb{E}/\mathbb{F})}{\mathrm{Gal}(\mathbb{E}/\mathbb{K})} \cong \mathrm{Gal}(\mathbb{K}/\mathbb{F}).$$

///

[**Remark:** The notation "G/H" for the **set** of cosets is motivated by Lagrange's Theorem:

$$\#(G/H) = \#G \,/\, \#H.$$

The notation "$\mathbb{K}/\mathbb{F}$"[107] for $\mathbb{K}$ as a **vector space** over $\mathbb{F}$ is motivated by a similar theorem, called Dedekind's Tower Law:

$$[\mathbb{E}/\mathbb{K}] = [\mathbb{E}/\mathbb{F}] \,/\, [\mathbb{K}/\mathbb{F}].$$

You will prove Dedekind's Law on the homework. In the situation where $\mathbb{E} \supseteq \mathbb{F}$ is a splitting field for some polynomial over $\mathbb{F}$, it follows from the Fundamental Theorem that Lagrange's Theorem and Dedekind's Law are equivalent.]

It is easy to imagine how the proof of Galois' Solvability Theorem might following from the Fundamental Theorem, since the solvability of field extensions and groups are both defined by the existence of certain kinds of chains. Sadly, there are still some difficulties. To go between solvable chains of subfields and solvable chains of subgroups we need to be careful about roots of unity. Never mind the details right now. Let me just show you an interesting example.

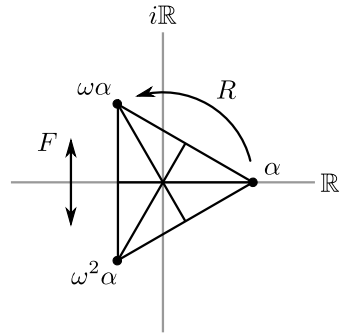
The Smallest Interesting Example. We saw last time that the splitting field $\mathbb{E} \supseteq \mathbb{Q}$ for the polynomial $x^3 - 2$ is given by

$$\mathbb{E} = \mathbb{Q}(\omega, \alpha) = \{a + b\alpha + c\alpha^2 + d\omega + e\omega\alpha + f\omega\alpha^2 : a, b, c, d, e, f \in \mathbb{Q}\} \supseteq \mathbb{Q}.$$

[106] As I mentioned above, the equation $\mathrm{Gal}(\mathbb{E}/\mathrm{Fix}_{\mathbb{E}}(H)) = H$ holds for any field $\mathbb{E}$ and for any finite group of automorphisms $H \subseteq \mathrm{Aut}(\mathbb{E})$. The proof only depends on Artin's Fixed Field Lemma (which, however, we did not prove in full generality). The other equation $\mathrm{Fix}_{\mathbb{E}}(\mathrm{Gal}(\mathbb{E}/\mathbb{K})) = \mathbb{K}$ is more interesting.

[107] The standard notation is $\mathbb{K} : \mathbb{F}$, but I think that the use of a colon to suggest a quotient doesn't read well to modern eyes. So I decided to update the notation.

Furthermore, we saw that the Galois group is isomorphic to the group of all permutations of the roots $\{\alpha, \omega\alpha, \omega^2\alpha\}$. To be concrete, let's identify this group with the dihedral group of symmetries of the equilateral triangle, as in the following picture:



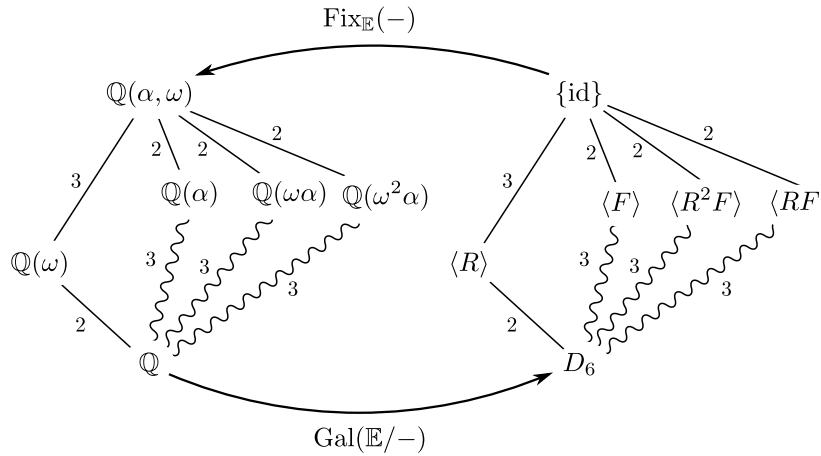
Recall that the six elements of this group can be expressed as follows:

$$D_6 = \{id, R, R^2, F, RF, R^2F\}.$$

The "rotation" $R : \mathbb{Q}(\omega, \alpha) \to \mathbb{Q}(\omega, \alpha)$ is defined on the generators by $R(\alpha) = \omega\alpha$ and $R(\omega) = \omega$, while the "reflection" $F : \mathbb{Q}(\omega, \alpha) \to \mathbb{Q}(\omega, \alpha)$ is defined by $F(\alpha) = \alpha$ and $F(\omega) = \omega^2$. (In fact, F is the restriction of "complex conjugation" to the subfield $\mathbb{E} \subseteq \mathbb{C}$.) By working with the coordinates $a, b, c, d, e, f \in \mathbb{Q}$ one can compute the fixed fields of the cyclic subgroups generated by R and F:

$$\mathrm{Fix}_{\mathbb{E}}(\langle R \rangle) = \mathbb{Q}(\omega) \qquad \text{and} \qquad \mathrm{Fix}_{\mathbb{E}}(\langle F \rangle) = \mathbb{Q}(\alpha).$$

With a bit more work we obtain the following bijection:



Since the finite group D_6 has finitely many subgroups, it follows from the Fundamental Theorem that the field extension $\mathbb{Q}(\omega, \alpha) \supseteq \mathbb{Q}$ has **finitely many intermediate fields**, which is certainly not obvious. I have labeled the edges

with the degree of the field extension (left) or the number of cosets (right). The Fundamental Theorem says that these numbers are equal.

Finally, I have labeled the **non-normal subgroups** with squiggly lines. These correspond on the left to field extensions that are **not splitting fields** for any polynomial. We say that the group D_6 is *solvable* because of the existence of the chain of normal subgroups

$$D_6 \trianglerighteq \langle R \rangle \trianglerighteq \{id\}$$

with abelian quotients $D_6/\langle R \rangle \cong \mathbb{Z}/2\mathbb{Z}$ and $\langle R \rangle/\{id\} \cong \mathbb{Z}/3\mathbb{Z}$. It is less clear what is special about the corresponding chain of fields. We might ask:

> Why is the chain $\mathbb{Q} \subseteq \mathbb{Q}(\omega) \subseteq \mathbb{Q}(\omega, \alpha)$ better than the chain $\mathbb{Q} \subseteq \mathbb{Q}(\alpha) \subseteq \mathbb{Q}(\omega, \alpha)$?

I'll just let you puzzle over this for now. Later we will see that the key to a "good" chain of field extensions is to

adjoin the roots of unity first.

///

My goal is to prove all of this before the end of the course. We will do this by building everything up slowly from the basic theory of "commutative rings". The study of commutative rings (which is called "commutative algebra") is an absolutely huge subject,[108] so we will only cover the material that is relevant to Galois' theorem.

Exercises

14.A Dedekind's Tower Law

Let $\mathbb{E} \supseteq \mathbb{F}$ be a field extension.

(a) There is an obvious "multiplication function" $\mathbb{F} \times \mathbb{E} \to \mathbb{E}$ defined by the rule $(a, b) \mapsto ab$. Verify that this multiplication makes $\mathbb{E}$ into a **vector space** over $\mathbb{F}$. We will denote this vector space by $\mathbb{E}/\mathbb{F}$. Its dimension is called the *degree* of the extension:

$$[\mathbb{E}/\mathbb{F}] := \dim(\mathbb{E}/\mathbb{F}).$$

[**Remark:** The fractional notation is a mnemonic device. Do not take it literally.]

[108] Eisenbud's textbook on *Commutative Algebra (with a View Toward Algebraic Geometry)* is 800 pages long and it assumes everything that we will say in this course as a pre-requisite.

(b) Now let $\mathbb{E} \supseteq \mathbb{K} \supseteq \mathbb{F}$ be any intermediate field. Prove that the degrees of the three extensions satisfy

$$[\mathbb{E}/\mathbb{F}] = [\mathbb{E}/\mathbb{K}] \cdot [\mathbb{K}/\mathbb{F}].$$

[Hint: Let $\{\alpha_i\}_i$ be a basis for $\mathbb{K}/\mathbb{F}$ and let $\{\beta_j\}_j$ be a basis for $\mathbb{E}/\mathbb{K}$. Prove that the set $\{\alpha_i\beta_j\}_{i,j}$ is a basis for $\mathbb{E}/\mathbb{F}$.] Does this remind you of Lagrange's Theorem?

14.B Axioms for the Galois Group

In this problem you will show that the hypothesis of invertibility is redundant in the definition of the Galois group. Let $\mathbb{E} \supseteq \mathbb{F}$ be a field extension and let $\sigma : \mathbb{E} \to \mathbb{E}$ be any function satisfying

$$\sigma(a+b) = \sigma(a) + \sigma(b) \quad \text{and} \quad \sigma(ab) = \sigma(a)\sigma(b) \quad \text{for all } a, b \in \mathbb{E}.$$

(a) Prove that σ is necessarily injective.

(b) If $\sigma(a) = a$ for all $a \in \mathbb{F}$, prove that $\sigma : \mathbb{E}/\mathbb{F} \to \mathbb{E}/\mathbb{F}$ is a linear function.

(c) If $\sigma(a) = a$ for all $a \in \mathbb{F}$ and if $[\mathbb{E}/\mathbb{F}] < \infty$,[109] combine parts (a) and (b) to prove that σ is necessarily bijective. [Hint: Use the Rank-Nullity Theorem.]

14.C Quadratic Field Extensions

Let $\mathbb{E} \supseteq \mathbb{F}$ be a field extension and let $\alpha \in \mathbb{E}$ be any element such that $\alpha \notin \mathbb{F}$ and $\alpha^2 \in \mathbb{F}$. Consider the subfield $\mathbb{F} \subseteq \mathbb{F}(\alpha) \subseteq \mathbb{E}$ generated by α.

(a) Prove that the set $\{1, \alpha\} \subseteq \mathbb{F}(\alpha)/\mathbb{F}$ is linearly independent.

(b) Prove that $\{1, \alpha\} \subseteq \mathbb{F}(\alpha)/\mathbb{F}$ is a spanning set. [Hint: Prove that $\{a+b\alpha : a, b \in \mathbb{F}\} \subseteq \mathbb{E}$ is a subfield by "rationalizing the denominator".] It follows that $\{1, \alpha\}$ is a basis for $\mathbb{F}(\alpha)/\mathbb{F}$ and hence $[\mathbb{F}(\alpha)/\mathbb{F}] = 2$.

(c) Use Dedekind's Tower Law to prove that there **does not exist** any intermediate field:

$$\mathbb{F} \subsetneq \mathbb{K} \subsetneq \mathbb{F}(\alpha).$$

(d) Prove that the function $\tau : \mathbb{F}(\alpha) \to \mathbb{F}(\alpha)$ defined by $\tau(a + b\alpha) \coloneqq a - b\alpha$ is a field automorphism. We call this operation *conjugation*.

14.D A Biquadratic Field Extension

Let $\sqrt{2}, \sqrt{3} \in \mathbb{R}$ be some specific square roots of 2 and 3, and consider the subfields $\mathbb{Q}(\sqrt{2}), \mathbb{Q}(\sqrt{3}) \subseteq \mathbb{R}$. We saw in class that the union $\mathbb{Q}(\sqrt{2}) \cup \mathbb{Q}(\sqrt{3}) \subseteq \mathbb{R}$ is **not** a subfield. So instead we will consider the join/compositum subfield:

$$\mathbb{Q}(\sqrt{2}) \vee \mathbb{Q}(\sqrt{3}) = \mathbb{Q}(\sqrt{2})\mathbb{Q}(\sqrt{3}) = \mathbb{Q}(\sqrt{2})(\sqrt{3}) = \mathbb{Q}(\sqrt{2}, \sqrt{3}) \subseteq \mathbb{R}.$$

[109] This will be true if $\mathbb{E}$ is the splitting field of a polynomial over $\mathbb{F}$.

(a) *A Basis.* Prove that elements of this field have the following explicit form:

$$\mathbb{Q}(\sqrt{2},\sqrt{3}) = \{a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} : a, b, c, d \in \mathbb{Q}\}.$$

[Hint: It's quite tricky to prove directly that the set on the right is a field. Use Dedekind's Tower Law for an indirect proof.]

(b) *The Galois Group.* Let $\sigma : \mathbb{Q}(\sqrt{2},\sqrt{3}) \to \mathbb{Q}(\sqrt{2},\sqrt{3})$ be any field automorphism. Prove that σ necessarily fixes the prime subfield $\mathbb{Q}$, and hence that σ is uniquely determined by the two values $\sigma(\sqrt{2})$ and $\sigma(\sqrt{3})$. Write down all of the possibilities and observe that you get a group isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

(c) *A Primitive Element.* Define the number $\gamma = \sqrt{2} + \sqrt{3}$ and prove that

$$\mathbb{Q}(\sqrt{2},\sqrt{3}) = \mathbb{Q}(\gamma).$$

[Hint: One inclusion is easy. For the other inclusion, expand γ^3 to show that $\sqrt{2}$ and $\sqrt{3}$ are in the field $\mathbb{Q}(\gamma)$.] You know from part (a) that $[\mathbb{Q}(\sqrt{2},\sqrt{3})/\mathbb{Q}] = 4$. It follows that the five elements $1, \gamma, \gamma^2, \gamma^3, \gamma^4 \in \mathbb{Q}(\sqrt{2},\sqrt{3})$ are **not** linearly independent over $\mathbb{Q}$, hence γ must satisfy a quartic equation of the form

$$a + b\gamma + c\gamma^2 + d\gamma^3 + e\gamma^4 = 0 \quad \text{for some nontrivial } a, b, c, d, e \in \mathbb{Q}.$$

Find this equation. [Hint: Expand γ^4 and work down.] If $\sigma : \mathbb{Q}(\sqrt{2},\sqrt{3}) \to \mathbb{Q}(\sqrt{2},\sqrt{3})$ is any field automorphism, prove that $\sigma(\gamma)$ is another solution of the same equation. Finally, use part (b) to obtain all four roots of the equation.

[**Remark:** In this problem we observe that there exists a bijection between the elements of the Galois group and the roots of the "minimal polynomial" for a "primitive element". This is actually how Galois **defined** the Galois group. But then one has to prove that different primitive elements lead to isomorphic groups. See Tignol's book for details. In order to prove that Dedekind's version of the Galois group is well-defined we will show later that the splitting field of a polynomial is unique up to isomorphism.]

Week 15

15.1 Definition of Rings

In this course I have followed a mostly chronological development of abstract algebra. The study of groups began with Galois in the 1820s and the study of fields began with Dedekind in the 1870s. The first work to study algebra from a purely axiomatic point of view was Ernst Steinitz' *Algebraic Theory of Fields* (1910). Inspired by this, several authors considered a weaker structure called "rings".[110] Abstract ring theory was standardized by Emmy Noether in the 1920s. Here is the modern definition.

Definition of Rings and Subrings/Extensions. Let R be a set equipped with two binary operations $+, \times : R \times R \to R$ and two special elements $0, 1 \in R$. We call this structure a *(commutative) ring* if the following axioms hold:

(R1) $(R, +, 0)$ is an abelian group.

(R2) $(R, \times, 1)$ is a (commutative) monoid.[111]

(R3) For all $a, b, c \in R$ we have $a(b + c) = ab + ac$.

Now let $S \subseteq R$ be any subset. We say that S is a *subring* of R if the following properties hold:

- The special elements $0, 1$ are in S.
- For all $a, b \in S$ we have $a \pm b \in S$ and $ab \in S$.

Equivalently, we say that R is a *ring extension* of S. ///

Remarks:

- The distributive law (R3) tells us how the two binary operations $+$ and $\times$ interact. From this we obtain some basic rules mixing additive and

[110] The word *ring* (or *Zahlring*) comes from David Hilbert's *Zahlbericht* (1897). The word "Zahlbericht" means "number report" and "Zahlring" means "number ring". Nobody knows why he chose the word "ring".

[111] This means that multiplication is an associative operation with identity element 1.

multiplicative concepts:

$$0a = 0,$$
$$a(-b) = (-a)b = -(ab),$$
$$(-a)(-b) = ab.$$

You will prove these on 15.B.

- We are are allowed to have $1 = 0$ in a ring. But in this case we also have

$$a = 1a = 0a = 0 \quad \text{for all } a \in R.$$

This structure is called the *zero ring* $R = 0$.

- If $R \neq 0$ (i.e., if $1 \neq 0$) then we define the set

$$R^\times \coloneqq \{a \in R : \exists b \in R, ab = 1\} \subseteq R - \{0\}.$$

It follows that $(R^\times, \times, 1)$ is a group, called the *group of units* of the ring.

- If $R^\times = R - \{0\}$ then we say that R is a *field*. Note that this implies $1 \neq 0$.[112]
- As with subgroups and subfields, the intersection of any collection of subrings is again a subring, and we can use this to define the "subring genererated by a subset". For example, if $E \supseteq R$ is any ring extension and if $\alpha \in E$ is any element then we will use the following "square-bracket" notation:

$$R[\alpha] \coloneqq \text{ the smallest subring of } E \text{ containing } R \cup \{\alpha\}.$$

More on this later. ///

The main innovation (CHANGE THIS) of Emmy Noether was to recognize the importance of homomorphisms in the study of rings. These are much more interesting than homomorphisms between fields.

Definition of Ring Homomorphisms. Let R and S be rings and let $\varphi : R \to S$ be any function. We say that φ is a *ring homomorphism* if the following properties hold:

(H1) $\varphi(a + b) = \varphi(a) + \varphi(b)$,

(H2) $\varphi(ab) = \varphi(a)\varphi(b)$,

(H3) $\varphi(1) = 1$.

///

Remarks:

[112] There is no "field with one element". I don't know why.

- The first axiom (H1) says that $\varphi : (R, +, 0) \to (S, +, 0)$ is a homomorphism of groups. Let me refresh your memory why this implies that

$$\varphi(0) = 0 \qquad \text{and} \qquad \varphi(-a) = -\varphi(a) \text{ for all } a \in R.$$

Proof. First note that $\varphi(0) = \varphi(0 + 0) = \varphi(0) + \varphi(0)$. Now subtract $\varphi(0)$ from both sides to obtain $0 = \varphi(0)$. Then for any $a \in R$ we have

$$0 = \varphi(0) = \varphi(a - a) = \varphi(a) + \varphi(-a),$$

which implies that $\varphi(-a) = -\varphi(a)$. □

- Unfortunately, the second axiom (H2) does **not** imply that $\varphi(1) = 1$. Indeed, the proof from above does not work because we are not allowed to divide. For this reason we must include axiom (H3). Alternatively, we could combine axioms (H2) and (H3) by saying that $\varphi : (R, \times, 1) \to (S, \times, 1)$ is a *homomorphism of monoids.* ///

Recall from last semester that a homomorphism of groups $\varphi : G \to H$ leads to a Correspondence Theorem and three Isomorphism Theorems. Our next goal is to extend all of this structure to rings. However, we will find that the situation is a bit more complicated.

First of all, we note that a subring is the same thing as the image of a homomorphism.

Subring = Image of a Ring Homomorphism. Let S be a ring and let $S' \subseteq S$ be a subset. I claim that $S' \subseteq S$ is a subring if and only if there exists a ring homomorphism $\varphi : R \to S$ such that $\operatorname{im} \varphi = S'$.

Proof. First let $\varphi : R \to S$ be a ring homomorphism and consider the image

$$\operatorname{im} \varphi \coloneqq \{\varphi(a) : a \in R\} \subseteq S.$$

Since $\varphi(0) = 0$ and $\varphi(1) = 1$ we find that $0, 1 \in \operatorname{im} \varphi$. Furthermore, if $\varphi(a), \varphi(b) \in \operatorname{im} \varphi$ are any two elements of the image then we have

$$\varphi(a) \pm \varphi(b) = \varphi(a \pm b) \in \operatorname{im} \varphi \qquad \text{and} \qquad \varphi(a)\varphi(b) = \varphi(ab) \in \operatorname{im} \varphi,$$

as desired. Conversely, let $S' \subseteq S$ be any subring and let $id|_{S'} : S' \to S$ be the restriction of the identity function $id : S \to S$. Clearly $id_{S'}$ is a ring homomorphism with image S'. □

Digression: The urge to translate all concepts (such as "subring") into the language of homomorphisms ultimately leads to the subject of *category theory.*[113]

[113] The language of categories emerged in the 1940s and 1950s in order to clarify the subject of topology. Since our course is focused on the years 1830–1930 we will not use this language. However, this course does contain two examples of categories ideas. One is the idea of a "universal property", which is a special case of limits and colimits. The other is the idea of an "abstract Galois connection", which is a special example of adjoint functors.

I will not actually define categories in this class, but I will point out a few category-theoretic ideas. Here's one.

Initial and Final Rings. Let R be any ring. Then there exists a unique ring homomorphism from the ring of integers and a unique ring homomorphism to the zero ring:

$$\mathbb{Z} \xrightarrow{\exists!} R \qquad \text{and} \qquad R \xrightarrow{\exists!} 0.$$

[**Jargon:** We say that $\mathbb{Z}$ is the *initial object* and 0 is the *final object* in the category of rings.]

Proof. There is a unique function $\varphi : R \to 0$ and this function is a ring homomorphism. On the other hand, recall that for any $a \in R$ and $n \in \mathbb{Z}$ we have defined the following notation:

$$n \cdot a \coloneqq \begin{cases} \overbrace{a + a + \cdots + a}^{n \text{ times}} & \text{if } n \geq 1, \\ 0 & \text{if } n = 0, \\ \underbrace{-a - a - \cdots - a}_{-n \text{ times}} & \text{if } n \leq 1. \end{cases}$$

When discussing cyclic groups we proved by induction that the map $n \mapsto n \cdot a$ is the unique group homomorphism $(\mathbb{Z}, +, 0) \to (R, +, 0)$ sending $1 \in \mathbb{Z}$ to $a \in R$. Now let $\varphi : \mathbb{Z} \to R$ be any ring homomorphism. In particular, since $\varphi : (\mathbb{Z}, +, 0) \to (R, +, 0)$ is a group homomorphism we must have $\varphi(n) = n \cdot 1$ for all $n \in \mathbb{Z}$. It only remains to prove that the function $\varphi(n) \coloneqq n \cdot 1$ preserves multiplication, and this must be done by induction. Here is the key step:

$$\begin{aligned} \varphi(m)\varphi(n+1) &= (m \cdot 1)\,[(n+1) \cdot 1] \\ &= (m \cdot 1)(n \cdot 1 + 1) \\ &= (m \cdot 1)(n \cdot 1) + m \cdot 1 \\ &= (mn) \cdot 1 + m \cdot 1 && \text{induction on } n \\ &= (mn + m) \cdot 1 \\ &= [m(n+1)] \cdot 1 = \varphi(m(n+1)). \end{aligned}$$

□

15.2 General Structure of Rings

Last time we saw that a subring is the same thing as the image of a ring homomorphism. As with group homomorphisms, there is also a notion of kernel for ring homomorphisms. This concept was implicit in the work of Kummer and Dedekind on unique factorization. (We will discuss this below.) Emmy Noether synthesized the ideas of number theory and algebraic geometry to obtain the modern definition.

Ideal = Additive Kernel of a Ring Homomorphism. Consider any ring homomorphism $\varphi : R \to S$. In particular this defines a homomorphism of additive groups $\varphi : (R, +, 0) \to (S, +, 0)$. Let $\ker \varphi \subseteq R$ denote the kernel of this group homomorphism:

$$\ker \varphi \coloneqq \{a \in R : \varphi(a) = 0\}.$$

Then the First Isomorphism Theorem for groups tells us that the (well-defined) function $a + \ker \varphi \mapsto \varphi(a)$ is an isomorphism of additive groups:

$$\left(\frac{R}{\ker \varphi}, +, 0 + \ker \varphi\right) \cong (\operatorname{im} \varphi, +, 0).$$

However, since $\varphi : R \to S$ satisfies the extra properties $\varphi(ab) = \varphi(a)\varphi(b)$ and $\varphi(1) = 1$ we know from the above that $\operatorname{im} \varphi \subseteq S$ actually a ring. By pulling back this structure we conclude that the quotient **group** $R/\ker \varphi$ is also a **ring** with multiplication defined by

$$(a + \ker \varphi)(b + \ker \varphi) \coloneqq (ab + \ker \varphi).$$

The fact that this operation is well-defined reflects a certain structural property of the kernel, which is analogous to the "normal subgroup" property of group kernels. In order to motivate the following definition I will state it as a theorem.

Theorem (Definition of Ideals and Quotient Rings). Let $(R, +, \times, 0, 1)$ be a ring and let $I \subseteq (R, +, 0)$ be any additive subgroup. Then the following are equivalent:

(I1) For all $a \in I$ and $b \in R$ we have $ab \in I$. In this case we say that I is an *ideal of* R.

(I2) There exists a ring homomorphism $\varphi : R \to S$ with $I = \ker \varphi$.

///

Proof. First we do the easy direction.

(I2)⇒(I1): Let $\varphi : R \to S$ be a ring homomorphism and consider the kernel $\ker \varphi \coloneqq \{a \in R : \varphi(a) = 0\}$. Then for all $a \in \ker \varphi$ and $b \in R$ we have

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

and hence $ab \in \ker \varphi$. In other words, $\ker \varphi \subseteq R$ is an ideal.

(I1)⇒(I2): Let $I \subseteq R$ be an ideal and consider the additive quotient group $(R/I, +, 0 + I)$. I claim that the following "multiplication operation" on R/I is well-defined:

$$(a + I)(b + I) \coloneqq (ab + I).$$

Indeed, suppose that we have $a + I = a' + I$ and $b + I = b' + I$. By definition this means that $a - a'$ and $b - b'$ are in I. But then we have

$$ab - a'b' = ab - ab' + ab' - a'b' = a(b - b') + (a - a')b' \in I$$

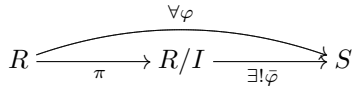
and hence $ab + I = a'b' + I$ as desired. It is easy to check that this defines a ring structure:

$$(R/I, +, \times, 0 + I, 1 + I).$$

Finally, consider the function $\pi : R \to R/I$ defined by $a \mapsto a + I$. It is easy to check that this function is a ring homomorphism with $\ker \pi = I$. □

Remarks:

- The construction of quotient rings appeared on 6.B last semester. I decided to repeat it here to emphasize the connection with ring homomorphisms.
- In category theory we emphasize that the quotient ring is really a pair $(R/I, \pi)$ where $\pi : R \to R/I$ is the *canonical projection* with $\ker \pi = I$. It satisfies the following so-called universal property:



  In words: For any ring homomorphism $\varphi : R \to S$ with $I \subseteq \ker \varphi$, there exists a unique ring homomorphism $\bar{\varphi} : R/I \to S$ satisfying $\varphi = \bar{\varphi} \circ \pi$. Feel free to ignore this remark.
- The word "ideal" comes from Ernst Kummer's concept of "ideal numbers". He introduced this concept in order to recover some version of unique prime factorization in rings such as $\mathbb{Z}[\sqrt{-5}]$ where the literal version fails. Dedekind shortened the name from "ideal number" to "ideal".

///

The whole point of the above theorem/definition was to generalize the First Isomorphism Theorem from additive groups to rings. Here is the statement.

The First Isomorphism Theorem for Rings. Let $\varphi : R \to S$ be any ring homomorphism. Since $\ker \varphi \subseteq R$ is an ideal we may consider the quotient ring $R/\ker \varphi$. Then the natural map $a + \ker \varphi \mapsto \varphi(a)$ is a well-defined isomorphism of rings:

$$R/\ker \varphi \cong \operatorname{im} \varphi.$$

///

And what about the Correspondence Theorem and the Second/Third Isomorphism Theorems? This is a bit more complicated because there are **three** lattices naturally associated to a ring:

- The lattice of additive subgroups.
- The lattice of subrings.
- The lattice of ideals.

Since subrings and ideals are both examples of additive subgroups, it seems most reasonable to use the notation $\mathscr{L}$ for the lattice of additive subgroups. Thus for any ring R we define

$$\mathscr{L}(R) \coloneqq \{\text{the lattice of subgroups of } (R,+,0)\}.^{114}$$

For any subgroups $A, B \subseteq R$ recall from last semester that the meet and join are given by the intersection and the sum, respectively:

$$A \wedge B = A \cap B \qquad \text{and} \qquad A \vee B = A + B \coloneqq \{a+b : a \in A, b \in B\}.$$

Furthermore, if $I \subseteq R$ is any subgroup (probably an ideal) then we will use the notation

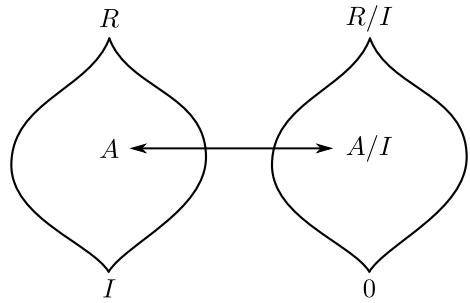
$$\mathscr{L}(R,I) \coloneqq \{\text{subgroups } A \text{ such that } I \subseteq A \subseteq R\}.$$

Then we have the following theorem.

The Correspondence Theorem for Rings. Let R be a ring and let $I \subseteq R$ be an ideal. In particular, I is an additive subgroup, so the Correspondence Theorem for Groups gives us an isomorphism of lattices:

$$\begin{array}{ccc} \mathscr{L}(R,I) & \xrightarrow{\sim} & \mathscr{L}(R/I) \\ A & \mapsto & A/I. \end{array}$$

Here is a picture:



So far this is just group theory. To incorporate ring theory, we first observe that R/I is a ring because $I \subseteq R$ is an ideal. Then for any subgroup $I \subseteq A \subseteq R$ I claim that

$$A \subseteq R \text{ is a subring} \qquad \Longleftrightarrow \qquad A/I \subseteq R/I \text{ is a subring}$$

[114] I apologize that this choice conflicts with my use of $\mathscr{L}(\mathbb{F})$ for the lattice of subfields of a field $\mathbb{F}$. Hopefully this will cause no confusion.

and

$$A \subseteq R \text{ is an ideal} \qquad \Longleftrightarrow \qquad A/I \subseteq R/I \text{ is an ideal.}$$

///

Proof. Last semester we gave a lengthy proof of the Correspondence Theorem for Groups. Luckily we don't have to prove it again. You will prove the final statements about subrings and ideals in Exercise 15.E, where you will also prove ring versions of the Second and Third Isomorphism Theorems. □

Exercises

15.A One Step Ideal Test

Let $(R, +, \times, 0, 1)$ be a ring and let $S \subseteq R$ be any subset. Prove that S is an ideal if and only if for all $a, b \in S$ and $r \in R$ we have $a + rb \in S$.

15.B Addition vs. Multiplication

Prove that following properties hold in any ring.

(a) $0a = 0$,

(b) $a(-b) = (-a)b = -(ab)$,

(c) $(-a)(-b) = ab$.

15.C The Characteristic of a Ring

Let R be a ring and let $R' \subseteq R$ be the smallest subring. Recall that there exists a unique ring homomorphism $\iota_R : \mathbb{Z} \to R$ from the integers.

(a) Prove that $R' \cong \mathbb{Z}/n\mathbb{Z}$ for some integer $n \geq 0$, which we call the *characteristic* of R:

$$\text{char}(R) = n.$$

[Hint: Apply the First Isomorphism Theorem to ι_R.]

(b) If $\varphi : R \to S$ is any ring homomorphism prove that $\text{char}(S)$ divides $\text{char}(R)$. [Hint: By uniqueness we know that $\iota_S = \varphi \circ \iota_R$. Consider the kernel.]

(c) Next let R be an *integral domain*, which means that R has no *zero-divisors*:

$$\forall a, b \in R, (ab = 0) \Rightarrow (a = 0 \text{ or } b = 0).$$

In this case prove that $\text{char}(R) = 0$ or $\text{char}(R) = p$ for some prime p.

(d) Finally, let $\mathbb{F}$ ba a field and let $\mathbb{F}' \subseteq \mathbb{F}$ be the smallest subfield. Prove that

$$\mathbb{F}' \cong \mathbb{Q} \qquad \text{or} \qquad \mathbb{F}' \cong \mathbb{Z}/p\mathbb{Z} \text{ for some prime } p.$$

15.D The Chinese Remainder Theorem, Part II

Let R be a ring. For any ideals $I, J \subseteq R$ we define the *product ideal*:

$$IJ \coloneqq \text{intersection of all ideals that contain } \{ab : a \in I, b \in J\}.$$

(a) Prove that $IJ \subseteq I \cap J$.

(b) We say that $I, J \subseteq R$ are *coprime* if $I + J = R$. In this case show that $I \cap J \subseteq IJ$, and hence $IJ = I \cap J$. [Hint: Since $1 \in I + J$ we have $1 = x + y$ for some $x \in I$ and $y \in J$.]

(c) If $I, J \subseteq R$ are coprime, prove that the obvious map $(a + IJ) \mapsto (a + I, a + J)$ defines an isomorphism of rings:

$$\frac{R}{IJ} \cong \frac{R}{I} \times \frac{R}{J}.$$

[Hint: The hardest part is surjectivity. Use the same trick that you used when $R = \mathbb{Z}$.]

15.E Ring Isomorphism Theorems

Let R be a ring and let $I \subseteq R$ be an ideal.

(a) For any additive subgroup $I \subseteq S \subseteq R$ prove that

$$S \subseteq R \text{ is a subring} \quad \iff \quad S/I \subseteq R/I \text{ is a subring.}$$

(b) For any subring $S \subseteq R$ prove that we have an isomorphism of rings:

$$\frac{S}{S \cap I} \cong \frac{S + I}{I}.$$

[Hint: Consider the ring homomorphism $\varphi : S \to R/I$ defined by $\varphi(a) = a + I$.]

(c) For any additive subgroup $I \subseteq J \subseteq R$ prove that

$$J \subseteq R \text{ is an ideal} \quad \iff \quad J/I \subseteq R/I \text{ is an ideal,}$$

in which case we have an isomorphism of rings:

$$\frac{R/I}{J/I} \cong \frac{R}{J}.$$

[Hint: Consider the ring homomorphism $\varphi : R/I \to R/J$ defined by $\varphi(a + I) = a + J$.]

Week 16

16.1 Ideal Theory of $\mathbb{F}$ and $\mathbb{Z}$

That was the general theory. This week we will start to focus on specific examples. But first, a bit of notational hygiene.

The Subring and the Ideal Generated by a Subset. Let R be a ring and let $S \subseteq R$ be any subset. Here's a question for you:

$$\text{What should the notation } \langle S \rangle \subseteq R \text{ represent?}$$

I can think of at least four possibilities:

- The smallest additive subgroup containing S.
- The smallest subring containing S.
- The smallest ideal containing S.
- The smallest subfield containing S (if any exist).

Because of this ambiguity I will try to avoid the notation $\langle S \rangle$ as much as possible.[115] Instead I will use the following notations, which are fairly standard.

First, let $E \supseteq R$ be any ring extension and let $S \subseteq E$ be any subset. Then we define

$$R[S] \coloneqq \bigcap \{\text{sub}\textbf{rings} \text{ of } E \text{ that contain the set } R \cup S\} \subseteq E.$$

In other words, $R[S]$ is the smallest subring of E that contains the set $R \cup S$. If E contains a subfield (for example, if E is a field) then we will also define

$$R(S) \coloneqq \bigcap \{\text{sub}\textbf{fields} \text{ of } E \text{ that contain the set } R \cup S\} \subseteq E.$$

Observe that this is consistent with our previous notation for field extensions. Since every subfield is itself a subring, observe that we always have

$$R \subseteq R[S] \subseteq R(S) \subseteq E.$$

[115] One major exception: When it comes to rings of polynomials I will tend to write $\langle S \rangle$ for the **ideal** generated by a subset S, since the alternative notations are too cumbersome.

In general the inclusion $R[S] \subseteq R(S)$ is strict. However, we will meet a nice class of examples below (when R is a field and S is a finite set that is "algebraic" over R) for which $R[S] = R(S)$.

Next let R be a ring and let $S \subseteq R$ be any subset. Then we define

$$RS \coloneqq \bigcap \{\text{ideals of } R \text{ that contain } S\} \subseteq R.$$

In other words, RS is the smallest ideal of R that contains the set S. You should be aware that many authors use the notations $\langle S \rangle$ or (S) for this ideal. But I prefer the "multiplicative" notation RS because of the following fact:

> The ideal generated by S equals the set of *finite R-linear combinations*:[116]
>
> $$RS = \{a_1\alpha_1 + a_2\alpha_2 + \cdots + a_k\alpha_k : a_1, \ldots, a_k \in R, \alpha_1, \ldots, \alpha_k \in S, k \geq 1\}.$$

Proof. Let I be the set of linear combinations above. Since RS is an ideal containing S we see that $I \subseteq RS$. On the other hand, for all $a_1\alpha_1 + \cdots a_k\alpha_k \in I$ and $b \in R$ we have

$$b(a_1\alpha_1 + \cdots a_k\alpha_k) = (ba_1)\alpha_1 + \cdots + (ba_k)\alpha_k \in I,$$

which implies that $I \subseteq R$ is an ideal. Furthermore, we have $S \subseteq I$ since $\alpha = 1\alpha \in I$ for all $\alpha \in S$. Finally, since RS is the **smallest** ideal containing S we conclude that $RS \subseteq I$. □

If R is a commutative ring (which for us it always will be), then we will also write $RS = SR$. However, when R is **non-commutative** then the notations RS and SR will denote the smallest **left ideal** and **right ideal** containing S, respectively. You can probably guess all the relevant definitions. ///

Ideals generated by a single element are very important so we give them a special name.

Definition of Principal Ideals. Let R be a commutative ring. The ideal generated by a single element is called a *principal ideal*. For $\alpha \in R$ we will use the notation

$$\alpha R = R\alpha \coloneqq R\{\alpha\} = \{a\alpha : a \in R\}.$$

The smallest and largest ideals are both principal, generated by 0 and 1, respectively:

$$\begin{array}{rcll} 0R & = & \{0\} & \text{is called the } \textit{zero ideal}. \\ 1R & = & R & \text{is called the } \textit{unit ideal}. \end{array}$$

///

By the way, the name "unit ideal" is motivated by the following fact:

[116] This fact suggests an analogy between "ideals" and "vector subspaces". The analogy can be made precise with the definition of *R-modules*. But again, we don't have a need for that level of abstraction in this course.

Let $I \subseteq R$ be an ideal. Then $I = R$ if and only if I contains a unit.

Proof. If $I = R$ then I contains the unit 1. Conversely, let $u \in I$ be a unit. By definition this means that $uu^{-1} = 1$ for some (unique) $u^{-1} \in R$. But then since I is an ideal we must have $1 = uu^{-1} \in I$, and it follows that

$$a = 1a \in I \text{ for all } a \in R.$$

□

Now for the examples.

Example: Fields. Let $\mathbb{F}$ be a field. Then I claim that $0\mathbb{F}$ and $1\mathbb{F}$ are the **only** ideals.

Proof. Let $I \neq 0\mathbb{F}$ be a non-zero ideal. Then I contains a non-zero element, which must be a unit because $\mathbb{F}$ is a field. It follows from the previous remark that $I = 1\mathbb{F}$. □

Conversely, if R is any ring that contains **exactly two ideals**, then I claim that R is a field.

Proof. Assume that $0R \neq 1R$ are the only two ideals of R and let $0 \neq a \in R$ be any nonzero element. Now consider the principal ideal aR. Since $aR \neq 0R$ we must have $aR = 1R$. In particular, since $1 \in aR$ there exists an element $b \in R$ such that $1 = ab$. □

We have shown that a field is a ring with **exactly two ideals**. But recall that a group G with **exactly two normal subgroups** is called a *simple group*. Since ideals are analogous to normal subgroups (being kernels of the relevant homomorphisms) we might say that

a field is a "simple ring".

///

Example: The Integers. Fields are in some sense "too simple". The prototypical example of a ring is the ring of integers:

$$(\mathbb{Z}, +, \times, 0, 1).$$

But the ring $\mathbb{Z}$ is still "rather simple" because of the following fact:

Every ideal of $\mathbb{Z}$ is principal.

Proof. Let $I \subseteq \mathbb{Z}$ be any ideal. If $I = 0\mathbb{Z}$ then it is principal. Otherwise, let $0 \neq n \in I$ be a nonzero element with minimal absolute value (which exists by the well-ordering). Since I is an ideal we have $n\mathbb{Z} \subseteq I$. I claim that in fact $I = n\mathbb{Z}$. To see this, consider any element $a \in I$ and divide by n to obtain

$$a = qn + r \text{ for some (unique) } q, r \in \mathbb{Z} \text{ with } 0 \leq r < |n|.^{117}$$

If $r \neq 0$ then the facts $r = a - qn \in I$ and $0 < |r| = r < |n|$ contradict the minimality of n. Thus we must have $r = 0$ and hence $a = qn \in n\mathbb{Z}$. Finally, since this is true for all $a \in I$ we conclude that $I \subseteq n\mathbb{Z}$. □

You may recall that we already proved this last semester under the guise of "cyclic groups". In fact, we now see that every cyclic group $\mathbb{Z}/n\mathbb{Z}$ has a natural ring structure defined by

$$(a + n\mathbb{Z})(b + n\mathbb{Z}) = (ab + n\mathbb{Z}).$$

I had to hold my tongue many times.[118] Here is another important property of integers:

For all $a, b \in \mathbb{Z}$, if $ab = 0$ then we must have $a = 0$ or $b = 0$.

This could be taken as an axiom, but it is usually proved using induction and the fact that $0 \neq 1$. Equivalently we could say that

the ring $\mathbb{Z}$ has no *zero-divisors*.

The technical term for a ring without zero-divisors is an *integral domain*.[119] You might at first assume that **every** ring is an integral domain, but consider the following fact:

If $n \in \mathbb{Z}$ is not zero or prime then the ring $\mathbb{Z}/n\mathbb{Z}$ is not an integral domain.

Proof. If n is not prime then by definition there exist $a, b \in \mathbb{Z}$ where $n = ab$ and neither of a, b is in the ideal $n\mathbb{Z}$. Then we have

$$(a + n\mathbb{Z})(b + n\mathbb{Z}) = (n + n\mathbb{Z}) = (0 + n\mathbb{Z}),$$

where neither of $a + n\mathbb{Z}$ or $b + n\mathbb{Z}$ is equal to the zero element $0 + n\mathbb{Z}$. □

On the other hand, you proved on a previous homework that the ring $\mathbb{Z}/p\mathbb{Z}$ for prime p is actually a **field** (hence also an integral domain). This fact is closely related to the concept of

[117] This is not usually regarded as an axiom, but it is close to being the defining property of the integers.

[118] And the tongue holding will continue, because abelian groups are the same as "$\mathbb{Z}$-modules" and rings are the same as "$\mathbb{Z}$-algebras". Modules and algebras are important types of algebraic structures but I consider them more suitable for a graduate course.

[119] This notation is a near-literal translation of Kronecker's term *Integralitätsbereich* [domain of integrality]. Compare this to the term *Rationalitätsbereich* [domain of rationality], which was his name for fields. Why are these terms so closely related? You will prove on a future homework that the concepts of "integral domain" and "subring of a field" are equivalent.

division with remainder.

///

The examples of fields and integers will motivate much of the theory going forward. Next time we will meet the third fundamental example of a ring: polynomials.

16.2 What is a Polynomial?

Today's lecture is a bit philosophical, but don't worry — we'll discuss examples soon.

What is a Polynomial? For any subset of a ring $S \subseteq R$ we saw that the the smallest ideal $S \subseteq RS \subseteq R$ containing S consists of all the finite R-linear combinations:

$$RS = \{a_1\alpha_1 + a_2\alpha_2 + \cdots + a_k\alpha_k : a_1, \ldots, a_k \in R, \alpha_1, \ldots, \alpha_k \in S, k \geq 0\}.$$

We can give a similar explicit characterization of the sub**ring** generated by a subset. To be specific, let $E \supseteq R$ be a ring extension and let $S \subseteq E$ be any subset. Then the smallest subring $(R \cup S) \subseteq R[S] \subseteq E$ containing $R \cup S$ equals the set of all "finite polynomial expressions":

$$R[S] = \left\{ \sum_{n_1,\ldots,n_k} a_{n_1,\ldots,n_k} \alpha_1^{n_1} \cdots \alpha_k^{n_k} : a_{n_1,\ldots,n_k} \in R, \alpha_i \in S, k, n_i \geq 0 \right\}.$$

The word "finite" means that only finitely many of the coefficients $a_{n_1,\ldots,n_k}$ are nonzero. I won't bother to prove this fact because the notation is atrocious. Instead we'll prove the special case when $S = \{\alpha\} \subseteq E$ has just one element. In this case I claim that we have

$$R[\alpha] = \{a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n : a_0, \ldots, a_n \in R, n \geq 0\}.$$

Proof. Let $P = \{a_0 + a_2\alpha^2 + \cdots + a_n\alpha^n\}$ be the set of "finite R-polynomial expressions in α". Since $R[\alpha] \subseteq E$ is a subring containing $R \cup \{\alpha\}$ we note that $P \subseteq R[\alpha]$. On the other hand, one can see that P contains $0, 1 \in E$ and is closed under addition and multiplication, hence $P \subseteq E$ is a subring. Then since P contains $R \cup \{\alpha\}$, and since $R[\alpha]$ is the **smallest** subring containing $R \cup \{\alpha\}$, we conclude that $R[\alpha] \subseteq P$. □

This fact motivates the following definition.

Definition of Polynomials in One Variable. Let R be a ring and let x be a formal symbol, called a "variable". We use this to define a sequence of formal symbols "x^n" for $n \geq 0$:

$$x^0 \coloneqq 1, \quad x^1 \coloneqq x, \quad x^2, \quad x^3, \quad x^4, \quad \ldots.$$

Then we define the set of *formal polynomials in* x:

$$R[x] \coloneqq \{a_0 + a_1 x + a_2 x^2 + \cdots a_n x^n : a_0, \ldots, a_n \in R, n \geq 0\}.$$

This set has a natural ring structure which we define by pretending that

$$a_n x^n = a_n \cdot \underbrace{x \cdot x \cdots x}_{n \text{ times}},$$

even though x is not a number so this multiplication is imaginary. In other words, for all polynomials $f(x) = \sum_i a_i x^i$ and $g(x) = \sum_i b_i x^i$ we define

$$\left(\sum_i a_i x^i\right) + \left(\sum_i b_i x^i\right) \coloneqq \sum_i (a_i + b_i) x^i$$

and

$$\left(\sum_i a_i x^i\right)\left(\sum_i b_i x^i\right) \coloneqq \sum_k \left(\sum_{i+j=k} a_i b_j\right) x^k.$$

To interpret these expressions one should assume that the indices run over all non-negative integers and that only finitely many coefficients are non-zero. ///

Remarks:

- Someone should prove that addition and multiplication of polynomials satisfy the ring axioms, but I won't do it here because the notation is too ugly. The proof that multiplication is associative comes down to the following identity:

$$\sum_{k+\ell=m} \left(\sum_{i+j=\ell} a_i b_j\right) c_k = \sum_{i+j+k=m} a_i b_j c_k = \sum_{i+\ell=m} a_i \left(\sum_{j+k=\ell} b_j c_k\right).$$

- What is the relationship between "formal polynomials" and "polynomial functions"? For any formal polynomial $f(x) = \sum_i a_i x^i \in R[x]$ we define a function $f : R \to R$ by evaluating $f(x)$ at $\alpha \in R$:

$$f(\alpha) \coloneqq \sum_i a_i \alpha^i \in R.$$

More generally, for any ring homomorphism $\varphi : R \to S$ we define a polynomial $f^\varphi(x) \in S[x]$ by applying φ to the coefficients and then we define a function $f^\varphi : S \to S$ by evaluating at $\alpha \in S$:

$$f^\varphi(\alpha) \coloneqq \sum_i \varphi(a_i) \alpha^i \in S.$$

For any fixed $\alpha \in S$ it turns out that the "evaluate at α" function $f(x) \mapsto f^{\varphi}(\alpha)$ is a **ring homomorphism** $R[x] \to S$, which plays an important role in Galois theory. (Don't worry, I'll remind you of the definitions later.) ///

The key to the structure of polynomial rings is the following "division algorithm", which is analogous to the division algorithm for integers. In order to state and prove this result we need a couple of definitions.

Definition: Degree of a Polynomial. Let R be a ring and consider any polynomial $f(x) \in R[x]$. The *degree* of $f(x)$ is defined as the highest power of x that occurs with a nonzero coefficient. In other words, we say that $\deg(f) = n$ when

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \quad \text{with } a_n \neq 0.$$

Note that a polynomial of degree 0 is the same thing as a nonzero constant:

$$\deg(f) = 0 \quad \Longleftrightarrow \quad f(x) = a_0 \text{ for some } 0 \neq a_0 \in R.$$

It is more difficult to define the degree of the *zero polynomial* $0 \in R[x]$. We could just say that $\deg(0)$ is undefined, but I prefer the following convention:

$$\deg(0) \coloneqq -\infty.$$

[After all, this **is** the highest power of x that occurs with a nonzero coefficient.] Note that the degree of a sum always satisfies the following property:

$$\deg(f+g) \leq \max\{\deg(f), \deg(g)\}.$$

What about the degree of a product? We would like to say that $\deg(fg) = \deg(f) + \deg(g)$, but this is not always true. For example, consider the polynomials $f(x) = 1 + 2x$ and $g(x) = 1 + 2x^2$ with coefficients in $\mathbb{Z}/4\mathbb{Z}$. Then we have

$$(1+2x)(1+2x^2) = 1 + 2x + 2x^2 + 4x^3 = 1 + 2x + 2x^2 + 0x^3 = 1 + 2x + 2x^2,$$

so that $2 = \deg(fg) \neq \deg(f) + \deg(g) = 1 + 2$. The problem here is that the leading coefficients of f and g are zero-divisors. If we assume that the leading coefficients of $f(x)$ and $g(x)$ are **not zero-divisors** then we will have

$$\deg(fg) = \deg(f) + \deg(g).$$

///

Now we are ready for the theorem.

The Division Theorem for Polynomials. Let R be a ring and consider polynomials $f(x), g(x) \in R[x]$. If the leading coefficient of $g(x)$ is a unit then there exist unique polynomials $q(x), r(x) \in R[x]$ such that

$$\begin{cases} f(x) = q(x)g(x) + r(x), \\ \deg(r) < \deg(g). \end{cases}$$

These $q(x)$ and $r(x)$ are called the *quotient* and the *remainder* of $f(x)$ mod $g(x)$. ///

Proof. First we will prove existence of $q(x)$ and $r(x)$, then we will prove uniqueness.

Existence. Consider the set of potential remainders:

$$S = \{f(x) - q(x)g(x) : q(x) \in R[x]\}.$$

Let $r(x) \in S$ be a potential remainder of **minimal degree**, which exists by well-ordering of the set of degrees $\{-\infty < 0 < 1 < 2 < \cdots\}$. By definition we have $f(x) = q(x)g(x) + r(x)$ for some $q(x) \in R[x]$ and it remains to prove that $\deg(r) < \deg(g)$. To do this, let us assume for contradiction that $\deg(r) \geq \deg(g)$. Then we must have

$$g(x) = a_0 + \cdots + a_m x^m \quad \text{and} \quad r(x) = b_0 + \cdots b_n x^n$$

where a_m is a unit, b_n is nonzero, and $0 \leq m \leq n$. We can use these facts to cook up a remainder with strictly lower degree. Specifically, we define the polynomial

$$h(x) \coloneqq r(x) - \frac{b_n}{a_m} x^{n-m} \cdot g(x) = \left(b_n - \frac{b_n}{a_m} a_m\right) x^n + \text{ lower terms.}$$

Note that $\deg(h) < \deg(r)$ by construction. But we also have

$$h(x) = f(x) - q(x)g(x) - \frac{b_n}{a_m} x^{n-m} \cdot g(x) = f(x) - \left(q(x) + \frac{b_n}{a_m} x^{n-m}\right) g(x) \in S,$$

which contradicts the minimality of $r(x)$. It follows that $\deg(r) < \deg(g)$ as desired.

Uniqueness. Suppose that we have polynomials $q_1, q_2, r_1, r_2 \in R[x]$ satisfying

$$\begin{cases} f(x) = q_1(x)g(x) + r_1(x), \\ \deg(r_1) < \deg(g), \end{cases} \qquad \begin{cases} f(x) = q_2(x)g(x) + r_2(x), \\ \deg(r_2) < \deg(g). \end{cases}$$

To prove that $r_1 = r_2$ and $q_1 = q_2$ we first equate expressions for f to obtain

$$\begin{aligned} q_1 g + r_1 &= q_2 g + r_2 \\ (q_1 - q_2)g &= (r_2 - r_1). \end{aligned}$$

If $r_2 - r_1 \neq 0$ then since the leading coefficient of g is a unit (in particular, not a zero-divisor) we also have $q_1 - q_2 \neq 0$, which implies that

$$\deg(r_2 - r_1) = \deg((q_1 - q_2)g) = \deg(q_1 - q_2) + \deg(g) \geq \deg(g).$$

But this contradicts the fact that $\deg(r_2 - r_1) \leq \max\{\deg(r_1), \deg(r_2)\} < \deg(g)$, so we must have $r_2 - r_1 = 0$. Finally, since $r_2 - r_1 = 0$ and since g has a unit leading coefficient we conclude that

$$(q_1 - q_2)g = 0 \qquad \Longrightarrow \qquad q_1 - q_2 = 0.$$

□

This theorem is most interesting when $R = \mathbb{F}$ is a field, because of the following fact:

Every nonzero polynomial in $\mathbb{F}[x]$ has a unit leading coefficient.

16.3 Descartes' Factor Theorem

The Division Theorem for Polynomials is really an algorithm. Here are two examples.

Example: Long Division Over $\mathbb{Z}$. Consider the following polynomials over $\mathbb{Z}$:

$$f(x) = 2x^3 + 3x + 1,$$
$$g(x) = x + 1.$$

Since the leading coefficient of $g(x)$ is the unit $1 \in \mathbb{Z}$ we we know that there exist (unique) polynomials $q(x), r(x)$ satisfying

$$\begin{cases} (2x^3 + 3x + 1) = q(x)(x + 1) + r(x), \\ \deg(r) < \deg(x + 1) = 1. \end{cases}$$

The proof of existence above leads to the following algorithm for computing $q(x)$ and $r(x)$:

$$\begin{array}{rr|llll} & & 2x^2 & -2x & +5 & \\ \hline x+1 & & 2x^3 & & +3x & +1 \\ & & -2x^3 & -2x & & \\ \hline & & & -2x^2 & +3x & +1 \\ & & & -2x^2 & -2x & \\ \hline & & & & 5x & +1 \\ & & & & -5x & -5 \\ \hline & & & & & -4 \end{array}$$

We conclude that

$$(2x^3 + 3x + 1) = q(x)(x + 1) + r(x) = (2x^2 - 2x + 5)(x + 1) - 4.$$

Note that the remainder $r(x) = -4$ satisfies $0 = \deg(-4) < \deg(x + 1) = 1$ as expected. ///

Example: Long Division Over $\mathbb{Q}$. Let's change the example slightly:

$$f(x) = 2x^3 + 3x + 1,$$
$$g(x) = 2x + 1.$$

This time the polynomial $g(x) = 2x + 1 \in \mathbb{Z}[x]$ has a non-unit leading coefficient. Thus there is no guarantee that the quotient and remainder exist in $\mathbb{Z}[x]$. In fact, we see that the division algorithm fails at the second step:

$$\begin{array}{rr|llll} & & 2x^2 & ? & & \\ \cline{2-6} 2x+1 & & 2x^3 & & +3x & +1 \\ & & -2x^3 & -x^2 & & \\ \cline{3-6} & & & -x^2 & +3x & +1 \\ & & & ? & & \end{array}$$

In order to cancel the "leading term" $-x^2$ we would need to multiply $2x+1$ by the "monomial" $-\frac{1}{2}x$, which does not exist in $\mathbb{Z}[x]$. However, since $\mathbb{Z} \subseteq \mathbb{Q}$ is a subring we could also think of $f(x)$ and $g(x)$ as elements of $\mathbb{Q}[x]$. Then since 2 **is a unit** in $\mathbb{Q}$ (with inverse $1/2$) the algorithm is guaranteed to succeed:

$$\begin{array}{rr|llll} & & 2x^2 & -\frac{1}{2}x & +\frac{7}{4} & \\ \cline{2-6} 2x+1 & & 2x^3 & & +3x & +1 \\ & & -2x^3 & -x^2 & & \\ \cline{3-6} & & & -x^2 & +3x & +1 \\ & & & x^2 & +\frac{1}{2}x & \\ \cline{4-6} & & & & \frac{7}{2}x & +1 \\ & & & & -\frac{7}{2}x & -\frac{7}{4} \\ \cline{5-6} & & & & & -\frac{3}{4} \end{array}$$

We conclude that the unique quotient and remainder in $\mathbb{Q}[x]$ are

$$q(x) = 2x^2 - \frac{1}{2}x + \frac{7}{4} \qquad \text{and} \qquad r(x) = -\frac{3}{4}.$$

Actually we don't have to extend all the way to $\mathbb{Q}$. The quotient and remainder already exist in the smaller ring $\mathbb{Z}[1/2][x] \subseteq \mathbb{Q}[x]$, where $\mathbb{Z}[1/2] \subseteq \mathbb{Q}$ is the subring of fractions whose denominators are powers of 2. This is the smallest subring of $\mathbb{Q}$ in which 2 is a unit. ///

Next I will show you two important corollaries of the division algorithm. The first result goes back to René Descartes in his work *La Géométrie* (1637). This is the same work in which he introduced the concepts of "analytic geometry" and "Cartesian coordinates". Here is Descartes' statement of the result:

> *It is evident from the above that [a polynomial equation] having several roots is always divisible by a binomial consisting of the unknown quantity diminished by the value of one of the true roots, or plus the value of one of the true roots. In this way, the degree of an equation can be lowered. On the other hand, if [a polynomial] is not divisible by a binomial consisting of the unknown quantity plus or minus some other quantity, then this latter quantity is not a root of the equation.*[120]

And here is the modern statement.

Corollary (Descartes' Factor Theorem). Let $E \supseteq R$ be an extension of (commutative) rings. Then for any polynomial $f(x) \in R[x]$ and for any element $\alpha \in E$ there exists a (unique) polynomial $g(x) \in E[x]$ of degree $\deg(f) - 1$ such that

$$f(x) = (x - \alpha)g(x) + f(\alpha).$$

It follows from this that

$$\alpha \in E \text{ is a root of } f(x) \quad \Longleftrightarrow \quad (x - \alpha) \text{ divides } f(x) \text{ in the ring } E[x].$$

///

Proof. Since $x - \alpha = 1x - \alpha \in E[x]$ has a unit leading coefficient we know that there exist (unique) polynomials $q(x), r(x) \in E[x]$ satisfying

$$\begin{cases} f(x) = (x - \alpha)q(x) + r(x), \\ \deg(r) < \deg(x - \alpha). \end{cases}$$

Since $\deg(x - \alpha) = 1$ this implies that $\deg(r) = 0$ or $\deg(r) = -\infty$. In other words, $r(x) = c \in E$ is a constant. To compute this constant we simply plug in $x = \alpha$ to obtain

$$f(\alpha) = (\alpha - \alpha)q(\alpha) + c = 0 \cdot q(\alpha) + c = 0 + c = c.$$

[120] Quoted from page 159–160 of *The Geometry of René Descartes* (1954).

[**Remark:** For this step we needed the fact that α commutes with all of the coefficients of $q(x)$. This is why we assumed that E is a **commutative** ring.] Then to compute the degree of $q(x)$ we use the fact that 1 is not a zero-divisor to obtain

$$\deg(f) = \deg((x-\alpha)q+c) = \deg((x-\alpha)q) = \deg(x-\alpha)+\deg(q) = 1+\deg(q).$$

Finally, if $f(x) = (x-\alpha)g(x)$ for some polynomial $g(x) \in E[x]$ then we have $f(\alpha) = (\alpha-\alpha)g(\alpha) = 0 \cdot g(\alpha) = 0$. [Again, we assume that E is commutative.] Conversely, if $f(\alpha) = 0$ then the above result implies that

$$f(x) = (x-\alpha)q(x) + f(\alpha) = (x-\alpha)q(x) + 0 = (x-\alpha)q(x)$$

for some $q(x) \in E[x]$. □

For example, recall from above that -4 is the remainder of $2x^3+3x+1$ mod $x+1$. On the other hand, by plugging $x=-1$ into $2x^3+3x+1$ we obtain

$$2(-1)^3 + 3(-1) + 1 = -4.$$

Remarks:

- You will give a more constructive proof of this result on the homework.
- You will also use induction to prove the following corollary: A polynomial of degree n has **at most** n **distinct roots** in any integral domain. This finally justifies some of our remarks from Weeks 13 and 14.
- A polynomial of degree n may have **more than** n **roots** in a non-commutative ring. For example, you may be familiar with the ring of *quaternions*:

  $$\mathbb{H} = \{a+bi+cj+dk : a,b,c,d \in \mathbb{R}, i^2 = j^2 = k^2 = ijk = -1\}.$$

  This ring is non-commutative because, for example, $ij = -ji$. Note that the polynomial $x^2+1 \in \mathbb{H}[x]$ has at least three roots: i, j, k. [In fact, one can show that x^2+1 has **uncountably many** roots in $\mathbb{H}$.]
- On the homework you will also show that a polynomial of degree n may have more than n roots in a non-integral domain. ///

To end this section I will present another important corollary of the division algorithm.

Corollary (Every Ideal of $\mathbb{F}[x]$ is Principal). Let $\mathbb{F}$ be a field and consider the ring of polynomials $\mathbb{F}[x]$. Then every ideal $I \subseteq \mathbb{F}[x]$ is principal.

Proof. Let $I \subseteq \mathbb{F}[x]$ be an ideal. If $I = 0\mathbb{F}[x]$ then it is principal. Otherwise, let $0 \neq m(x) \in I$ be a nonzero element of minimal degree (which exists by well-ordering of degrees). Since I is an ideal we have $m(x)\mathbb{F}[x] \subseteq I$. I claim that in fact $I = m(x)\mathbb{F}[x]$. To see this, consider any element $f(x) \in I$. Since $\mathbb{F}$ is a field and $m(x) \neq 0$ we know that the leading coefficient of $m(x)$ is a unit, hence there exist $q(x), r(x) \in \mathbb{F}[x]$ satisfying

$$\begin{cases} f(x) = q(x)m(x) + r(x), \\ \deg(r) < \deg(m). \end{cases}$$

If $r(x) \neq 0$ then the fact that $r(x) = f(x) - q(x)m(x) \in I$ contradicts the minimality of $m(x)$. Hence we must have $r(x)$ and it follows that $f(x) = m(x)q(x) \in m(x)\mathbb{F}[x]$. Finally, since this is true for all $f(x) \in I$ we conclude that $I \subseteq m(x)\mathbb{F}[x]$ as desired. □

Remarks:

- Note that this proof is almost identical to our proof that every ideal of $\mathbb{Z}$ is principal. They both depended on the existence of a division algorithm.
- In addition to having only principal ideals, both of the rings $\mathbb{Z}$ and $\mathbb{F}[x]$ are integral domains. For this reason we will call them PIDs (Principal Ideal Domains).
- The definition of PIDs is not completely obvious, but it turns out that this is a very natural class of rings with many nice properties. In particular, every PID satisfies "unique prime factorization". We will discuss this next week.

Exercises

16.A Invariance of Quotient and Remainder

CHANGE S TO E. Let $R \subseteq S$ be a subring, so that $R[x] \subseteq S[x]$ is also a subring. Consider any two polynomials $f(x), g(x) \in R[x]$ where $g(x)$ has a unit leading coefficient $u \in R^\times$.

(a) Suppose that there exist polynomials $q(x), r(x) \in S[x]$ such that $\deg(r) < \deg(g)$ and $f(x) = q(x)g(x) + r(x)$. In this case prove that we actually have $q(x), r(x) \in R[x]$.

(b) Prove that $g(x)|f(x)$ in $R[x]$ if and only if $g(x)|f(x)$ in $S[x]$.

16.B Descartes' Factor Theorem Again

Let $E \supseteq R$ be any ring extension and let $f(x) \in R[x]$ be any polynomial with coefficients in R.

(a) For any element $\alpha \in E$ prove that $f(\alpha) = 0$ if and only if there exists a polynomial $h(x) \in E[x]$ with coefficients in E such that $f(x) = (x - \alpha)h(x)$ and $\deg(h) = \deg(f) - 1$. [Hint: For all integers $n \geq 2$ observe that

$$x^n - \alpha^n = (x - \alpha)(x^{n-1} + x^{n-2}\alpha + \cdots + x\alpha^{n-2} + \alpha^{n-1}) \in E[x].$$

Now consider the polynomial $f(x) - f(\alpha) \in E[x]$.]

(b) *Counting Roots.* If E is an *integral domain*, use the result of part (a) to prove that any polynomial $f(x) \in R[x]$ has at most $\deg(f)$ distinct roots in E.

(c) *A Non-Example.* Let $E = R = \mathbb{Z}/8\mathbb{Z}$ and consider the polynomial $x^2 - 1$. How many roots does this polynomial have? Why does this not contradict part (b)?

16.C Prime and Maximal Ideals

Let R be a ring and let $I \subseteq R$ be an ideal.

(a) We say that I is a *maximal ideal* if

$$\text{for any ideal } J \subseteq R \text{ we have } (I \subsetneq J) \Rightarrow (J = R).$$

Prove that R/I is a field if and only if I is maximal.

(b) We say that I is a *prime ideal* if

$$\text{for any } a, b \in R \text{ we have } (ab \in I) \Rightarrow (a \in I \text{ or } b \in I).$$

Prove that R/I is an integral domain if and only if I is prime.

(c) Prove that every maximal ideal is prime.

(d) Let $\mathbb{Z}[x]$ be the ring of polynomials over $\mathbb{Z}$ and consider the principal ideal

$$x\mathbb{Z}[x] = \{xf(x) : f(x) \in \mathbb{Z}[x]\}.$$

Prove that $x\mathbb{Z}[x]$ is prime but not maximal. [Hint: $\mathbb{Z}[x]/x\mathbb{Z}[x] \cong \mathbb{Z}$.]

Week 17

17.1 Definition of PIDs

Abstract ring theory is a big subject with too many definitions, but I believe that most of the theory is motivated by an analogy between the following two kinds of rings:

$$\textit{integers } \mathbb{Z} \quad \approx \quad \textit{polynomials in one variable over a field } \mathbb{F}[x]$$

This week we will explore the basics of this analogy up to the theory of unique prime factorization. I will try not to get too distracted by pathological counterexamples.

The first similarity between $\mathbb{Z}$ and $\mathbb{F}[x]$ is the property of being "integral domains". We have already seen some of the following concepts but I want to collect them in one place for posterity.

Theorem (Definition of Integral Domains). Let R be a ring. Then the following three conditions are equivalent:

(D1) The zero ideal $0R \subseteq R$ is prime:

$$(a \notin 0R \text{ and } b \notin 0R) \Rightarrow (ab \notin 0R).$$

(D2) The ring R has no zero-divisors:

$$(a \neq 0 \text{ and } b \neq 0) \Rightarrow (ab \neq 0).$$

(D3) The ring R satisfies multiplicative cancellation:

$$(a \neq 0 \text{ and } ab = ac) \Rightarrow (b = c).$$

Any ring satisfying one (and hence all) of these conditions is called an *integral domain.* ///

Proof. Note that (D1)⇔(D2) because $a \in 0R \Leftrightarrow a = 0$.

(D2)⇒(D3): Suppose that R has no zero-divisors and consider $a, b, c \in R$ with $a \neq 0$ and $ab = ac$. Then we have

$$\begin{aligned} ab &= ac \\ ab - ac &= 0 \\ a(b-c) &= 0 \\ (b-c) &= 0 \qquad \text{because } a \neq 0 \\ b &= c. \end{aligned}$$

(D3)⇒ (D2): Let R satisfy multiplicative cancellation and assume for contradiction that there exists a pair of zero-divisors: $a \neq 0$, $b \neq 0$ and $ab = 0$. Then we have $ab = a0$ which since $a \neq 0$ implies that $b = 0$. Contradiction. □

Remarks:

- Recall from the homework that a general ideal $I \subseteq R$ is called *prime* when its complement is closed under multiplication:

  $$(a \notin I \text{ and } b \notin I) \Rightarrow (ab \notin I).$$

  If $I \subseteq J \subseteq R$ are any ideals, then one can show that "primeness" is preserved by the Correspondence Theorem:

  $$(J/I \subseteq R/I \text{ is prime}) \Leftrightarrow (J \subseteq R \text{ is prime}).$$

  It follows that

  $$(R/I \text{ is a domain}) \Leftrightarrow (I/I \subseteq R/I \text{ is prime}) \Leftrightarrow (I \subseteq R \text{ is prime}).$$

- As for the name "prime", suppose that $p, a, b \in \mathbb{Z}$ are integers with p prime. Then Euclid's Lemma says that

  $$(p|ab) \Rightarrow (p|a \text{ or } p|b).$$

  Note that this is the same as

  $$(ab \in p\mathbb{Z}) \Rightarrow (a \in p\mathbb{Z} \text{ or } b \in p\mathbb{Z}).$$

  In other words, $p\mathbb{Z} \subseteq \mathbb{Z}$ is a prime ideal. We will generalize this idea below.

- There are only two basic ways[121] that a ring can **fail** to be an integral domain, both of which are illustrated by the rings $\mathbb{Z}/n\mathbb{Z}$:

[121] I'm lying a bit here. The real theorem says that a ring with zero-divisors has a nilpotent element or more than one minimal prime ideal. Having more than one minimal prime ideal is closely related to the existence of idempotents, and both of these are related to the idea of being "disconnected" in algebraic geometry.

(1) We say that $a \in R$ is *nilpotent* if $a \neq 0$ and $a^m = 0$ for some minimal $m \geq 2$. Then $a \cdot a^{m-1} = 0$ shows that R is not an integral domain. For example, the element $2 \in \mathbb{Z}/2^k\mathbb{Z}$ is nilpotent.

(2) We say that $e \in R$ is *idempotent* if $e \notin \{0, 1\}$ and $e^2 = e$. Then $e(1-e) = 0$ shows that R is not an integral domain. For example, note that $e = 3 \in \mathbb{Z}/6\mathbb{Z}$ is idempotent with $1 - e = 4$ and $e(1-e) = 3 \cdot 4 = 0$. Ultimately this comes from the Chinese Remainder isomorphism

$$\begin{array}{ccc} \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} & \to & \mathbb{Z}/6\mathbb{Z} \\ (a, b) & \mapsto & 3a + 4b \end{array}$$

since the elements $(1, 0), (0, 1)$ of the direct product $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ are idempotents.

And that's probably enough about that. ///

As the name suggests, the ring $\mathbb{Z}$ is an integral domain. Some authors adopt (D2) or (D3) as an axiom for the integers, but usually these are proved from the following axioms of order:

- For all $a, b \in \mathbb{Z}$ exactly one of following holds: $a < b$, $a = b$ or $a > b$.
- For all $a, b, c \in \mathbb{Z}$ with $a < b$ we have $(c > 0) \Rightarrow (ac < bc)$ and $(c < 0) \Rightarrow (ac > bc)$.

These axioms, in turn, can be derived from Peano's Axioms. Basically, every property of the ring $\mathbb{Z}$ is a property of induction.

The fact that $\mathbb{F}[x]$ is an integral domain is implied by the following more general fact:

$$(R \text{ is an integral domain}) \Rightarrow (R[x] \text{ is an integral domain}).$$

Proof. Let R be an integral domain let $f(x), g(x) \in R[x]$ be nonzero polynomials. By definition this means that

$$f(x) = a_0 + a_1 x + \cdots + a_m x^m \quad \text{and} \quad g(x) = b_0 + b_1 x + \cdots + b_n x^n,$$

where the leading coefficients $a_m, b_n \in R$ are nonzero. But then we have

$$f(x)g(x) = a_m b_n x^{n+m} + \text{ lower terms.}$$

Since R is a domain it follows that $a_m b_n \neq 0$ and hence $f(x)g(x) \neq 0$. □

The second similarity between $\mathbb{Z}$ and $\mathbb{F}[x]$ is the fact that each has a "division algorithm". The following definition is a bit awkward and we only really care about it as a stepping-stone to the next theorem.

Definition of Euclidean Ring. We say that a ring R is *Euclidean* if there exists a well-ordered[122] set $(\Omega, \leq)$ and a function $\nu : R \to \Omega$ satisfying the following property:

For all $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that

$$\begin{cases} a = qb + r, \\ \nu(r) < \nu(b). \end{cases}$$

We do not assume that the elements q, r are unique.

///

Note that $\mathbb{Z}$ is Euclidean with $\nu : \mathbb{Z} \to \{0 < 1 < 2 < \cdots\}$ given by the absolute value and that $\mathbb{F}[x]$ is Euclidean with $\nu : \mathbb{F}[x] \to \{-\infty < 0 < 1 < \cdots\}$ given by the degree. Here is the only reason we care about Euclidean rings.

Theorem (Euclidean Implies Principal Ideals). Let R be a ring. Then

$$(R \text{ is Euclidean}) \Rightarrow (\text{Every ideal } I \subseteq R \text{ is principal}).$$

///

Proof. We already proved this twice but let's do it one last time. If $I = 0R$ then we're done. Otherwise, choose $0 \neq m \in I$ with minimal $\nu(m)$. Then $mR \subseteq I$ and I claim that $mR = I$. Indeed, for any $a \in I$ we have

$$\begin{cases} a = qm + r, \\ \nu(r) < \nu(m). \end{cases}$$

If $r \neq 0$ then since $r \in I$ we get a contradiction to minimality. It follows that $r = 0$ and hence $a = qm \in mR$. Since this is true for all $a \in I$ we get $I \subseteq mR$. □

We summarize all of these properties with the following definition.

Definition of PIDs. Let R be a (commutative) ring. We say that R is a *principal ideal domain* (PID) if the following two properties hold:

- R is an integral domain,
- every ideal of R is principal.

///

In a subject with too many bad definitions, I believe that the definition of PIDs is good.[123]

[122] This means that every non-empty subset of Ω has a smallest element.

[123] What do I mean by this? There are many theorems of the form PID $\Rightarrow$ X for which we surprisingly also have X $\Rightarrow$ PID or (X + something small) $\Rightarrow$ PID. (The most basic example says that if $R[x]$ is a PID then R is a field.) This is rare in commutative algebra. When it happens you know you have a good definition.

17.2 Ideal Theory of $\mathbb{F}[x]$

Commutative algebra began with the study of number theory. In particular, many concepts of the subject were motivated by attempts to prove the following theorem:

> for all integers $x, y, z, n \in \mathbb{Z}$ with $n \geq 3$ and $xyz \neq 0$ we have
>
> $$x^n + y^n \neq z^n.$$

In 1637 (the same year as Descartes' *Géométrie*), Pierre de Fermat scribbled this result in the margin of his copy of Diophantus' *Arithmetica*, together with the following remark:

> *I have a truly marvelous demonstration of this proposition, which this margin is too narrow to contain.*

Through his correspondence Fermat tried to interest his contemporaries in number theory and he often challenged them by stating results without proof. However, this was the during the heart of the scientific revolution and it is likely that his contemporaries were more interested in applied areas of mathematics. For example, Christian Huygens made the following remark about Fermat's challenges in a 1658 letter to John Wallis:

> *There is no lack of better things for us to do.*[124]

It was approximately 100 years later when Leonhard Euler became interested in Fermat's number-theoretic ideas. Euler provided proofs for some of Fermat's unproved theorems (e.g., Fermat's Little Theorem) and he disproved others (e.g., Fermat's assertion that every number of the form $2^{2^n} + 1$ is prime).[125] But the result stated above resisted Euler's attempts and hence became known as "Fermat's Last Theorem" (FLT). In 1847 Gabriel Lamé gave a false proof of FLT in which he assumed that the ring $\mathbb{Z}[e^{2\pi i/n}]$ always has unique prime factorization. However, in 1844 Ernst Kummer had discovered the surprising fact that

> *the ring* $\mathbb{Z}[e^{2\pi i/23}]$ *does* **not** *have unique prime factorization.*

This motivated Kummer to develop a theory of "ideal prime factorization", which sadly did not repair Lamé's proof of FLT[126] but it did lead to the abstract theory of ideals.

In modern language, the motivation for Kummer's theory is to replace each **element** a in a ring R by the **principal ideal** $aR \subseteq R$ that it generates. The

[124] See Weil, *Number Theory: An approach through history from Hammurapi to Legendre*, (1984, page 119).

[125] On the homework you will investigate another result of Fermat on integers that can be expressed as a sum of two squares.

[126] It was eventually proved by Andrew Wiles in 1994.

first observation is that

$$aR \supseteq bR \quad \iff \quad a|b.$$

Proof. Suppose that $aR \supseteq bR$. Then since $b \in aR$ we have $ac = b$ for some $c \in R$. Conversely, suppose that $ac = b$ for some $c \in R$. Then for all $d \in R$ we have $bd = (ac)d = a(cd) \in aR$, and hence $bR \subseteq aR$. □

Many books use the following mnemonic:

to contain is to divide.

For the next observation we assume that R is an **integral domain**. Then we have

$$aR = bR \quad \iff \quad au = b \text{ for some unit } u \in R^\times.$$

Proof. One direction is done. For the other direction, suppose that $aR = bR$. If $a = 0$ or $b = 0$ then we have $a \cdot 1 = 0 = b$ with $1 \in R^\times$ as desired. So assume that $a \neq 0$. From the previous result we have $a|b$ and $b|a$, which implies that $ak = b$ and $b\ell = a$ for some $k, \ell \in R$. Finally, since R is an integral domain we have

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

and it follows that $k, \ell \in R^\times$. □

In general, we define a relation on the elements of a ring called "association".

Definition of Association. Let R be a ring. For all elements $a, b \in R$ we define

$$a \sim b \quad \iff \quad au = b \text{ for some unit } u \in R^\times.$$

[**Exercise:** Check that this is an equivalence relation.] When $a \sim b$ holds we say that a and b are *associates*. We will write $aR^\times := \{au : u \in R^\times\}$ for the equivalence class of $a \in R$, so that

$$a \sim b \quad \iff \quad aR^\times = bR^\times,$$

and we will use the notation $R/R^\times := \{aR^\times : a \in R\}$ for the set of equivalence classes. Warning: This is **not** a set of cosets because $R^\times \subseteq R$ is **not** a multiplicative subgroup. ///

Remarks:

- We always have $0R^\times = \{0\}$ and $1R^\times = R^\times$. If $a \in R$ is **non-zero-divisor** then the function $R^\times \to aR^\times$ defined by $u \mapsto au$ is a bijection $R^\times \leftrightarrow aR^\times$.
- If R is an **integral domain** then the previous result says that

$$aR = bR \quad \Longleftrightarrow \quad aR^\times = bR^\times.$$

  Hence we obtain a bijection between principal ideals and classes of associates.
- The most basic example is the ring of integers $\mathbb{Z}$ with group of units $\mathbb{Z}^\times$. In this case the classes of associates are

$$\mathbb{Z}/\mathbb{Z}^\times = \{\{0\}, \{1, -1\}, \{2, -2\}, \{3, -3\}, \ldots\}.$$

  By choosing the non-negative integer from each class we obtain the well-known bijection

$$\begin{array}{ccccccc} \mathbb{N} & \leftrightarrow & \mathbb{Z}/\mathbb{Z}^\times & \leftrightarrow & \{\text{principal ideals of } \mathbb{Z}\} & = & \{\text{ideals of } \mathbb{Z}\} \\ n & & & & & & n\mathbb{Z}. \end{array}$$

- If R is not an integral domain then strange things can happen. For example, consider the ring $\mathbb{Z}/12\mathbb{Z}$ with group of units $(\mathbb{Z}/12\mathbb{Z})^\times = \{1, 5, 7, 11\}$. One can check that

$$(\mathbb{Z}/12\mathbb{Z})/(\mathbb{Z}/12\mathbb{Z})^\times = \{\{0\}, \{1, 5, 7, 11\}, \{2, 10\}, \{3, 9\}, \{4, 8\}, \{6\}\}.$$

  Note that each class contains a unique divisor of 12. [Challenge Problem: Prove that the same holds in general for $\mathbb{Z}/n\mathbb{Z}$.] ///

Next we want to classify the principal ideals (hence all ideals) of the ring $\mathbb{F}[x]$. More generally, let R be any integral domain. Then I claim that

$$R[x]^\times = R^\times.$$

Proof. We think of $R \subseteq R[x]$ as the subring of constant polynomials. Note that this implies $R^\times \subseteq R[x]^\times$. Conversely, consider any elements $f(x), g(x) \in R[x]$ with $f(x)g(x) = 1$. Since R is a domain this implies that $\deg(f) + \deg(g) = \deg(1) = 0$. Then since degrees are non-negative we conclude that $\deg(f) = \deg(g) = 0$ and hence $f(x), g(x) \in R^\times$. □

[**Remark:** If the ring R has a nilpotent element then the inclusion $R^\times \subsetneq R[x]^\times$ is strict. For example, suppose we have $a \neq 0$ and $a^n = 0$ for some $n \geq 2$. Then the polynomial $1 - ax \in R[x]$ is a unit because

$$1 = 1 - 0x = 1 - a^n x^n = (1 - ax)(1 + ax + a^2x^2 + \cdots + a^{n-1}x^{n-1}).$$

This is yet another reason to prefer integral domains.]

This leads to the following theorem/definition.

Theorem (Definition of Monic Polynomials). We say that a polynomial is *monic* when its leading coefficient equals 1. If $\mathbb{F}$ is a field then we have a bijection

$$\{\text{ideals of } \mathbb{F}[x]\} \longleftrightarrow \{0\} \cup \{\text{monic polynomials in } \mathbb{F}[x]\}.$$

Proof. We know that $f(x)\mathbb{F}[x]$ is a PID, hence every ideal has the form $f(x)\mathbb{F}[x]$ for some polynomial $f(x) \in \mathbb{F}[x]$. If $f(x) \neq 0$ then we have $f(x) = a_0 + a_1x + \cdots a_nx^n$ for some $a_n \neq 0$. Since $\mathbb{F}$ is a field we can divide by a_n to obtain a **monic** polynomial $g(x) \coloneqq f(x)/a_n$. Then since $f(x)|g(x)$ and $g(x)|f(x)$ we have $f(x)\mathbb{F}[x] = g(x)\mathbb{F}[x]$. Conversely, suppose that $g(x)\mathbb{F}[x] = h(x)\mathbb{F}[x]$ where $g(x)$ and $h(x)$ are both monic. Since $\mathbb{F}[x]$ is a domain this implies that $g(x)u(x) = h(x)$ for some unit $u(x) \in \mathbb{F}[x]^\times$, which from the previous result equals a **nonzero constant** $u(x) = u \in \mathbb{F}^\times$. But then

$$u = (\text{leading coefficient of } gu) = (\text{leading coefficient of } h) = 1,$$

and we conclude that $g(x) = h(x)$. □

In summary, for any field $\mathbb{F}$ we have an isomorphism of lattices:

$$\left\{ \begin{array}{c} \text{ideals of } \mathbb{F}[x] \text{ under} \\ \text{reverse containment} \end{array} \right\} \cong \left\{ \begin{array}{c} \text{monic polynomials} \cup 0 \\ \text{under divisibility} \end{array} \right\}.$$

17.3 Every PID is a UFD

To end this week I will prove that each of the rings $\mathbb{Z}$ and $\mathbb{F}[x]$ has "unique prime factorization". The hardest part of the proof is to find the correct definition of "prime".

Let's begin with some examples. In the ring $\mathbb{Z}$ we can factor 12 in many ways:

$$\begin{aligned} 12 &= 2 \cdot 2 \cdot 3 \\ &= 2 \cdot 3 \cdot 2 \\ &= (-2) \cdot (-3) \cdot 2 \\ &= 3 \cdot 2 \cdot (-1)(-1) \cdot 2 \cdot 1 \cdot 1 \cdot 1 \\ &\text{etc.} \end{aligned}$$

We say that 12 has prime factors ± 2 and ± 3 with multiplicities 2 and 1, respectively. We don't want to say that ± 1 are prime because this will ruin uniqueness.

In the ring $\mathbb{Q}[x]$ the polynomial $x^2 + 1$ has many trivial factorizations:

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

$$= 3\left(\frac{1}{3}x^2 + \frac{1}{3}\right)$$

etc.

But I claim that $x^2 + 1$ cannot be factored in a non-trivial way.

Proof. Recall that a polynomial $f(x) \in \mathbb{Q}[x]$ is non-constant if and only if $\deg(f) \geq 1$. Now suppose for contradiction that we have $x^2 + 1 = f(x)g(x)$ for some non-constant polynomials $f(x)g(x) \in \mathbb{Q}[x]$. Since $\deg(f) + \deg(g) = \deg(fg) = 2$ this implies that $\deg(f) = \deg(g) = 1$. In particular we must have $f(x) = \alpha x + \beta$ for some $\alpha, \beta \in \mathbb{Q}$ with $\alpha \neq 0$. But then $f(-\beta/\alpha) = 0$, which implies that

$$(-\beta/\alpha)^2 + 1 = f(-\beta/\alpha)g(-\beta/\alpha) = 0g(-\beta/\alpha) = 0.$$

But I claim that this is impossible. Indeed, since $-\beta/\alpha \in \mathbb{Q}$ we must have $-\beta/\alpha = c/d$ for some $c, d \in \mathbb{Z}$ with $d \neq 0$. But then we have

$$\begin{aligned}(-c/d)^2 + 1 &= 0\\ c^2/d^2 &= -1\\ c^2 &= -d^2 < 0,\end{aligned}$$

which contradicts the fact that $c^2 \geq 0$. □

We say that the polynomial $x^2 + 1$ is *irreducible over* $\mathbb{Q}$. However, the same polynomial is *reducible* over the field extension $\mathbb{C} \supseteq \mathbb{Q}$ because

$$x^2 + 1 = (x - i)(x + i).$$

Hopefully these examples will motivate the following definition. The definition is a bit awkward because it only applies to integral domains, and because it rather arbitrarily excludes the zero element and the units. We exclude units because they lead to silly non-unique factorizations. As for the zero element: I say that 0 is irreducible in a domain, but that's just my opinion.

Definition of Irreducible Elements. Let R be an integral domain. We say that an element $a \in R$ is *irreducible* if the principal ideal $aR \subseteq R$ is "nontrivial and maximal among principal ideals". In other words, when the following two conditions hold:

- $0R \subsetneq aR \subsetneq 1R$,
- for all $b \in R$ we have $(aR \subseteq bR \subseteq 1R) \Rightarrow (aR = bR \text{ or } bR = 1R)$.

Equivalently, these two conditions say that

- a is not zero and not a unit,

- if $a = bc$ for some $b, c \in R$ then we have $c \in R^\times$ $(aR = bR)$ or $b \in R^\times$ $(bR = 1R)$.

///

Our first goal is to show that every element in a domain can be factored as a (finite) product of irreducible elements, times a unit. Sadly, there exist pathological examples where this is false. For example, consider the ring of polynomials over $\mathbb{Q}$ with constant term in $\mathbb{Z}$:

$$\mathbb{Z} + x\mathbb{Q}[x] = \{f(x) \in \mathbb{Q}[x] : f(0) \in \mathbb{Z}\} \subseteq \mathbb{Q}[x].$$

The units of this ring are just $\mathbb{Z}^\times = \{\pm 1\}$. But the polynomial x can never be factored into irreducibles because

$$x = 2 \cdot \frac{x}{2} = 2 \cdot 2 \cdot \frac{x}{4} = 2 \cdot 2 \cdot 2 \cdot \frac{x}{8} = \cdots .$$

Ultimately, the problem is that we have an infinite increasing chain of principal ideals:

$$\langle x \rangle \subsetneq \langle x/2 \rangle \subsetneq \langle x/4 \rangle \subsetneq \langle x/8 \rangle \subsetneq \cdots .$$

We will show that this problem does not occur in a PID.

Theorem (PID $\Rightarrow$ Factorization Terminates). Every element in a PID can be expressed as a (finite)[127] product of irreducible elements, times a unit.

Proof. Let $a_0 \in R$ be a non-zero non-unit[128] and assume for contradiction that the factoring process does not terminate. Then we obtain an infinite increasing chain of principal ideals:

$$a_0R \subsetneq a_1R \subsetneq a_2R \subsetneq a_3R \subsetneq \cdots .$$

I claim that the infinite union $I = \cup_i a_i R \subseteq R$ is an ideal. To see this, consider any $b, c \in I$ and $r \in R$. By definition there exist indices i, j such that $b \in a_i R$ and $c \in a_j R$. If $k = \max\{i, j\}$ then we have $b, c \in a_k R$ and it follows that

$$b - rc \in a_k R \subseteq I.$$

Since R is a PID we must have $I = aR$ for some $a \in I$. But then by definition we have $a \in a_k R$ for some k and it follows that

$$I = aR \subseteq a_k R \subsetneq a_{k+1} R \subseteq I.$$

Contradiction. □

[127] Products in a general ring are necessarily finite. To definite an infinite product one would need some notion of "convergence", which we do not have.

[128] I say that 0 is irreducible (times 1) and that every unit is a product of itself times no irreducibles.

[**Jargon:** We say that a ring R is *Noetherian* if it does not contain an infinite strictly increasing chain of ideals. We can rephrase the above theorem by saying that every PID is Noetherian. Emmy Noether showed that this condition is a very convenient abstract substitute for the well-ordering principle.]

And what about uniqueness? Consider the ring $\mathbb{Z}[\sqrt{-3}] = \{a + b\sqrt{-3} : a, b \in \mathbb{Z}\}$ and note that the element 4 has two seemingly different factorizations:

$$4 = 2 \cdot 2 = (1 + \sqrt{-3})(1 - \sqrt{-3}).$$

Indeed, you will show on the homework that the elements 2, $(1 + \sqrt{-3})$ and $(1 - \sqrt{-3})$ are irreducible but that 2 is not associate to either of $(1 + \sqrt{-3})$ or $(1 - \sqrt{-3})$. The problem is that none of these irreducible elements is "prime" in the following sense.

Definition of Prime Elements. Let $p \in R$ be a non-unit element of a ring. We say that

$$p \text{ is prime} \quad \iff \quad pR \subseteq R \text{ is a prime ideal.}$$

In other words, we say that $p \notin R^\times$ is prime if for all $a, b \in R$ we have

$$(p|ab) \quad \implies \quad (p|a \text{ or } p|b).$$

///

If R is a **domain** then for all elements $p \in R$ we have

$$p \text{ is prime} \quad \implies \quad p \text{ is irreducible.}$$

Proof. Let $p \in R$ be prime. If $p = 0$ then we're done because (in my opinion) 0 is irreducible. So let $p \neq 0$ and assume for contradiction that we have $p = ab$, where a and b are non-zero non-units. Since p is prime we have $p|a$ or $p|b$. Without loss of generality, suppose that $p|a$. Then the facts that $p|a$ and $a|p$ imply that $p = au$ for some unit $u \in R^\times$. Finally, we have

$$\begin{aligned} \not{p}b &= \not{p}u \\ b &= u, \end{aligned}$$

which contradicts the fact that $b \notin R^\times$. □

But irreducible elements are not prime in general. For example, consider again the domain $\mathbb{Z} + x\mathbb{Q}[x]$. I claim that the element x is irreducible but not prime. Indeed, a polynomial of degree 1 over a domain is always irreducible. To see that x is not prime, first note that x divides the product $2 \cdot (x/2)$. But $x \nmid 2$ (for reasons of degree) and $x \nmid (x/2)$ because $1/2$ is not in the ring. This example was necessarily rather artificial, because of the following theorem.

I call this Euclid's Lemma because he proved it for integers.

Theorem (Euclid's Lemma). Let R be a PID. Then for all $p \in R$ we have

$$p \text{ is prime} \iff p \text{ is irreducible.}$$

Fancy Proof. We already proved that every prime element in a domain is irreducible. For the other direction, let $p \in R$ be irreducible. By definition this means that the ideal $pR \subsetneq R$ is maximal among principal ideals. Since R is a PID this means that pR is maximal among all ideals. Finally, since every maximal ideal is prime we conclude that pR is a prime ideal, hence $p \in R$ is a prime element. □

Euclid's Proof. Let $p \in R$ be irreducible and assume that $p|ab$ for some $a, b, \in R$, say $pk = ab$. We will show that $p \nmid a$ implies $p|b$. So suppose that $a \notin pR$, which means that $pR \subsetneq pR + aR$. Since R is a PID we know that pR is a maximal ideal, hence $pR + aR = R$. In other words, there exist elements $x, y \in R$ such that $px + ay = 1$. Now multiply both sides by b to obtain

$$\begin{aligned} px + ay &= 1 \\ pbx + (ab)y &= b \\ pbx + (pk)y &= b \\ p(bx + ky) &= b. \end{aligned}$$

We conclude that $p|b$ as desired. □

[**Jargon:** We say that elements $a, b \in R$ are *coprime* if $aR + bR = R$, or, in other words, if there exist elements $x, y \in R$ such that $ax + by = 1$. If the ring R is Euclidean then one can use the so-called Euclidean Algorithm to find some specific elements x, y.]

Finally, we can prove that every element in a PID has a unique factorization into irreducibles. The proof of this result should be taken as the motivation for all of the previous definitions.

Theorem (PID ⇒ UFD). Let R be a PID. We showed previously that every element can be factored as a product of irreducibles, times a unit. Now suppose that we have

$$p_1 p_2 \cdots p_k = u \cdot q_1 q_2 \cdots q_\ell$$

where u is a unit and where the elements $p_1, \ldots, p_k, q_1, \ldots, q_\ell$ are irreducible. Then I claim that $k = \ell$ and we can relabel the factors so that $p_i \sim q_i$ are associate for all i.

Proof. We use induction on $\min\{k, \ell\}$. For the base case, let $\ell = 0$, so we have $p_1 \cdots p_k = u$. If $k \neq 0$ then $p_1 | u$ and $u | p_1$ imply that p_1 is a unit, contradicting the fact that p_1 is irreducible. For the general case, assume that

$$p_1 p_2 \cdots p_k = u \cdot q_1 q_2 \cdots q_\ell.$$

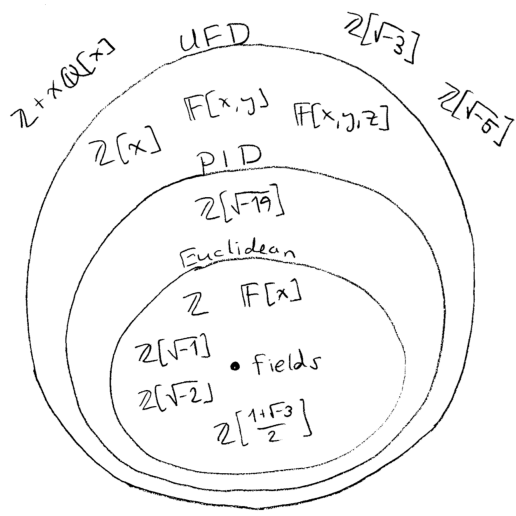
Since $p_1 | q_1 \cdots q_\ell$ it follows from Euclid's Lemma that $p_1 | q_i$ for some i. Without loss, suppose that $p_1 | q_1$. Since q_1 is irreducible and p_1 is not a unit we must have $q_1 = p_1 u'$ for some unit $u' \in R^\times$. Then since $p_1 \neq 0$ we can cancel p_1 from both sides to obtain

$$\begin{aligned} \cancel{p_1} p_2 \cdots p_k &= uu' \cdot \cancel{p_1} q_2 \cdots q_\ell \\ p_2 \cdots p_k &= uu' \cdot q_2 \cdots q_\ell \end{aligned}$$

Since $\min\{k-1, \ell-1\} < \min\{k, \ell\}$, we have by induction that $k - 1 = \ell - 1$ and we can reorder the factors so that $p_i \sim q_i$ are associate for all $i \geq 2$. □

By the way, any domain that satisfies the conclusion of this theorem is called a *unique factorization domain* (UFD).

In summary, here is a sketch of the different kinds of integral domains:



Remarks:

- As you see, each of the inclusions is strict:

$$\{\text{Fields}\} \subsetneq \{\text{Euclidean Domains}\} \subsetneq \{\text{PIDs}\} \subsetneq \{\text{UFDs}\} \subsetneq \{\text{Domains}\}.$$

- Carl Friedrich Gauss proved that the ring $\mathbb{Z}[x]$ is also a UFD, even though $\mathbb{Z}$ is not a field. See Exercise 18.C for the surprisingly tricky proof. More generally, one can use the same argument to show that $R[x]$ is a UFD whenever R is a UFD. Then since $\mathbb{F}[x]$ is a UFD it follows that $\mathbb{F}[x,y] = \mathbb{F}[x][y]$ is a UFD, and by induction the ring of polynomials in any (finite) number of variables over a field is a UFD.
- The rings $\mathbb{Z}[\sqrt{d}]$ for negative integers $d < 0$ are well understood. (Technically: If $d = 1 \bmod 4$ then we should replace $\mathbb{Z}[\sqrt{d}]$ by the ring $\mathbb{Z}[(1+\sqrt{d})/2]$, which has nicer properties. One such nice property is that PID $\Leftrightarrow$ UFD.) Gauss proved that these rings have unique factorization when

$$d \in \{-1, -2, -3, -7, -11, -19, -43, -67, -163\}.$$

  It is a modern theorem of Baker, Heenger and Stark that for all other $d < 0$ the ring of integers does **not** have unique factorization.
- For $d > 0$ it is an unsolved problem to determine when $\mathbb{Z}[\sqrt{d}]$ has unique factorization. Number theory is hard.
- Understanding polynomials in one variable over a field is easier, so we return to that topic next week.

Exercises

17.A The Definition of PIDs is Good

For any ring R prove that

$$(R \text{ is a field}) \quad \iff \quad (R[x] \text{ is a PID}).$$

17.B Quadratic Field Extensions, Part II

Let $\mathbb{E} = \mathbb{F}(\iota) \supseteq \mathbb{F}$ for some element $\iota \in \mathbb{E}$ satisfying $\iota \notin \mathbb{F}$ and $\iota^2 \in \mathbb{F}$. Recall that the vector space $\mathbb{E}/\mathbb{F}$ has basis $\{1, \iota\}$ and the Galois group $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ is generated by the "conjugation" automorphism $(a+b\iota)^* \coloneqq a - b\iota$.

(a) For any $\alpha \in \mathbb{E}$ show that $\alpha \in \mathbb{F}$ if and only if $\alpha^* = \alpha$. Use this to show that $\alpha\alpha^*$ and $\alpha + \alpha^*$ are in $\mathbb{F}$ for all $\alpha \in \mathbb{E}$.

(b) For any polynomial $f(x) = \sum_i \alpha_i x^i \in \mathbb{E}[x]$ we define $f^*(x) \coloneqq \sum_i \alpha_i^* x^i$. Show that this is a ring automorphism $* : \mathbb{E}[x] \to \mathbb{E}[x]$. Use this to prove that $f(x)f^*(x)$ and $f(x) + f^*(x)$ are in $\mathbb{F}[x]$ for all $f(x) \in \mathbb{E}[x]$.

(c) For all $f(x) \in \mathbb{F}[x]$ show that the roots of $f(x)$ in $\mathbb{E} - \mathbb{F}$ come in conjugate pairs.

(d) *Application.* Let $f(x) \in \mathbb{F}[x]$ have degree 3. If f has a root in $\mathbb{E}$, prove that f also has a root in $\mathbb{F}$. [Hint: Use Descartes' Factor Theorem.]

17.C Wilson's Theorem

We saw in the previous problem that any ring homomorphism $\varphi : R \to S$ extends to a ring homomorphism $\varphi : R[x] \to S[x]$ by acting on coefficients. Now let $p \in \mathbb{Z}$ be prime and consider the following polynomial with integer coefficients:

$$f(x) \coloneqq (x^{p-1} - 1) - \prod_{k=1}^{p-1} (x - k) \in \mathbb{Z}[x].$$

(a) Let $\pi : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ be the quotient homomorphism. Prove that the polynomial $f^\pi(x) \in (\mathbb{Z}/p\mathbb{Z})[x]$ has $p-1$ distinct roots and degree $< p-1$. [Hint: Fermat's Little Theorem.]

(b) Use Descartes' Factor Theorem to show that every coefficient of $f(x) \in \mathbb{Z}[x]$ is a multiple of p. Show that this implies $(p-1)! = -1 \bmod p$.

17.D Gaussian Integers

The following theorem is mostly due to Fermat:

> An integer $n \in \mathbb{N}$ is a sum of two squares if and only if any prime factor $p|n$ satisfying $p = 3 \bmod 4$ occurs to an even power.

In this problem we will give a mostly algebraic proof due to Gauss. Let $i \in \mathbb{C}$ be any square root of -1 and consider the following ring extension of $\mathbb{Z}$, called the ring of *Gaussian integers*:

$$\mathbb{Z} \subseteq \mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\} \subseteq \mathbb{C}.$$

(a) Let $N : \mathbb{Z}[i] \to \mathbb{N}$ be the "norm" function defined by $N(a+ib) \coloneqq a^2 + b^2$. Prove that $(\mathbb{Z}[i], N)$ is a Euclidean domain, hence $\mathbb{Z}[i]$ is a UFD. [Hint: For any $\alpha, \beta \in \mathbb{Z}[i]$ with $\beta \neq 0$, the ideal $\beta\mathbb{Z}[i]$ is a "square lattice" in $\mathbb{C}$ with (squared) side length $N(\beta)$. Let $\beta\zeta$ be the closest element of $\beta\mathbb{Z}[i]$ to α and observe that $N(\alpha - \beta\zeta) < N(\beta)$.]

(b) For all $\alpha, \beta \in \mathbb{Z}[i]$ prove that $N(\alpha\beta) = N(\alpha)N(\beta)$. Use this to show that

$$\mathbb{Z}[i]^\times = \{\alpha \in \mathbb{Z}[i] : N(\alpha) = 1\} = \{\pm 1, \pm i\}.$$

(c) For all $n \in \mathbb{N}$ show that $n = 3 \bmod 4$ implies $n \notin \operatorname{im} N$. [Hint: What are the square elements of the ring $\mathbb{Z}/4\mathbb{Z}$?]

(d) Use induction on n to prove the following statement:

$$n \in \operatorname{im} N \Rightarrow (\text{every prime } p|n \text{ with } p = 3 \bmod 4 \text{ occurs to an even power}).$$

[Hint: Let $n = a^2 + b^2 \in \operatorname{im} N$ and let $p \in \mathbb{Z}$ be prime. If $p = 3 \bmod 4$ use (b) and (c) to show that p is irreducible in $\mathbb{Z}[i]$. Then if $p|n$ use (a)

to show that $p|(a+bi)$ or $p|(a-bi)$ in $\mathbb{Z}[i]$. In either case show that $p|a$ and $p|b$, hence $n/p^2 \in \operatorname{im} N$.]

(e) Conversely, for prime $p \in \mathbb{N}$ show that $p = 1 \bmod 4$ implies $p \in \operatorname{im} N$. [Hint: Let $p = 4k+1$ and assume for contradiction that $p \notin \operatorname{im} N$. Use (a) and (b) to show that p is irreducible and hence prime in $\mathbb{Z}[i]$. On the other hand, set $m := (2k)!$ and use Wilson's Theorem to show that $p|(m-i)(m+i)$.]

(f) Finish the proof.

[**Remark:** The expression as a sum of squares is not necessarily unique. Lagrange actually gave a formula for the number of distinct representations of $n \in \mathbb{N}$ as a sum of squares:

$$2\left(1+\left(\frac{-1}{n}\right)\right)\sum_{d|n}\left(\frac{-1}{d}\right).$$

Here the notation $\left(\frac{a}{b}\right)$ is called the *Jacobi symbol* and I am not going to define it.]

17.E $\mathbb{Z}[\sqrt{-3}]$ is not a UFD

Let $\sqrt{-3} \in \mathbb{C}$ be a fixed square root of -3 and consider the ring

$$\mathbb{Z} \subseteq \mathbb{Z}[\sqrt{-3}] = \{a + b\sqrt{-3} : a, b \in \mathbb{Z}\} \subseteq \mathbb{C}.$$

(a) Let $N : \mathbb{Z}[\sqrt{-3}] \to \mathbb{N}$ be defined by $N(a+b\sqrt{-3}) := a^2 + 3b^2$. For all $\alpha, \beta \in \mathbb{Z}[\sqrt{-3}]$ prove that $N(\alpha\beta) = N(\alpha)N(\beta)$ and use this to show that

$$\mathbb{Z}[\sqrt{-3}]^\times = \{\alpha \in \mathbb{Z}[\sqrt{-3}] : N(\alpha) = 1\} = \{\pm 1\}.$$

(b) Prove that there is no element $\alpha \in \mathbb{Z}[\sqrt{-3}]$ with $N(\alpha) = 2$. Use this to show that any element with $N(\alpha) = 4$ is irreducible. In particular, $2 \in \mathbb{Z}[\sqrt{-3}]$ is irreducible.

(c) But show that $2 \in \mathbb{Z}[\sqrt{-3}]$ is **not prime** because

$$2|(1+\sqrt{-3})(1-\sqrt{-3}) \text{ and } 2 \nmid (1+\sqrt{-3}) \text{ and } 2 \nmid (1-\sqrt{-3}).$$

(d) Use this to prove that the following ideal is **not principal**:

$$\{2\alpha + (1+\sqrt{-3})\beta : \alpha, \beta \in \mathbb{Z}[\sqrt{-3}]\} \subseteq \mathbb{Z}[\sqrt{-3}].$$

[**Remark:** The previous two problems are part of a tricky subject called *algebraic number theory*. We will now leave this subject behind, since any further discussion would lead us away from the main goals of this course.]

Week 18

18.1 Universal Property of Polynomials

In the 1700s, Enlightment mathematicians such as Leonhard Euler took for granted the nature and existence of the basic number systems:

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}.$$

These mathematical objects were regarded as real in the same sense that the physical world is real. In the 1800s, mathematicians began to doubt their intuition[129] and they started to ask for more rigorous definitions of basic concepts.

The first major post-Enlightment mathematician was Augustin-Louis Cauchy (1789–1857). His textbook *Cours d'Analyse* (1821) gave the first rigorous treatment of calculus. Later, in 1847, he gave the first rigorous definition of the complex numbers. Assuming that the real numbers exist, Cauchy defined the complex numbers as a quotient ring:

$$\mathbb{C} \coloneqq \frac{\mathbb{R}[x]}{(x^2+1)\mathbb{R}[x]}.$$

Since the polynomial x^2+1 is an irreducible element of the PID $\mathbb{R}[x]$ we know that the ideal $(x^2+1)\mathbb{R}[x] \subseteq \mathbb{R}[x]$ is maximal, hence the quotient ring is a field. The role of the imaginary unit "$\sqrt{-1}$" is played by the coset of x:

$$\sqrt{-1} \coloneqq x + (x^2+1)\mathbb{R}[x] = \{x + (x^2+1)f(x) : f(x) \in \mathbb{R}[x]\}.$$

Indeed, this coset is nonzero and we can check that it is a square root of the coset of -1:

$$\big(x + (x^2+1)\mathbb{R}[x]\big)^2 = \big(x^2 + (x^2+1)\mathbb{R}[x]\big) = \big(-1 + (x^2+1)\mathbb{R}[x]\big).$$

This level of abstraction was too much for Cauchy's contemporaries but it was later taken up in the 1880s by Leopold Kronecker. This week I will present Kronecker's proof that every polynomial over a field has a root in some extension field.

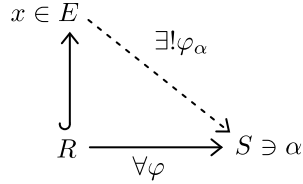
[129] There were many reasons, but perhaps the most important was the discovery of self-consistent "non-Euclidean geometries" by Gauss, Bolyai and Lobachevsky.

In order to do this we need a more modern definition of polynomials.

Theorem/Definition (Universal Property of Polynomials). Let R be a ring and let x be an indeterminate. We say that a ring E is a/the *free R-algebra generated by x* if the following two properties hold:

(1) We have $x \in E$ and we have a subring $R \subseteq E$ isomorphic to R.

(2) For any ring homomorphism $\varphi : R \to S$ and for any element $\alpha \in S$ there exists a unique ring homomorphism $\varphi_\alpha : E \to S$ satisfying $\varphi_\alpha(x) = \alpha$ and $\varphi_\alpha(a) = a$ for all $a \in R$. In other words, there exists a unique φ_α making the following diagram commute:



Note that the polynomial ring $R[x]$ satisfies (1) and (2). Furthermore, if E and E' are any two rings satisfying (1) and (2) then there exists a unique ring isomorphism $E \cong E'$ fixing the subset $R \cup \{x\}$. In this sense, we can say that

$$R[x] \text{ is the unique free } R\text{-algebra generated by } x.$$

///

Proof. The fact that $R[x]$ satisfies (1) and (2) is easy. For (1) we can think of $R \subseteq R[x]$ as the subring of constant polynomials. For (2) suppose that $\varphi_\alpha : R[x] \to S$ is any ring homomorphism satisfying $\varphi_\alpha(x) = \alpha$ and $\varphi_\alpha(a) = a$ for all $a \in R$. Then we must have

$$\varphi_\alpha\left(\sum_i a_i x^i\right) = \sum_i \varphi_\alpha(a_i)\varphi_\alpha(x)^i = \sum_i \varphi(a_i)\alpha^i.$$

And it is easy to check that this function is indeed a ring homomorphism, as long as $\alpha \in S$ commutes with the elements of the subring $\operatorname{im}\varphi \subseteq S$.[130]

The surprising thing is that properties (1) and (2) determine the ring $R[x]$ up to isomorphism. I will only sketch the proof of this and you are free to skip it. So let E and E' be two rings satisfying (1) and (2). From (1) we note that $R \cup \{x\}$ is a subset of E and E' and from (2) we note that the identity maps $id_E : E \to E$ and $id_{E'} : E' \to E'$ are the **unique** ring homomorphisms $E \to E$ and $E' \to E'$ fixing the subset $R \cup \{x\}$. Also from (2) we know that there **exist** ring homomorphisms $\phi : E \to E'$ and $\psi : E' \to E$ fixing $R \cup \{x\}$.

[130] For us this condition is automatic because we assume that S is a commutative ring.

Since the compositions fix $R \cup \{x\}$, we conclude from the previous remark that $\phi \circ \psi = id_E$ and $\psi \circ \phi = id_{E'}$, hence $E \cong E'$. □

Remarks:

- This definition is an example of a "universal property". You will see another example on the homework when you study the field of fractions of a domain. The concept of universal properties was promoted by Saunders Mac Lane in the 1940s and 1950s, so it is strictly speaking a bit too modern for this course.[131] Nevertheless, the universal property of polynomials is important for Galois theory.
- Instead of telling us what a polynomial "is", the universal property tells us what a polynomial "does". Sometimes this is more important.
- To be explicit, the purpose of a polynomial is to be "evaluated". Given a ring homomorphism $\varphi : R \to S$, there exists a unique ring homomorphism $\varphi : R[x] \to S[x]$ acting by φ on the coefficients and sending $x \mapsto \alpha$. We denote this map by $f(x) \mapsto f^{\varphi}(x)$. Then for any element $\alpha \in S$ there exists a unique ring homomorphism $\varphi_\alpha : R[x] \to S[x] \to S$ defined by "evaluating the polynomial $f^{\varphi}(x)$ at the argument $x = \alpha$". ///

18.2 The Minimal Polynomial Theorem

Last time we proved that for any ring homomorphism $\varphi : R \to S$ and for any element $\alpha \in S$ there exists a ring homomorphism $\varphi_\alpha : R[x] \to S$ acting on the coefficients by φ and sending $x \mapsto \alpha$. Today we will focus on the special case when φ is just the identity homomorphism on a subring $R \subseteq S$.

Definition of Evaluation. Let $R \subseteq S$ be a subring and let $id : R \hookrightarrow S$ be the restriction of the identity homomorphism $S \to S$. Then for any element $\alpha \in S$ we have a homomorphism $id_\alpha : R[x] \to S$ defined by fixing the coefficients and sending $x \mapsto \alpha$. For any polynomial $f(x) = \sum_i a_i x^i \in R[x]$ we will use the notation

$$f(\alpha) \coloneqq id_\alpha(f(x)) = id_\alpha\left(\sum_i a_i x^i\right) = \sum_i a_i \alpha^i.$$

We call id_α the *evaluation homomorphism at* $x = \alpha$. ///

I claim that the image of the evaluation $id_\alpha : R[x] \to S$ is equal to the smallest subring of S that contains the set $R \cup \{\alpha\}$:

$$\mathrm{im}(id_\alpha) = R[\alpha] \subseteq S.$$

[131] Saunders Mac Lane was an American mathematician who studied at Göttingen in the 1930s. After the war he co-founded with Samuel Eilenberg the subject of *category theory*. The extreme abstraction of categories was shortly taken up by French mathematicians and became part of the mathematical mainstream in the 1960s.

Proof. Since $R[\alpha]$ is a subring containing $R \cup \{\alpha\}$ and since every element of $\mathrm{im}(id_\alpha)$ is formed from $R \cup \{\alpha\}$ using a finite number of ring operations, we have $\mathrm{im}(id_\alpha) \subseteq R[\alpha]$. Conversely, since $\mathrm{im}(id_\alpha) \subseteq S$ is a subring containing $R \cup \{\alpha\}$ and since $R[\alpha] \subseteq S$ is the **smallest** subring containing $R \cup \{\alpha\}$ we must have $R[\alpha] \subseteq \mathrm{im}(id_\alpha)$. □

The kernel of an evaluation is more complicated, so we will restrict our attention to polynomials over a **field**. Then there are two basic cases, called *transcendental* and *algebraic*.

Definition of the Minimal Polynomial. Let $\alpha \in \mathbb{E} \supseteq \mathbb{F}$ be an element of a field extension and consider the evaluation homomorphism $id_\alpha : \mathbb{F}[x] \to \mathbb{E}$. By the First Isomorphism Theorem we have

$$\frac{\mathbb{F}[x]}{\ker(id_\alpha)} \cong \mathrm{im}(id_\alpha) = \mathbb{F}[\alpha] \subseteq \mathbb{E}.$$

If $\ker(id_\alpha) = 0$ then we say that α *is transcendental over* $\mathbb{F}$ and we obtain an isomorphism

$$\mathbb{F}[x] \cong \mathbb{F}[\alpha].$$

In other words, we can think of a transcendental element as a "variable".

If $\ker(id_\alpha) \neq 0$ then we say that α *is algebraic over* $\mathbb{F}$. In this case, since $\mathbb{F}[x]$ is a PID there exists a unique monic polynomial $m_{\alpha/\mathbb{F}}(x) \in \mathbb{F}[x]$ satisfying

$$\ker(id_\alpha) = \langle m_{\alpha/\mathbb{F}}(x) \rangle = m_{\alpha/\mathbb{F}}(x)\mathbb{F}[x] = \{m_{\alpha/\mathbb{F}}(x)g(x) : g(x) \in \mathbb{F}[x]\}.$$

We call this $m_{\alpha/\mathbb{F}}(x) \in \mathbb{F}[x]$ the *minimal polynomial of* α *over* $\mathbb{F}$. ///

In less algebraic terms we can say that $m_{\alpha/\mathbb{F}}(x) \in \mathbb{F}[x]$ is the unique monic polynomial of minimal degree that has α as a root.

Proof. Indeed, we have $m_{\alpha/\mathbb{F}}(\alpha) = 0$ by definition. And for any polynomial $f(x) \in \mathbb{F}[x]$ we have

$$f(\alpha) = 0 \quad \Leftrightarrow \quad f(x) \in \ker(id_\alpha) \quad \Leftrightarrow \quad f(x) \in \langle m_{\alpha/\mathbb{F}}(x) \rangle \quad \Leftrightarrow \quad m_{\alpha/\mathbb{F}}(x) | f(x).$$

If $f(\alpha) = 0$ with $f(x) \in \mathbb{F}[x]$ monic, then since $m_{\alpha/\mathbb{F}}(x) | f(x)$ and $f(x) \neq 0$ we conclude that $\deg(m_{\alpha/\mathbb{F}}) \leq \deg(f)$. □

The most basic case occurs when $\alpha \in \mathbb{F}$. In this case I claim that the minimal polynomial is just $m_{\alpha/\mathbb{F}}(x) = x - \alpha \in \mathbb{F}[x]$. Indeed, we already know from Descartes' Factor Theorem that for all $f(x) \in \mathbb{F}[x]$ we have

$$f(\alpha) = 0 \quad \Longleftrightarrow \quad (x - \alpha) | f(x).$$

Thus we can view the concept of the minimal polynomial as some kind of generalization of Descartes' Theorem.

The following theorem is one of our main tools for studying field extensions.

The Minimal Polynomial Theorem. Let $\mathbb{E} \supseteq \mathbb{F}$ be a field extension and let $\alpha \in \mathbb{E}$ be algebraic over $\mathbb{F}$ with minimal polynomial $m_{\alpha/\mathbb{F}}(x) \in \mathbb{F}[x]$.

(1) If $f(x) \in \mathbb{F}[x]$ is irreducible and monic with $f(\alpha) = 0$ then $f(x) = m_{\alpha/\mathbb{F}}(x)$.

(2) The minimal polynomial $m_{\alpha/\mathbb{F}}(x) \in \mathbb{F}[x]$ is irreducible and it follows that $\mathbb{F}[\alpha]$ is a field. In other words, we have

$$\mathbb{F}[\alpha] = \mathbb{F}(\alpha).$$

(3) If $\deg(m_{\alpha/\mathbb{F}}) = n$ then I claim that $\{1, \alpha, \ldots, \alpha^{n-1}\}$ is a basis for the vector space $\mathbb{F}(\alpha)$ over $\mathbb{F}$. It follows that

$$[\mathbb{F}(\alpha)/\mathbb{F}] = \deg(m_{\alpha/\mathbb{F}}).$$

///

Proof. (1) Suppose that we have $f(\alpha) = 0$ for some $f(x) \in \mathbb{F}[x]$. By definition this means that $m_{\alpha/\mathbb{F}}(x) | f(x)$. If $f(x)$ is irreducible then this implies that $f(x) = \lambda \cdot m_{\alpha/\mathbb{F}}(x)$ for some nonzero constant $\lambda \in \mathbb{F}$ and if $f(x)$ is monic then we must have $\lambda = 1$, hence $f(x) = m_{\alpha/\mathbb{F}}(x)$.

(2) To prove that $m_{\alpha/\mathbb{F}}(x)$ is irreducible, suppose that we have $m_{\alpha/\mathbb{F}}(x) = f(x)g(x)$ for some non-constant polynomials $f(x), g(x) \in \mathbb{F}[x]$. In particular, we have $\deg(f) < \deg(m_{\alpha/\mathbb{F}})$ and $\deg(g) < \deg(m_{\alpha/\mathbb{F}})$. Then evaluating at $x = \alpha$ gives

$$f(\alpha)g(\alpha) = m_{\alpha/\mathbb{F}}(\alpha) = 0.$$

Since $\mathbb{F}$ is a domain this implies that $f(\alpha) = 0$ or $g(\alpha) = 0$. Without loss of generality suppose that $f(\alpha) = 0$, so that $m_{\alpha/\mathbb{F}}(x) | f(x)$. But then since $f(x) \neq 0$ we must have $\deg(m_{\alpha/\mathbb{F}}) \leq \deg(f)$, which is a contradiction.

Now recall that $\mathbb{F}[\alpha]$ and $\mathbb{F}(\alpha)$ are by definition the smallest sub**ring** and sub**field** of $\mathbb{E}$ that contain the set $\mathbb{F} \cup \{\alpha\}$. Since every subfield is a subring we have $\mathbb{F}[\alpha] \subseteq \mathbb{F}(\alpha)$. Conversely, since $m_{\alpha/\mathbb{F}}(x) \in \mathbb{F}[x]$ is irreducible in a PID we know that $\langle m_{\alpha/\mathbb{F}}(x) \rangle \subseteq \mathbb{F}[x]$ is a maximal ideal and hence

$$\frac{\mathbb{F}[x]}{\langle m_{\alpha/\mathbb{F}}(x) \rangle} = \frac{\mathbb{F}[x]}{\ker(id_\alpha)} \cong \mathrm{im}(id_\alpha) = \mathbb{F}[\alpha] \text{ is a field.}$$

Then since $\mathbb{F}[\alpha] \subseteq \mathbb{E}$ is a subfield that contains $\mathbb{F} \cup \{\alpha\}$ we conclude that $\mathbb{F}(\alpha) \subseteq \mathbb{F}[\alpha]$.

(3) Let $\deg(m_{\alpha/\mathbb{F}}) = n$ and consider the set $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\} \subseteq \mathbb{F}(\alpha)$. To show that this set **spans** $\mathbb{F}(\alpha)$ over $\mathbb{F}$ we observe that every element of $\mathbb{F}(\alpha) = \mathbb{F}[\alpha] = \text{im}(id_\alpha)$ has the form $f(\alpha)$ for some polynomial $f(x) \in \mathbb{F}[x]$. From the Division Theorem there exist polynomials $q(x), r(x) \in \mathbb{F}[x]$ with

$$f(x) = m_{\alpha/\mathbb{F}}(x)q(x) + r(x) \quad \text{and} \quad \deg(r) < \deg(m_{\alpha/\mathbb{F}}).$$

Since $\deg(r) < \deg(m_{\alpha/\mathbb{F}}) = n$ we can write $r(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1}$ for some $a_0, \ldots, a_{n-1} \in \mathbb{F}$ and then evaluating at $x = \alpha$ gives

$$\begin{aligned} f(\alpha) &= m_{\alpha/\mathbb{F}}(\alpha)q(\alpha) + r(\alpha) \\ &= 0 \cdot q(\alpha) + r(\alpha) \\ &= r(\alpha) \\ &= a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}. \end{aligned}$$

Finally, to show that $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ is **independent** over $\mathbb{F}$, suppose that

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} = 0 \quad \text{for some } a_0, \ldots, a_{n-1} \in \mathbb{F}.$$

In other words, suppose we have $f(\alpha) = 0$ for some polynomial $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} \in \mathbb{F}[x]$. In this case I claim that $f(x) = 0$ is the zero polynomial and hence $a_0 = a_1 = \cdots = a_{n-1} = 0$. Indeed, since $f(\alpha) = 0$ we have $m_{\alpha/\mathbb{F}}(x) | f(x)$. If $f(x) \neq 0$ then this implies that $\deg(m_{\alpha/\mathbb{F}}) \leq \deg(f)$, which contradicts the fact that $\deg(f) < n$. □

You investigated a special case of this theorem on a previous homework. Now we will relate this example to the theory of minimal polynomials.

Example: Quadratic Field Extensions. Let $\mathbb{E} \supseteq \mathbb{F}$ be a field extension and consider an element $\alpha \in \mathbb{E}$ with $\alpha^2 \in \mathbb{F}$ and $\alpha \notin \mathbb{F}$. Then I claim that $x^2 - \alpha^2 \in \mathbb{F}[x]$ is the minimal polynomial of α over $\mathbb{F}$. Since $x^2 - \alpha^2$ is monic and has α as a root, it suffices to show that this polynomial is irreducible. So assume for contradiction that we have $x^2 - \alpha^2 = f(x)g(x)$ for some non-constant polynomials $f(x), g(x) \in \mathbb{F}[x]$. This implies that $\deg(f) = \deg(g) = 1$. If $f(x) = ax + b$ with $a, b \in \mathbb{F}$ and $a \neq 0$ then we see that $-b/a \in \mathbb{F}$ is a root of $f(x)$, hence also a root of $x^2 - \alpha^2$. But from Descartes' Factor Theorem we know that $\pm\alpha$ are the **only** roots of $x^2 - \alpha^2$ and by assumption these are not in $\mathbb{F}$.

Then since $m_{\alpha/\mathbb{F}}(x) = x^2 - \alpha^2$ has degree 2 we conclude from the Minimal Polynomial Theorem that $[\mathbb{F}(\alpha)/\mathbb{F}] = 2$ with basis $\{1, \alpha\}$, and it follows that

$$\mathbb{F}(\alpha) = \{a + b\alpha : a, b \in \mathbb{F}\}.$$

Recall that we originally proved this result by "rationalizing the denominator". The new method is better because it extends to more general situations. ///

And here is one of those more general situations.

Example: The Minimal Polynomial of $\sqrt[3]{2}$ Over $\mathbb{Q}$. Let $\alpha = \sqrt[3]{2} \in \mathbb{R}$ be the unique real cube root of 2. I claim that the minimal polynomial of α over $\mathbb{Q}$ is

$$m_{\alpha/\mathbb{Q}}(x) = x^3 - 2.$$

Proof. Since α is a root of $x^3 - 2$, we only need to check that $x^3 - 2 \in \mathbb{Q}[x]$ is irreducible. So suppose for contradiction that $x^3 - 2 = f(x)g(x)$ for some non-constant polynomials $f(x), g(x) \in \mathbb{Q}[x]$. By considering degrees we must have $\deg(f) = 1$ or $\deg(g) = 1$, and then it follows as in the previous example that $x^3 - 2$ has a root in $\mathbb{Q}$. To be specific, suppose that $(a/b)^3 - 2 = 0$ for some $a, b \in \mathbb{Z}$ with $\gcd(a, b) = 1$. Then we have

$$\begin{aligned} a^3/b^3 - 2 &= 0 \\ a^3/b^3 &= 2 \\ a^3 &= 2b^3. \end{aligned}$$

Since $b|a^3$ with $\gcd(a, b) = 1$ we must have $b \in \{\pm 1\}$. And since $a|2b^3$ with $\gcd(a, b) = 1$ we must have $a \in \{\pm 1, \pm 2\}$. It follows that

$$a/b \in \{\pm 1, \pm 2\},$$

and one can check that that none of these is a root of $x^3 - 2$. [**Remark:** This method is called the Rational Root Test. We will give the general statement below.] □

Finally, since $m_{\alpha/\mathbb{Q}}(x) = x^3 - 2$ has degree 3 we conclude from the Minimal Polynomial Theorem that $\{1, \alpha, \alpha^2\}$ is a basis for the field $\mathbb{Q}(\alpha)$ over $\mathbb{Q}$, and it follows that

$$\mathbb{Q}(\alpha) = \{a + b\alpha + c\alpha^2 : a, b, c \in \mathbb{Q}\}.$$

In particular, we conclude that the set on the right is a **field**. However, you would find it very difficult to "rationalize the denominator" by hand:

$$\frac{1}{a + b\alpha + c\alpha^2} = (?) + (?)\alpha + (?)\alpha^2.$$

I set up a 3×3 linear system and used my computer to find that

$$\frac{1}{a + b\alpha + c\alpha^2} = \left(\frac{a^2 - 2bc}{\Delta}\right) + \left(\frac{2c^2 - ab}{\Delta}\right)\alpha + \left(\frac{b^2 - ac}{\Delta}\right)\alpha^2,$$

with $\Delta = a^3 + 2b^3 + 4c^3 - 6abc$. This formula is probably not useful for anything. ///

18.3 Kronecker's Theorem

Let $f(x) \in \mathbb{F}[x]$ be any irreducible polynomial and suppose that there exists a field extension $\mathbb{E} \supseteq \mathbb{F}$ and an element $\alpha \in \mathbb{E}$ such that $f(\alpha) = 0$. Last time we proved that $\langle f(x) \rangle$ is the kernel of the evaluation homomorphism $id_\alpha : \mathbb{F}[x] \to \mathbb{E}$. Then since $\langle f(x) \rangle \subseteq \mathbb{F}[x]$ is a maximal ideal we obtain an isomorphism of fields:

$$\mathbb{F}[x]/\langle f(x) \rangle = \mathbb{F}[x]/\ker(id_\alpha) \cong \mathrm{im}(id_\alpha) = \mathbb{F}(\alpha) \subseteq \mathbb{E}.$$

Furthermore, we observe that the coset $x + \langle f(x) \rangle \in \mathbb{F}[x]/\langle f(x) \rangle$ gets identified with the root $\alpha \in \mathbb{E}$. But what if we don't know any roots of $f(x)$? Today we will reverse this construction and use it to **create a root** for any given polynomial over a field.

Leopold Kronecker is known as a "constructivist" mathematician, meaning that he would not accept the existence of a mathematical entity unless he could give a finite algorithm for constructing it. His contemporary Dedekind, on the other hand, was happy to accept infinite sets implicitly defined by satisfying certain conditions. Dedekind's point of view eventually became standard but the following construction of Kronecker is still imporant. Kronecker's original goal was to give a concrete way to work with algebraic irrational numbers such as $\sqrt{2}$.[132]

Kronecker's Theorem (Every Polynomial Has a Root Somewhere). Let $\mathbb{F}$ be a field[133] and let $f(x) \in \mathbb{F}[x]$ be a polynomial of degree ≥ 1. Then there exists a field extension $\mathbb{E} \supseteq \mathbb{F}$ in which $f(x)$ has a root.

Proof. Let $f(x) \in \mathbb{F}[x]$ be a polynomial of degree ≥ 1. Since $\mathbb{F}[x]$ is a PID we know that we can write $f(x) = p(x)g(x)$ with $p(x), g(x) \in \mathbb{F}[x]$ and with $p(x)$ **irreducible**. Suppose that we can find an extension $\mathbb{E} \supseteq \mathbb{F}$ and an element $\alpha \in \mathbb{E}$ such that $p(\alpha) = 0$. Then this α is also a root of $f(x)$ because "evaluation at $x = \alpha$" is a ring homomorphism:

$$f(\alpha) = p(\alpha)g(\alpha) = 0g(\alpha) = 0.$$

In order to construct such a field $\mathbb{E}$ and element $\alpha \in \mathbb{E}$ we consider the principal ideal

$$\langle p(x) \rangle = p(x)\mathbb{F}[x] = \{p(x)g(x) : g(x) \in \mathbb{F}[x]\}.$$

Again, since $p(x) \in \mathbb{F}[x]$ is irreducible in a PID we know that $\langle p(x) \rangle \subseteq \mathbb{F}[x]$ is a **maximal ideal**, hence the quotient ring is a **field**. We will call it $\mathbb{E}$:

$$\mathbb{E} \coloneqq \mathbb{F}[x]/\langle p(x) \rangle.$$

[132] The method doesn't help with transcendental numbers such as π. It is said that Kronecker did not believe in such numbers.

[133] The result also applies to polynomials over an integral domain R by taking $\mathbb{F} = \mathrm{Frac}(R) \supseteq R$. As always, non-domains are a different story. I'm starting to think that the concept of "rings" is too broad.

I claim that the coset $\alpha = x + \langle p(x)\rangle \in \mathbb{E}$ is the desired root of $f(x)$. Wait a minute, that sounds ridiculous. How can a coset be a root?

First we need to view $\mathbb{F}$ as a subfield of $\mathbb{E}$. So consider the following ring homomorphism:

$$\begin{array}{rccc} \iota : \mathbb{F} & \to & \mathbb{E} \\ a & \mapsto & a + \langle p(x)\rangle. \end{array}$$

Note that this function is injective. Indeed, if $\iota(a) = \iota(b)$ for some $a, b \in \mathbb{F}$ then we have

$$\begin{aligned} \iota(a) &= \iota(b) \\ a + \langle p(x)\rangle &= b + \langle p(x)\rangle \\ a - b &\in \langle p(x)\rangle \\ a - b &= p(x)g(x) \text{ for some } g(x) \in \mathbb{F}[x]. \end{aligned}$$

But note that $\deg(p) \geq 1$ (since $p(x)$ is an irreducible polynomial) and $\deg(a - b) \leq 0$ (since $a - b$ is a constant). Then since $\deg(pg) = \deg(p) + \deg(g)$, the only possible solution is $g(x) = 0$, which implies that $a - b = 0$ as desired.

From the First Isomorphism Theorem we conclude that $\operatorname{im} \iota = \{a + \langle p(x)\rangle : a \in \mathbb{F}\} \subseteq \mathbb{E}$ is a subfield isomorphic to $\mathbb{F}$. Now pay close attention to the following remark:

we choose to **identify** $\mathbb{F}$ *with the subfield* $\{a + \langle p(x)\rangle : a \in \mathbb{F}\} \subseteq \mathbb{E}$.

Now it only remains to show that the element $\alpha = x + \langle p(x)\rangle \in \mathbb{E}$ is a root of the polynomial $p(x) \in \mathbb{F}[x]$. More generally, consider any polynomial $h(x) = \sum_i a_i x^i = \sum_i (a_i + \langle p(x)\rangle) x^i$. Then by definition we have

$$\begin{aligned} h(\alpha) &= h(x + \langle p(x)\rangle) \\ &= \sum_i (a_i + \langle p(x)\rangle)(x + \langle p(x)\rangle)^i \\ &= (\sum_i a_i x^i) + \langle p(x)\rangle \\ &= h(x) + \langle p(x)\rangle. \end{aligned}$$

In particular, this implies that $p(\alpha) = p(x) + \langle p(x)\rangle = 0 + \langle p(x)\rangle$ as desired.[134] □

I apologize for level of abstraction in that proof. Kronecker's Theorem is similar in spirit to the construction of fractions from integers, or the construction of

[134] If you don't like that, here's a different argument. From the universal property of polynomials we know that there exists a unique ring homomorphism $\mathbb{F}[x] \to \mathbb{E}$ sending $x \mapsto x + \langle p(x)\rangle$, called "evaluation at $x + \langle p(x)\rangle$". But note that the quotient map $\mathbb{F}[x] \to \mathbb{F}[x]/\langle p(x)\rangle$ also satisfies this condition! Hence by uniqueness we conclude that the evaluation map equals the quotient map: $h(x) \mapsto h(x) + \langle p(x)\rangle$.

real numbers from fractions. At first we think of a fraction as an infinite equivalence class of ordered pairs of integers. Similarly, we first think of a real number as either a "Dedekind cut" (ordered pair of infinite sets of fractions) or an infinite equivalence class of "Cauchy sequences". However, after we are satisfied with the existence of these objects we always revert to a more concrete point of view.

For example, see the following corollary.

Corollary/Definition (Every Polynomial Has a Splitting Field). Let $\mathbb{F}$ be a field and let $f(x) \in \mathbb{F}[x]$ be a polynomial of degree $n \geq 1$. Then there exists a field $\mathbb{E} \supseteq \mathbb{F}$ and elements $\alpha_1, \ldots, \alpha_n \in \mathbb{E}$ such that

$$f(x) = (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n) \text{ in } \mathbb{E}[x].$$

Recall that we define $\mathbb{F} \subseteq \mathbb{F}(\alpha_1, \ldots, \alpha_n) \subseteq \mathbb{E}$ as the smallest subfield of $\mathbb{E}$ containing the set $\mathbb{F} \cup \{\alpha_1, \ldots, \alpha_n\}$. In the case that

$$\mathbb{F}(\alpha_1, \ldots, \alpha_n) = \mathbb{E}$$

we will say that $\mathbb{E}$ is *a splitting field for* $f(x)$. ///

Proof by Induction. The base case is Kronecker's Theorem. So let $n \geq 2$ and consider a polynomial $f(x) \in \mathbb{F}[x]$ of degree $n \geq 2$. By Kronecker's Theorem there exists a field $\mathbb{K}_1 \supseteq \mathbb{F}$ and an element $\alpha_1 \in \mathbb{K}_1$ such that $f(\alpha_1) = 0$, and then by Descartes' Factor Theorem we have

$$f(x) = (x - \alpha_1)g(x) \text{ for some polynomial } g(x) \in \mathbb{K}_1[x] \text{ of degree } n - 1.$$

Now by induction there exists a field $\mathbb{K} \supseteq \mathbb{K}_1 \supseteq \mathbb{F}$ and elements $\alpha_2, \ldots, \alpha_n \in \mathbb{K}$ such that

$$\begin{aligned} g(x) &= (x - \alpha_2) \cdots (x - \alpha_n) \\ f(x) = (x - \alpha_1)g(x) &= (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n) \text{ in } \mathbb{K}[x]. \end{aligned}$$

Finally, note that $\mathbb{E} \coloneqq \mathbb{F}(\alpha_1, \ldots, \alpha_n) \subseteq \mathbb{K}$ is a splitting field for $f(x)$. □

You may recall from Week 14 that splitting fields are central to the Fundamental Theorem of Galois Theory. Later we will show that if $\mathbb{E} \supseteq \mathbb{F}$ and $\mathbb{E}' \supseteq \mathbb{F}$ are two splitting fields for the same polynomial $f(x) \in \mathbb{F}[x]$ then there exists an isomorphism $\mathbb{E} \cong \mathbb{E}'$ fixing $\mathbb{F}$. However, this isomorphism is **not** unique.[135] In fact, you already know this. The collection of such isomorphisms $\mathbb{E} \cong \mathbb{E}'$ is called the Galois group of $f(x)$ over $\mathbb{F}$.

[135] In other words, the splitting field does not satisfy a "universal property".

Exercises

18.A Invariance of the GCD

Let $\mathbb{F}$ be a field and consider two polynomials $f(x), g(x) \in \mathbb{F}[x]$, not both zero. Since $\mathbb{F}[x]$ is a PID we know that there exists a unique monic polynomial $d(x) \in \mathbb{F}[x]$ such that

$$f(x)\mathbb{F}[x] + g(x)\mathbb{F}[x] = d(x)\mathbb{F}[x],$$

which we call the *greatest common divisor* of $f(x)$ and $g(x)$ in $\mathbb{F}[x]$.

(a) For any field extension $\mathbb{E} \supseteq \mathbb{F}$ prove that the greatest common divisor of $f(x), g(x) \in \mathbb{F}[x]$ is the same, whether computed in $\mathbb{F}[x]$ or $\mathbb{E}[x]$.

(b) We say that polynomials $f(x), g(x) \in \mathbb{F}[x]$ are *coprime* when $\gcd(f, g) = 1$. Prove that

$$\gcd(f, g) \neq 1 \quad \iff \quad f(x) \text{ and } g(x) \text{ have a common root in some field extens}$$

18.B Field of Fractions

In this problem you will show that "integral domain" and "subring of a field" are the same concept. Let R be an integral domain and consider the following set of abstract symbols, called *fractions*:

$$\mathrm{Frac}(R) \coloneqq \left\{ \frac{a}{b} : a, b \in R, b \neq 0 \right\}.$$

(a) Prove that the following relation is an equivalence on the set of fractions:

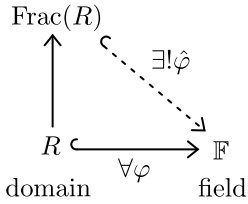
$$\frac{a}{b} \sim \frac{a'}{b'} \quad \iff \quad ab' = a'b.$$

(b) Prove that the following operations are well-defined on equivalence classes:

$$\frac{a}{b} \cdot \frac{c}{d} \coloneqq \frac{ac}{bd} \qquad \text{and} \qquad \frac{a}{b} + \frac{c}{d} \coloneqq \frac{ad + bc}{bd}.$$

It follows that the set of equivalence classes $\mathrm{Frac}(R)/\sim$ is a field. Following tradition, we will just call it $\mathrm{Frac}(R)$ and we will write $=$ instead of $\sim$. Furthermore, we will write $R \subseteq \mathrm{Frac}(R)$ for the image of the injective ring homomorphism $a \mapsto a/1$.

(c) *Universal Property.* Let $\mathbb{F}$ be a field and let $\varphi : R \to \mathbb{F}$ be an **injective** ring homomorphism. Prove that this extends to a unique ring homomorphism $\varphi : \mathrm{Frac}(R) \to \mathbb{F}$, which is also injective. [Hint: Show that $\hat{\varphi}(a/b) \coloneqq \varphi(a)/\varphi(b)$ is well-defined.] Here is a picture:



(d) *Application.* If a field $\mathbb{F}$ contains a subring isomorphic to $\mathbb{Z}$, prove that $\mathbb{F}$ also contains a subfield isomorphic to $\mathbb{Q}$.

[**Remark:** Part (d) fills a gap in our earlier proof characterizing prime subfields. This problem illustrates that the rigorous theory of fractions is subtle.[136] We will usually just follow our intuition.]

18.C Gauss' Lemma

In this problem you will prove that $\mathbb{Z}[x]$ is a unique factorization domain. The key idea of the proof is to consider the greatest common divisor of the coefficients of an integer polynomial. To be specific, for any polynomial $f(x) = a_0 + a_1 + \cdots + a_n x^n \in \mathbb{Z}[x]$ we define the *content*[137] as follows:

$$I(f) = c(a_0 + a_1 x + \cdots + a_n x^n) \coloneqq \gcd(a_0, a_1, \ldots, a_n) \in \mathbb{N}.$$

(a) Let $d = \gcd(a_0, a_1, \ldots, a_n)$ with $a_i = da_i'$ for all i. Prove that

$$\gcd(a_0', a_1', \ldots, a_n') = 1.$$

(b) If $f(x) \in \mathbb{Q}[x]$ is monic, prove that there exists an integer $k \in \mathbb{N}$ with $kf(x) \in \mathbb{Z}[x]$ and $I(kf) = 1$. [Hint: Choose any $n \in \mathbb{N}$ such that $nf(x) \in \mathbb{Z}[x]$ and let $d = I(nf)$.]

(c) For all $f(x), g(x) \in \mathbb{Z}[x]$ and prime $p \in \mathbb{Z}$ prove that $p | f(x)g(x)$ implies that $p | f(x)$ or $p | g(x)$. Use this to conclude that $I(f) = I(g) = 1$ implies $I(fg) = 1$. [Hint: Consider the ring homomorphism $\mathbb{Z}[x] \to \mathbb{Z}/p\mathbb{Z}[x]$ defined by reducing each coefficient mod p.]

(d) If $f(x), g(x) \in \mathbb{Q}[x]$ are **monic** with $f(x)g(x) \in \mathbb{Z}[x]$, prove that $f(x), g(x) \in \mathbb{Z}[x]$. [Hint: From (b) we have $k, \ell \in \mathbb{N}$ with $kf(x), \ell g(x) \in \mathbb{Z}[x]$ and $I(kf) = I(\ell g) = 1$. Now use (c) to show that $k\ell = 1$.]

(e) Use the previous results to prove that $\mathbb{Z}[x]$ is a UFD. [Hint: It suffices to prove that every irreducible element of $\mathbb{Z}[x]$ is prime. There are two cases: (1) irreducible constants $p \in \mathbb{Z}$ and (2) non-constant irreducible polynomials $p(x) \in \mathbb{Z}[x]$.]

[136] And it is deeper than it looks. The general construction of fractions is called "localization". It somehow corresponds to "zooming in" on a point of an algebraic variety.

[137] The letter I stands for *Inhalt* (German for "content").

18.D Waring's Theorem on Symmetric Polynomials

Given a ring R and a set of "independent variables" $\mathbf{x} = \{x_1, \ldots, x_n\}$ we define *multivariate polynomials* by induction:

$$R[\mathbf{x}] = R[x_1, \ldots, x_n] \coloneqq R[x_1, \ldots, x_{n-1}][x_n] = \left\{ f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}^n} a_{\mathbf{k}} \mathbf{x}^{\mathbf{k}} : a_{\mathbf{k}} \in R \right\}.$$

To save space we use the notations $\mathbf{k} = (k_1, \ldots, k_n) \in \mathbb{N}^k$ and $\mathbf{x}^{\mathbf{k}} = x_1^{k_1} \cdots x_n^{k_n}$. We assume that all but finitely many of the coefficients $a_{\mathbf{k}} \in R$ are zero.

(a) We say that a polynomial $f(\mathbf{x}) = R[\mathbf{x}]$ is *symmetric* if for all $\sigma \in S_n$ we have

$$f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = f(x_1, \ldots, x_n).$$

Observe that the symmetric polynomials are a subring of $R[\mathbf{x}]$.

(b) *Waring's Theorem.* Recall the definition of the *elementary symmetric polynomials*:

$$e_k(x_1, \ldots, x_n) \coloneqq \sum_{1 \le i_1 < \cdots < i_k \le n} x_{i_1} \cdots x_{i_k}.$$

For convenience, let's define $\mathbf{e}^{\mathbf{k}} \coloneqq e_1^{k_1} \cdots e_n^{k_n}$. For any symmetric polynomial $f(\mathbf{x}) = \sum_{\mathbf{k}} a_{\mathbf{k}} \mathbf{x}^{\mathbf{k}} \in R[\mathbf{x}]$, prove that there exist some $b_{\mathbf{k}} \in R$ such that $f(x) = \sum_{\mathbf{k}} b_{\mathbf{k}} \mathbf{e}^{\mathbf{k}}$. [Hint: Order the degree vectors $\mathbf{k} \in \mathbb{N}^n$ by "dictionary order" and let $a_{\mathbf{k}} \mathbf{x}^{\mathbf{k}}$ be the "leading term". By symmetry of f we must have $k_1 \ge k_2 \ge \cdots \ge k_n$. Show that there exists $\mathbf{k}' \in \mathbb{N}^k$ so that $a_{\mathbf{k}} \mathbf{e}^{\mathbf{k}'}$ has the same leading term, hence $f(\mathbf{x}) - a_{\mathbf{k}} \mathbf{e}^{\mathbf{k}'}$ is a symmetric polynomial of "smaller degree".]

(c) *Important Corollary.* Suppose that a polynomial $f(x) \in R[x]$ of degree n splits in some ring extension $E \supseteq R$. That is, suppose that we have

$$f(x) = x^n - e_1 x^{n-1} + e_2 x^{n-2} - \cdots + (-1)^n e_n = (x - \alpha_1) \cdots (x - \alpha_n) \in E[x].$$

Prove that any "symmetric expression of the roots" is in R.

(d) *Application: Discriminant of a Cubic.* Let $f(x) = x^3 + ax^2 + bx + c \in R[x]$ and let $E \supseteq R$ be a ring extension such that

$$x^3 + ax^2 + bx + c = (x - \alpha)(x - \beta)(x - \gamma) \in E[x].$$

From part (c) we know that the following element of E (called the *discriminant* of f) is actually in R:

$$\mathrm{Disc}(f) \coloneqq (\alpha - \beta)^2 (\alpha - \gamma)^2 (\beta - \gamma)^2.$$

Use the algorithm from part (b) to express $\mathrm{Disc}(f)$ as a specific polynomial in the coefficients. [I'll get you started: Note that $\mathrm{Disc}(f) = (\alpha^4 \beta^2 + \text{lower terms})$ and $a^2 b^2 = (\alpha^4 \beta^2 + \text{lower terms})$. Now find the leading term of $\mathrm{Disc}(f) - a^2 b^2$.]

Week 19

19.1 Irreducible Polynomials

Last week we developed the following tool for studying field extensions:

> Let $\alpha \in \mathbb{E} \supseteq \mathbb{F}$ be an element of a field extension and consider the intermediate field $\mathbb{E} \supseteq \mathbb{F}(\alpha) \supseteq \mathbb{F}$. If $f(x) \in \mathbb{F}[x]$ is an **irreducible** polynomial with $f(\alpha) = 0$ then $1, \alpha, \alpha^2, \ldots, \alpha^{\deg(f)-1}$ is a basis for $\mathbb{F}(\alpha)$ as a vector space over $\mathbb{F}$, hence
>
> $$[\mathbb{F}(\alpha)/\mathbb{F}] = \deg(f).$$

But this tool is only useful if we have some way to prove that a given polynomial is irreducible. Here are a couple of basic tricks.

Low-Degree Trick. Let $f(x) \in \mathbb{F}[x]$ have degree 2 or 3. Then

$$f(x) \in \mathbb{F}[x] \text{ is reducible} \quad \Longleftrightarrow \quad f(x) \text{ has a root in } \mathbb{F}.$$

Proof. Since $\mathbb{F}$ is a field we know from Descartes' Theorem that

$$f(x) \in \mathbb{F}[x] \text{ has a factor of degree 1} \quad \Longleftrightarrow \quad f(x) \text{ has a root in } \mathbb{F}.$$

Indeed, if $\alpha \in \mathbb{F}$ is a root then we have $f(x) = (x-\alpha)g(x)$ for some $g(x) \in \mathbb{F}[x]$. Conversely, if $f(x) = (ax+b)g(x)$ for some $a, b \in \mathbb{F}$ with $a \neq 0$ then $-b/a \in \mathbb{F}$ is a root. Finally, if $f(x) \in \mathbb{F}[x]$ has degree 2 or 3 then we observe that $f(x)$ is reducible if and only if $f(x)$ has a factor of degree 1. □

Here is an example to show that the trick does not work for polynomials of degree four.

Example: Leibniz' Mistake. One of the first problems of Calculus was to compute the antiderivative for any given rational function $f(x)/g(x)$ with $f(x), g(x) \in \mathbb{R}[x]$. By 1675, Leibniz knew that

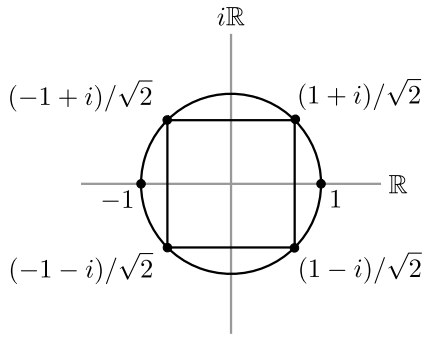
$$\int x^n dx = \frac{x^{n+1}}{n+1} \text{ (if } n \neq 1), \quad \int \frac{dx}{x} = \log(x), \quad \int \frac{dx}{x^2+1} = \arctan(x).$$

He also knew that if the denominator $g(x)$ can be factored into polynomials of degree 1 and 2, then the function $f(x)/g(x)$ can be expanded by partial fractions and hence the antiderivative can be computed from the above three formulas.

Today we know that every irreducible polynomial in $\mathbb{R}[x]$ has degree 1 or 2. (This is one way to state the Fundamental Theorem of Algebra.) But this fact is certainly not obvious. Indeed, Leibniz made a famous mistake in 1702 when he claimed that polynomials of the form $x^4 + a^4$ with $a \in \mathbb{R}$ and $a \neq 0$ are irreducible over $\mathbb{R}$. Here is his exact quote:

> *Therefore,* $\int \frac{dx}{x^4+a^4}$ *cannot be reduced to the squaring of the circle or the hyperbola by our analysis above, but founds a new kind of its own.*[138]

Leibniz' problem was that he didn't have a good understanding of the complex 4th roots of -1. Today we know that these roots are the vertices of a square in the complex plane:



Then by grouping the complex roots of $x^4 + a^4$ into conjugate pairs we obtain

$$\begin{aligned} x^4 + a^4 &= \left(x - \frac{a(1+i)}{\sqrt{2}}\right)\left(x - \frac{a(1-i)}{\sqrt{2}}\right)\left(x - \frac{a(-1+i)}{\sqrt{2}}\right)\left(x - \frac{a(-1-i)}{\sqrt{2}}\right) \\ &= \left(x^2 - a\sqrt{2}x + a^2\right)\left(x^2 + a\sqrt{2}x + a^2\right), \end{aligned}$$

and it follows from this that the antiderivative of $1/(x^4+a^4)$ **can** be expressed in terms of log and arctan (but I won't write the formula because it's too terrible). However, for our purposes, the main point of this example is that the polynomial $x^4 + a^4 \in \mathbb{R}[x]$ (for $a \neq 0$) is **reducible over** $\mathbb{R}$ but has **no roots in** $\mathbb{R}$. ///

To apply the Low-Degree Trick we still need some method to prove that a polynomial has no roots in a certain field. The following trick works when we are looking for roots in the field of fractions of a UFD.

[138] See Tignol, page 75. By "squaring of the circle" Leibniz means $\int \frac{dx}{x^2+1} = \arctan(x)$ and "squaring of the hyperbola" he means $\int \frac{dx}{x} = \log(x)$.

The Rational Root Test. Let R be a UFD (for example, $\mathbb{Z}$) and consider a polynomial

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n \in R[x].$$

Since R is a UFD we can write any fraction $p/q \in \mathrm{Frac}(R)$ in "lowest terms", i.e., with $\gcd(p, q) = 1$. If $f(p/q) = 0$ then we must have

$$p|a_0 \quad \text{and} \quad q|a_n.$$

And these restrictions give us a finite list of possible roots p/q that we can check by hand.

Proof. Multiplying both sides of the equation $f(p/q) = 0$ by q^n gives

$$\begin{aligned} a_0 + a_1(p/q) + \cdots + a_{n-1}(p/q)^{n-1} + a_n(p/q)^n &= 0 \\ a_0 q^n + a_1 p q^{n-1} + \cdots + a_{n-1} p^{n-1} q + a_n p^n &= 0. \end{aligned}$$

Pulling $a_0 q^n$ to one side gives

$$a_0 q^n = -p(a_1 q^{n-1} + \cdots + a_{n-1} p^{n-2} q + a_n p^{n-1}) \quad \Longrightarrow \quad p|a_0 q^n,$$

which implies that $p|a_0$ because $\gcd(p, q) = 1$. Similarly, pulling $a_n p^n$ to one side gives

$$a_n p^n = -q(a_0 q^{n-1} + a_1 p q^{n-2} + \cdots + a_{n-1} p^{n-1}) \quad \Longrightarrow \quad q|a_n p^n,$$

which implies that $q|a_n$ because $\gcd(p, q) = 1$. □

[**Remark:** We just used the fact that $a|bc$ and $\gcd(a, b) = 1$ imply $a|c$. In a PID we can prove this by writing $ax + by = 1$ and then multiplying both sides by c. In a general UFD these x, y might not exist, but we can still prove the result by comparing prime factorizations. The details are not important.]

19.2 Gauss and Cyclotomy

The tricks from the previous lecture are surprisingly useful. Here is an example that fills in a gap from our discussion in the introduction.

Example: The Splitting Field of $x^3 - 2$. The polynomial $x^3 - 2 \in \mathbb{Q}[x]$ has three distinct complex roots. To be specific, if $\alpha := \sqrt[3]{2} \in \mathbb{R}$ and $\omega := e^{2\pi i/3} \in \mathbb{C}$ then we can write

$$x^3 - 2 = (x - \alpha)(x - \omega\alpha)(x - \omega^2\alpha) \in \mathbb{C}[x].$$

Let $\mathbb{E} := \mathbb{Q}(\alpha, \omega\alpha, \omega^2\alpha) \subseteq \mathbb{C}$ be the splitting field. In the introduction I claimed that $[\mathbb{E}/\mathbb{Q}] = 6$ with basis $\{1, \alpha, \alpha^2, \omega, \omega\alpha, \omega\alpha^2\}$ but we were not able to prove this at the time. Now we can.

Proof. First observe that $\mathbb{Q}(\alpha, \omega\alpha, \omega^2\alpha) = \mathbb{Q}(\alpha, \omega)$ because $\{\alpha, \omega\alpha, \omega^2\alpha\}$ can be obtained from $\{\alpha, \omega\}$ through field operations and, conversely, $\{\alpha, \omega\}$ can be obtained from $\{\alpha, \omega\alpha, \omega^2\alpha\}$ through field operations. Consider the following chain of field extensions:

$$\mathbb{Q} \subseteq \mathbb{Q}(\alpha) \subseteq \mathbb{Q}(\alpha)(\omega) = \mathbb{E}.$$

Our goal is to compute a vector space basis for each extension and then combine them using Dedekind's Tower Law. We have already seen that $\mathbb{Q}(\alpha)/\mathbb{Q}$ has basis $\{1, \alpha, \alpha^2\}$ but let me prove this again quickly. Note that α is a root of $f(x) \coloneqq x^3 - 2 \in \mathbb{Q}$. If $f(p/q) = 0$ is a rational root in lowest terms then the Rational Root Trick says that $p|2$ and $q|1$. But we can check by hand that ± 2 are **not** roots of $f(x)$. Since $f(x)$ has degree 3 and no rational roots we conclude that $f(x) = m_{\alpha/\mathbb{Q}}(x)$ is the minimal polynomial for α over $\mathbb{Q}$, and since $\deg(m_{\alpha/\mathbb{Q}}) = 3$ we conclude that $\{1, \alpha, \alpha^2\}$ is a basis for $\mathbb{Q}(\alpha)/\mathbb{Q}$.

To find a basis for $\mathbb{Q}(\alpha)(\omega)/\mathbb{Q}(\alpha)$ we need to compute the minimal polynomial:

$$m_{\omega/\mathbb{Q}(\alpha)}(x) \in \mathbb{Q}(\alpha)[x].$$

First of all, note that ω is a root of $x^3 - 1 \in \mathbb{Q}[x] \subseteq \mathbb{Q}(\alpha)[x]$. But this polynomial is **not irreducible** because

$$x^3 - 1 = (x - 1)(x^2 + x + 1) \in \mathbb{Q}[x] \subseteq \mathbb{Q}(\alpha)[x].$$

Furthermore, since

$$(\omega - 1)(\omega^2 + \omega + 1) = \omega^3 - 1 = 0$$

we must have $\omega^2 + \omega + 1 = 0$. I claim that $g(x) \coloneqq x^2 + x + 1 \in \mathbb{Q}(\alpha)[x]$ is irreducible and hence is the minimal polynomial for ω over $\mathbb{Q}(\alpha)$. Indeed, since $g(x)$ has degree 2 we only need to check that it has **no roots in the field** $\mathbb{Q}(\alpha)$. But we know that $g(x)$ has two **non-real roots** $\omega, \omega^2 \in \mathbb{C} - \mathbb{R}$ and since $\alpha \in \mathbb{R}$ we know that $\mathbb{Q}(\alpha)$ is contained in $\mathbb{R}$, hence

$$m_{\omega/\mathbb{Q}(\alpha)}(x) = x^2 + x + 1.$$

Then since $\deg(m_{\omega/\mathbb{Q}(\alpha)}) = 2$ we conclude that $\{1, \omega\}$ is a basis for $\mathbb{E} = \mathbb{Q}(\alpha)(\omega)/\mathbb{Q}(\alpha)$. Finally, by applying Dedekind's Tower Law we obtain the basis

$$\{1, \alpha, \alpha^2\} \cdot \{1, \omega\} = \{1 \cdot 1, \alpha \cdot 1, \alpha^2 \cdot 1, 1 \cdot \omega, \alpha \cdot \omega, \alpha^2 \cdot \omega\} = \{1, \alpha, \alpha^2, \omega, \omega\alpha, \omega\alpha^2\}$$

for the splitting field $\mathbb{E}/\mathbb{Q}$, and it follows that $[\mathbb{E}/\mathbb{Q}] = 6$. □

Remarks:

- After seeing that $\mathbb{Q}(\alpha, \omega\alpha, \omega^2\alpha) = \mathbb{Q}(\alpha, \omega)$ you might wonder if the splitting field can be generated by a single element:

$$\mathbb{Q}(\alpha, \omega) = \mathbb{Q}(\gamma) \text{ for some } \gamma \in \mathbb{Q}(\alpha, \omega)?$$

If this is possible then we will call γ a *primitive element* for the field extension.[139] Later we will prove that any finite dimensional extension over $\mathbb{Q}$ has a primitive element (in fact, infinitely many). However, it is not easy to find one by hand. For this example I used my computer to verify that $\gamma := \alpha + \omega$ is a primitive element with minimal polynomial[140]

$$m_{\gamma/\mathbb{Q}}(x) = x^6 + 3x^5 + 6x^4 + 3x^3 + 0x^2 + 9x + 9 \in \mathbb{Q}[x].$$

Note this polynomial has degree 6 as expected.

- Suppose that we have field extensions $\mathbb{E} \supseteq \mathbb{K} \supseteq \mathbb{F}$ and an element $\alpha \in \mathbb{E}$. If the minimal polynomial $m_{\alpha/\mathbb{K}}(x) \in \mathbb{K}[x]$ has coefficients in $\mathbb{F}$ then it necessarily follows that

$$m_{\alpha/\mathbb{K}}(x) = m_{\alpha/\mathbb{F}}(x) \in \mathbb{F}[x].$$

Proof. Any polynomial that is irreducible over $\mathbb{K}$ is still irreducible over $\mathbb{F}$. Then the result follows since $m_{\alpha/\mathbb{K}}(x) \in \mathbb{F}[x]$ is monic, irreducible and has α as a root. □

Thus from the above example we have

$$m_{\omega/\mathbb{Q}(\alpha)}(x) = m_{\omega/\mathbb{Q}}(x) = x^2 + x + 1.$$

///

We have seen that x^2+x+1 is the minimal polynomial over $\mathbb{Q}$ for the primitive third roots of unity. More generally, I claim that the following definition gives the minimal polynomial over $\mathbb{Q}$ for any primitive n-th root of unity.

Definition of Cyclotomic Polynomials. For any integer $n \geq 1$ we define

$$\Phi_n(x) := \prod_{\substack{0 \leq k < n \\ \gcd(k,n)=1}} (x - e^{2\pi i k/n}) \in \mathbb{C}[x].$$

///

At first it seems that cyclotomic polynomials have complex coefficients. However, on the next homework you will prove by induction that $\Phi_n(x) \in \mathbb{Z}[x] \subseteq \mathbb{Q}[x]$ for all $n \geq 1$. We have already seen the first few examples:

$$\Phi_1(x) = x - 1,$$
$$\Phi_2(x) = x + 1,$$

[139] Another name for a primitive element is a *Galois resolvent*, hence the letter γ.

[140] In particular, this polynomial is irreducible over $\mathbb{Q}$. But I would never know that if you showed it to me out of context.

$$\Phi_3(x) = x^2 + x + 1.$$

For the next case, observe that the primitive 4th roots of unity are $\{\pm i\}$, hence

$$\Phi_4(x) = (x - i)(x + i) = x^2 + 1.$$

So far it is clear that each of these polynomials is irreducible over $\mathbb{Q}$. But you will show that

$$\Phi_5(x) = x^4 + x^3 + x^2 + x + 1,$$

and more generally that

$$\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x^2 + x + 1 \quad \text{for any prime } p.$$

It is not clear why these polynomials should be irreducible over $\mathbb{Q}$. Gauss proved in the *Disquisitiones* that $\Phi_p(x) \in \mathbb{Z}[x]$ is irreducible for any prime p. The typical textbook proof of this uses a clever trick called "Eisenstein's Criterion", which was communicated in an 1850 letter from Gotthold Eisenstein to Gauss. It is also true, but quite tricky to prove (see Exercise 22.C), that $\Phi_n(x) \in \mathbb{Q}[x]$ is irreducible for any n. Then it follows that $\Phi_n(x) \in \mathbb{Q}[x]$ is the minimal polynomial for any primitive root of n over $\mathbb{Q}$, and hence the dimension of the *cyclotomic field* $\mathbb{Q}(e^{2\pi i/n})/\mathbb{Q}$ is equal to Euler's totient function:

$$[\mathbb{Q}(e^{2\pi i/n})/\mathbb{Q}] = \phi(n) = \#\{0 \le k < n : \gcd(k, n) = 1\}.$$

Any why did Gauss care about this? His original goal was to investigate whether the n-th roots of unity can be expressed in terms of square roots.

Definition of Constructible Numbers. We say that a complex number $\alpha \in \mathbb{C}$ is *constructible* if it can be obtained from $\mathbb{Q}$ by solving a sequence of quadratic equations, i.e., if there exists a chain of fields

$$\alpha \in \mathbb{F}_k \supseteq \mathbb{F}_{k-1} \supseteq \cdots \supseteq \mathbb{F}_1 \supseteq \mathbb{F}_0 = \mathbb{Q}$$

satisfying $[\mathbb{F}_{i+1}/\mathbb{F}_i] = 2$ for all i. ///

The motivation for the word "constructible" comes from Euclidean geometry. Suppose that we start with the points $(0, 0)$ and $(1, 0)$ in the Cartesian plane $\mathbb{R}^2$. From these two points we are allowed to construct new points via Euclid's Postulates:

- We are allowed to draw the straight line through any two points.
- Given points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ we are allowed to draw the circle through $\mathbf{y}$ with center at $\mathbf{x}$.
- We are allowed to draw the points of intersection for any constructed lines and circles.

One can check that the points of intersection of any two lines and circles can always be computed by a quadratic equation,[141] hence any point $(\alpha, \beta) \in \mathbb{R}^2$ that is constructible in the geometric sense will have coordinates $\alpha, \beta \in \mathbb{R}$ that are constructible in the algebraic sense.[142]

The young Gauss applied this reasoning to the construction of regular polygons. He completed the *Disquisitiones Arithmeticae* in 1798, at the age of 21. The final chapter of this work contains a study of "cyclotomy". We can summarize the main points as follows:

the regular n-gon is constructible
$\iff$ the point $(\cos(2\pi/n), \sin(2\pi/n)) \in \mathbb{R}^2$ is constructible
$\iff$ the numbers $\cos(2\pi/n), \sin(2\pi/n) \in \mathbb{R}$ are constructible
$\iff$ the number $e^{2\pi i/n} \in \mathbb{C}$ is constructible
$\iff$ the dimension $[\mathbb{Q}(e^{2\pi i/n})/\mathbb{Q}]$ is a power of 2
$\iff$ $e^{2\pi i/n} \in \mathbb{C}$ has minimal polynomial with degree a power of 2
$\iff$ the cyclotomic polynomial $\Phi_n(x)$ has degree a power of 2
$\iff$ Euler's totient $\phi(n)$ is a power of 2.

Some of these implications were filled in by Pierre Wantzel in 1837, when he was 23 years old.[143] Hence this result is sometimes called the **Gauss-Wantzel Theorem**.

For example, by observing that $\phi(7) = 6$ is not a power of 2 we can explain why the ancient Greeks were never able to construct a regular heptagon with straightedge and compass. (You will give a more elementary proof of this fact on the next homework.) However, the more surprising result is the existence of constructible polygons that the ancient Greeks missed. By observing that $\phi(17) = 16$ is a power of 2, Gauss was able to prove (indirectly) that

the regular 17*-gon is constructible with straightedge and compass!*

Exercises

19.A Computing Minimal Polynomials

Define $\alpha \coloneqq \sqrt[3]{2} \in \mathbb{R}$ and $\omega \coloneqq e^{2\pi i/3} \in \mathbb{C}$.

(a) Prove that $x^3 - 2$ is the minimal polynomial for α over $\mathbb{Q}(\omega)$.

(b) Prove that $x^2 + x + 1$ is the minimal polynomial for ω over $\mathbb{Q}(\alpha\omega)$.

[141] The hardest case is the intersection of two circles.
[142] The converse is also true but I feel no need to discuss this.
[143] Abel died in 1829 at age 26 and Galois died in 1832 at age 20. For some reason there were a lot of precocious mathematicians in the early 1800s.

(c) Prove that $x^2 + (\alpha\omega)x + (\alpha\omega)^2$ is the minimal polynomial for α over $\mathbb{Q}(\alpha\omega)$.

[Hint: Consider any $\beta \in \mathbb{E} \supseteq \mathbb{F}$ and let $f(x) \in \mathbb{F}[x]$ be a polynomial satisfying $\deg(f) = [\mathbb{E}/\mathbb{F}]$. Suppose also that $f(x)$ is monic and satisfies $f(\beta) = 0$, hence $m_{\beta/\mathbb{F}}(x)|f(x)$. Then since $m_{\beta/\mathbb{F}}(x)$ and $f(x)$ are monic of the same degree we conclude that $m_{\beta/\mathbb{F}}(x) = f(x)$.]

19.B Cyclotomic Polynomials

Fix an integer n and consider the polynomial $x^n - 1 \in \mathbb{Z}[x]$.

(a) Factor $x^n - 1$ into irreducible polynomials over $\mathbb{C}$. [Hint: Let $\omega \coloneqq e^{2\pi/n}$.]

(b) Factor $x^n - 1$ into irreducible polynomials over $\mathbb{R}$. [Hint: For all integers $k \in \mathbb{Z}$ we have $\omega^k + \omega^{-k} = 2\cos(2\pi k/n)$.]

(c) We define the n-th *cyclotomic polynomial* $\Phi_n(x) \in \mathbb{C}[x]$ as follows:

$$\Phi_n(x) \coloneqq \prod_{\omega \in \Omega'_n} (x - \omega) \quad \text{where} \quad \Omega'_n \coloneqq \{e^{2\pi i k/n} : 0 \leq k < n, \gcd(k, n) = 1\}.$$

Prove that

$$x^n - 1 = \prod_{d|n} \Phi_d(x) = \prod_{d|n} \Phi_{n/d}(x).$$

[Hint: The elements of Ω'_n are called the *primitive* nth roots of unity. Prove that the set of **all** nth roots of unity can be expressed as a disjoint union $\coprod_{d|n} \Omega'_d$.]

(d) Use part (c) and induction to prove that actually $\Phi_n(x) \in \mathbb{Z}[x]$. [Hint: For any $f(x), g(x) \in \mathbb{Z}[x]$ with $g(x)$ monic there exist **unique** polynomials $q(x), r(x) \in \mathbb{Z}[x]$ such that $f(x) = q(x)g(x) + r(x)$ and $\deg(r) < \deg(g)$.]

[**Remark:** We will prove in Exercise 22.C that in fact $\Phi_n(x)$ is **irreducible** over $\mathbb{Q}$, hence is the minimal polynomial over $\mathbb{Q}$ for any primitive n-th root of unity ω. It follows that the cyclotomic field $\mathbb{Q}(\omega)/\mathbb{Q}$ has dimension equal to Euler's totient $\phi(n)$.]

19.C Quadratic Field Extensions, Part III

Prove that $[\mathbb{E}/\mathbb{F}] = 2$ implies $\mathbb{E} = \mathbb{F}(\iota)$ for some $\iota \in \mathbb{E} - \mathbb{F}$ with $\iota^2 \in \mathbb{F}$. [Hint: For any $\alpha \in \mathbb{E} - \mathbb{F}$ note that the set $1, \alpha, \alpha^2$ is linearly dependent, hence we have $f(\alpha) = 0$ for some polynomial $f(x) \in \mathbb{F}[x]$ of degree 2. Let $\beta \in \mathbb{E}$ be the other root of $f(x)$ and define $\iota \coloneqq \alpha - \beta \in \mathbb{E}$.]

19.D Impossible Constructions

We say that a number $\alpha \in \mathbb{R}$ is *constructible over* $\mathbb{Q}$ if there exists a chain of field extensions

$$\alpha \in \mathbb{F}_k \supseteq \mathbb{F}_{k-1} \supseteq \cdots \supseteq \mathbb{F}_1 \supseteq \mathbb{F}_0 = \mathbb{Q}$$

such that $[\mathbb{F}_{i+1}/\mathbb{F}_i] = 2$ for all i. [Reason: A point of $\mathbb{R}^2$ is "constructible with straightedge and compass" if and only if both of its coordinates are constructible in the above sense.]

(a) Let $f(x) \in \mathbb{Q}[x]$ be any polynomial of degree 3. Prove that

$$f \text{ has a constructible root } \alpha \in \mathbb{R} \quad \Longrightarrow \quad f \text{ has a root in } \mathbb{Q}.$$

[Hint: You proved the induction step on the previous homework.]

(b) Prove that the real numbers $\sqrt[3]{2}$, $\cos(2\pi/18)$ and $\cos(2\pi/7)$ are not constructible. It follows from this that the classical problems of "doubling the cube", "trisecting the angle", and "constructing the regular heptagon" are impossible. [Hint: Show that each is a root of some irreducible polynomial $f(x) \in \mathbb{Q}[x]$ of degree 3.]

Week 20

20.1 Fields of Size Four and Eight

This week we will apply our knowledge of irreducible polynomials to the construction of finite fields. We already know that finite fields exist since $\mathbb{F}_p \coloneqq \mathbb{Z}/p\mathbb{Z}$ is a field for any prime $p \in \mathbb{Z}$. But are there any other finite fields?

Suppose that $\mathbb{E}$ is a finite field. This implies that $\mathbb{E}$ has characteristic $p > 0$ since otherwise the prime subfield would be $\mathbb{Q}$, which is infinite. So let $\mathbb{F}_p \subseteq \mathbb{E}$ be the prime subfield and consider the vector space $\mathbb{E}/\mathbb{F}_p$. Since $\mathbb{E}$ is **finite** we know that this vector space is **finite-dimensional**, say $[\mathbb{E}/\mathbb{F}_p] = k$. In this case I claim that $\#\mathbb{E} = p^k$.

Proof. Let $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{E}$ be a basis for $\mathbb{E}$ as a vector space over $\mathbb{F}_p$. By definition, every element of $\mathbb{E}$ can be expressed uniquely in the form

$$a_1\alpha_1 + a_2\alpha_2 + \cdots + a_k\alpha_k \text{ for some } a_1, a_2, \ldots, a_k \in \mathbb{F}_p.$$

Then since there are p ways to choose each coefficient we conclude that

$$\#E = (\#\text{ choices for } a_1)(\#\text{ choices for } a_2)\cdots(\#\text{ choices for } a_k) = p^k.$$

□

But we still have not seen any fields of size p^k with $k \geq 2$. Here is our first example.

Example: A Field of Size Four. Consider the polynomial $x^2+x+1 \in \mathbb{F}_2[x]$ with coefficients in the field of two elements $\mathbb{F}_2 = \{0, 1\}$. It is easy to see that this polynomial is irreducible over $\mathbb{F}_2$ because it has no roots in $\mathbb{F}_2$:

x	0	1
$x^2 + x + 1$	1	1

Since $\text{char}(\mathbb{F}_2) = 2 \neq 0$ the Fundamental Theorem of Algebra doesn't tell us anything about the existence of roots, so we have to apply Kronecker's

Theorem. Specifically, since $\mathbb{F}_2[x]$ is a PID it follows that the ideal $\langle x^2+x+1\rangle \subseteq \mathbb{F}_2[x]$ is maximal, hence we obtain a field:

$$\mathbb{E} \coloneqq \frac{\mathbb{F}_2[x]}{\langle x^2+x+1\rangle}.$$

If we identify $\mathbb{F}_2$ with the subfield $\{a+\langle x^2+x+1\rangle : a \in \mathbb{F}_2\} \subseteq \mathbb{E}$ then we can think of $\mathbb{E} \supseteq \mathbb{F}_2$ as a field extension which contains an element $\alpha \in \mathbb{E}$ satisfying

$$\alpha^2+\alpha+1=0.^{144}$$

In fact, since $x^2+x+1 \in \mathbb{F}_2[x]$ is monic, irreducible and has $\alpha \in \mathbb{E}$ as a root, we conclude that it is the minimal polynomial for α over $\mathbb{F}_2$ and it follows from this that $\mathbb{E}=\mathbb{F}_2(\alpha)$:

$$\mathbb{E} = \frac{\mathbb{F}_2[x]}{\langle x^2+x+1\rangle} = \frac{\mathbb{F}_2[x]}{\langle m_{\alpha/\mathbb{F}_2}(x)\rangle} \cong \mathbb{F}_2(\alpha) \subseteq \mathbb{E}.$$

Furthermore, since the minimal polynomial has degree 2 we conclude that $\{1,\alpha\}$ is a basis for $\mathbb{E}$ over $\mathbb{F}_2$ and it follows that

$$\mathbb{E} = \{0+0\alpha, 1+0\alpha, 0+1\alpha, 1+1\alpha\} = \{0,1,\alpha,1+\alpha\}.$$

Thus we have constructed a field of size four. The addition table is just inherited from the vector space structure of $\mathbb{E}/\mathbb{F}_2$:

$+$	0	1	α	$1+\alpha$
0	0	1	α	$1+\alpha$
1	1	0	$1+\alpha$	α
α	α	$1+\alpha$	0	1
$1+\alpha$	$1+\alpha$	α	1	0

But the multiplication table is more interesting because it uses the polynomial relation

$$\alpha^2+\alpha+1=0 \quad\Longrightarrow\quad \alpha^2=-1-\alpha=1+\alpha.$$

For example, we have $(1+\alpha)^2=1^2+2\alpha+\alpha^2=1+0+(1+\alpha)=\alpha$. Here is the full table:

$\times$	0	1	α	$1+\alpha$
0	0	0	0	0
1	0	1	α	$1+\alpha$
α	0	α	$1+\alpha$	1
$1+\alpha$	0	$1+\alpha$	1	α

In fact one could use this table as the **definition of multiplication** in $\mathbb{E}$ and then check by hand that all of the field axioms are satisfied. However, that would leave the existence of $\mathbb{E}$ completely unexplained. ///

[144] Technically, $\alpha = x+\langle x^2+x+a\rangle$ is the coset generated by x but from this point on we will just call it α. Prior to Kronecker's Theorem the "imaginary" roots of polynomials over finite fields were called "Galois imaginaries" and their nature was somewhat mysterious.

The construction of the field $\mathbb{E} = \{0, 1, \alpha, 1+\alpha\}$ above might have seemed rather arbitrary, but I claim that there were no other options.

Theorem (There is Only One Field of Size Four). If $\mathbb{E}'$ is any field of size four then we have a ring isomorphism $\mathbb{E}' \cong \mathbb{E}$.

Proof. Let $\#\mathbb{E}' = 4$. Then from previous remarks we know that $\mathbb{E}'$ is a 2-dimensional vector space over $\mathbb{F}_2$. Extend the set $\{1\}$ to a basis $\{1, \gamma\}$. Then by definition we must have

$$\mathbb{E}' = \{0, 1, \gamma, 1+\gamma\}.$$

Clearly we have a vector space isomorphism identifying $\alpha \leftrightarrow \gamma$. But does this isomorphism also preserve multiplication? For this we need to prove that $\gamma^2 = 1+\gamma$. So consider the element $\gamma^2 \in \{0, 1, \gamma, 1+\gamma\}$. Since $\gamma \neq 0$ in a field we must have $\gamma^2 \neq 0$. Then since $\gamma \notin \{0, 1\}$ in a field we must have $\gamma^2 \neq \gamma$. Finally, assume for contradiction that we have $\gamma^2 = 1$, so that $1 - \gamma^2 = 0$. But then we have

$$0 = 1 - \gamma^2 = (1-\gamma)(1+\gamma) = (1+\gamma)(1+\gamma) = (1+\gamma)^2,$$

which contradicts the fact that $1+\gamma \neq 0$. By process of elimination we conclude that $\gamma^2 = 1+\gamma$ and hence $\mathbb{E}' \cong \mathbb{E}$ as rings. □

The next-smallest non-trivial power of a prime is $2^3 = 8$.

Example: Two Fields of Size Eight? Based on the previous example, we will be able to construct a field of size 8 if we can find an irreducible polynomial in $\mathbb{F}_2[x]$ of degree 3. In fact, there are two such polynomials! Indeed, the polynomials $x^3 + x^2 + 1$ and $x^3 + x + 1$ are both irreducible over $\mathbb{F}_2$ since they each have degree 3 and no roots in $\mathbb{F}_2$:

x	0	1
$x^3 + x^2 + 1$	1	1
$x^3 + x + 1$	1	1

This guarantees that the following two vector spaces over $\mathbb{F}_2$ are fields:

$$\mathbb{E} \coloneqq \{a + b\alpha + c\alpha : a, b, c \in \mathbb{F}_2, \alpha^3 + \alpha^2 + 1 = 0\},$$
$$\mathbb{E}' \coloneqq \{a + b\beta + c\beta^2 : a, b, c \in \mathbb{F}_2, \beta^3 + \beta + 1 = 0\}.$$

For your information, here is the multiplication table of the field $\mathbb{E}$:

$\times$	0	1	α	$1+\alpha$	α^2	$1+\alpha^2$	$\alpha+\alpha^2$	$1+\alpha+\alpha^2$
0	0	0	0	0	0	0	0	0
1	0	1	α	$1+\alpha$	α^2	$1+\alpha^2$	$\alpha+\alpha^2$	$1+\alpha+\alpha^2$
α	0	α	α^2	$\alpha+\alpha^2$	$1+\alpha^2$	$1+\alpha+\alpha^2$	1	$1+\alpha$
$1+\alpha$	0	$1+\alpha$	$\alpha+\alpha^2$	$1+\alpha^2$	1	α	$1+\alpha+\alpha^2$	α^2
α^2	0	α^2	$1+\alpha^2$	1	$1+\alpha+\alpha^2$	$1+\alpha$	α	$\alpha+\alpha^2$
$1+\alpha^2$	0	$1+\alpha^2$	$1+\alpha+\alpha^2$	α	$1+\alpha$	$\alpha+\alpha^2$	α^2	1
$\alpha+\alpha^2$	0	$\alpha+\alpha^2$	1	$1+\alpha+\alpha^2$	α	α^2	$1+\alpha$	$1+\alpha^2$
$1+\alpha+\alpha^2$	0	$1+\alpha+\alpha^2$	$1+\alpha$	α^2	$\alpha+\alpha^2$	1	$1+\alpha^2$	α

And here is the multiplication table of $\mathbb{E}'$:

$\times$	0	1	β	$1+\beta$	β^2	$1+\beta^2$	$\beta+\beta^2$	$1+\beta+\beta^2$
0	0	0	0	0	0	0	0	0
1	0	1	β	$1+\beta$	β^2	$1+\beta^2$	$\beta+\beta^2$	$1+\beta+\beta^2$
β	0	β	β^2	$\beta+\beta^2$	$1+\beta$	1	$1+\beta+\beta^2$	$1+\beta^2$
$1+\beta$	0	$1+\beta$	$\beta+\beta^2$	$1+\beta^2$	$1+\beta+\beta^2$	β^2	1	β
β^2	0	β^2	$1+\beta$	$1+\beta+\beta^2$	$\beta+\beta^2$	β	$1+\beta^2$	1
$1+\beta^2$	0	$1+\beta^2$	1	β^2	β	$1+\beta+\beta^2$	$1+\beta$	$\beta+\beta^2$
$\beta+\beta^2$	0	$\beta+\beta^2$	$1+\beta+\beta^2$	1	$1+\beta^2$	$1+\beta$	β	β^2
$1+\beta+\beta^2$	0	$1+\beta+\beta^2$	$1+\beta^2$	β	1	$\beta+\beta^2$	β^2	$1+\beta$

///

Even though these two multiplication tables look completely different I claim that

$$\mathbb{E} \cong \mathbb{E}'.$$

Proof. It is difficult to find an isomorphism by hand so we will use an indirect method. First observe that $m_{\alpha/\mathbb{F}_2}(x) = x^3 + x^2 + 1$ is the minimal polynomial for $\alpha/\mathbb{F}_2$, so that

$$\mathbb{E} = \mathbb{F}_2(\alpha) \cong \frac{\mathbb{F}_2[x]}{\langle m_{\alpha/\mathbb{F}_2}(x)\rangle} = \frac{\mathbb{F}_2[x]}{\langle x^3 + x^2 + 1\rangle}.$$

Sadly, β does not satisfy the same equation. However, if we can prove that there exists **some** element $\gamma \in \mathbb{E}'$ satisfying $\gamma^3+\gamma^2+1 = 0$ then since x^3+x^2+1 is irreducible over $\mathbb{F}_2$ we will conclude that $m_{\gamma/\mathbb{F}_2}(x) = x^3 + x^2 + 1$ and hence

$$\mathbb{E} = \mathbb{F}_2(\alpha) \cong \frac{\mathbb{F}_2[x]}{\langle m_{\alpha/\mathbb{F}_2}(x)\rangle} = \frac{\mathbb{F}_2[x]}{\langle m_{\gamma/\mathbb{F}_2}(x)\rangle} \cong \mathbb{F}_2(\gamma) \subseteq \mathbb{E}'.$$

Finally, since $\mathbb{F}_2(\gamma)$ and $\mathbb{E}'$ both have size 8 we will conclude that $\mathbb{E} \cong \mathbb{F}_2(\gamma) = \mathbb{E}'$.

To prove the existence of such an element we consider the group of units $(\mathbb{E}^\times, \times, 1)$. Since $\#\mathbb{E}^\times$ has size 7, Lagrange's Theorem tells us that $v^7 = 1$ for all $v \in \mathbb{E}^\times$. In particular, since $\alpha \in \mathbb{E}^\times$ we must have $\alpha^7 - 1 = 0$. Then since $x^3 + x^2 + 1$ is the minimal polynomial for $\alpha/\mathbb{F}_2$ we conclude that

$$(x^3 + x^2 + 1)f(x) = (x^7 - 1) \text{ for some } f(x) \in \mathbb{F}_2[x] \text{ of degree } 4.$$

Next consider any non-zero element $\gamma \in \mathbb{E}'$. Since the group of units of $\mathbb{E}'$ also has size 7 we conclude again from Lagrange's Theorem that $\gamma^7 = 1$ and hence

$$(\gamma^3 + \gamma^2 + 1)f(\gamma) = (\gamma^7 - 1) = 0.$$

Since this is true for all $0 \neq \gamma \in \mathbb{E}'$ and since $f(x)$ has at most 4 roots in $\mathbb{E}'$, we conclude that there exist at least three (hence exactly three) elements $\gamma \in \mathbb{E}'$ such that $\gamma^3 + \gamma^2 + 1 = 0$. □

That's the best I can do by hand. To be more explicit, I used my computer to check that

$$\gamma = 1 + \beta, \quad 1 + \beta^2, \quad 1 + \beta + \beta^2$$

are the three promised roots of $x^3 + x^2 + 1$ in the field $\mathbb{E}'$. Then by sending $\alpha \mapsto \gamma$ we obtain the following three explicit isomorphisms:

0	1	α	$1+\alpha$	α^2	$1+\alpha^2$	$\alpha+\alpha^2$	$1+\alpha+\alpha^2$
0	1	$1+\beta$	β	$1+\beta^2$	β^2	$\beta+\beta^2$	$1+\beta+\beta^2$
0	1	$1+\beta^2$	β^2	$1+\beta+\beta^2$	$\beta+\beta^2$	β	$1+\beta$
0	1	$1+\beta+\beta^2$	$\beta+\beta^2$	$1+\beta$	β	β^2	$1+\beta^2$

We will prove later that there can be no other isomorphisms $\mathbb{E} \cong \mathbb{E}'$ because the Galois group $\mathrm{Gal}(\mathbb{E}/\mathbb{F}_2)$ has size $3 = [\mathbb{E}/\mathbb{F}_2]$.

Note that the above proof does not imply that **all** fields of size 8 are isomorphic, just these two **particular** fields of size 8. It happens to be true that any two finite fields of the same size are isomorphic but in order to prove this we need an extra ingredient called the Primitive Root Theorem.

20.2 Uniqueness of Finite Fields

The existence of finite fields beyond $\mathbb{Z}/p\mathbb{Z}$ was discovered by Galois.[145] However, the concept of isomorphism is more modern. E. H. Moore first stated and proved the uniqueness of finite fields in *A Doubly-Infinite System of Simple Groups* (1896), which was read at the International Mathematical Congress in Chicago in 1893.[146] This is the same paper in which he introduced the English term "field" for the German "Körper". Moore denoted the unique field of size p^k by GF$[p^k]$ for "Galois field", but I will use the modern notation $\mathbb{F}_{p^k}$.

In the next two lectures we will complete our discussion of finite fields by proving that for all $p, k \in \mathbb{Z}$ with p prime and $k \geq 1$, there **exists** a field of size p^k which is **unique** up to isomorphism. The full proof will require three lemmas, two of which you will prove on the homework. The first lemma shows that any finite field whatsoever has the form $\mathbb{F}_p[x]/\langle f(x)\rangle$ for some irreducible polynomial $f(x) \in \mathbb{F}_p[x]$.

Lemma (Primitive Root Theorem). If $\mathbb{E}$ is a finite field then $(\mathbb{E}^\times, \times, 1)$ is a cyclic group.

[145] Gauss probably discovered them independently but he didn't publish the results. Gauss' approach to publication was described by his motto: *Pauca sed matura* (Few, but ripe). His extensive mathematical notebooks were published after his death and complicated many issues of priority.

[146] The main topic of the paper is the family $PSL_2(p^k)$ of finite simple groups.

To be specific, let $\mathbb{F}_p \subseteq \mathbb{E}$ be the prime subfield and suppose that $[\mathbb{E}/\mathbb{F}_p] = k$, hence $\#\mathbb{E} = p^k$. Since $\mathbb{E}^\times$ is cyclic we can write $\mathbb{E} = \{0, 1, \alpha, \alpha^2, \ldots, \alpha^{p^k-2}\}$ for some $\alpha \in \mathbb{E}$. Then since every element of $\mathbb{E}$ can be expressed in terms of $\mathbb{F}_p \cup \{\alpha\}$ using field operations we conclude that

$$\mathbb{E} = \mathbb{F}_p(\alpha) \cong \frac{\mathbb{F}_p[x]}{\langle m_{\alpha/\mathbb{F}_p}(x)\rangle} \qquad \text{with } \deg(m_{\alpha/\mathbb{F}_p}) = k.$$

Proof. Homework. □

Remarks:

- Recall that a generator of the group $\langle e^{2\pi i/n}\rangle \subseteq \mathbb{C}^\times$ is called a *primitive n-th root of unity*. The number of generators is $\phi(n)$ and they are given by $e^{2\pi ik/n}$ for $\gcd(k, n) = 1$. Since $\langle e^{2\pi i/n}\rangle \cong \mathbb{Z}/n\mathbb{Z}$ the term "primitive roots" can also be applied to the additive generators of $\mathbb{Z}/n\mathbb{Z}$.
- In the *Disquisitiones Arithmeticae* (1801) Gauss proved for any prime p that the multiplicative group $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic. He applied the term *primitive root* to any generator of this group and said that he was following Euler's notation.
- Next, Galois claimed without showing any details that Gauss' proof can be extended to show that the fields $GF(p^k)$ also have cyclic groups of units. He followed Gauss in calling the generators *primitive roots*.
- So now the term "primitive root" had three different meanings. We can extend the meaning yet again by observing that if $\alpha \in \mathbb{E}^\times$ is a primitive root (multiplicative generator for the units of a finite field) then it follows that $\mathbb{E} = \mathbb{F}_p(\alpha)$, so that α is a generator of the field extension $\mathbb{E} \supseteq \mathbb{F}_p$.
- Finally, if $\mathbb{E} = \mathbb{F}(\gamma) \supseteq \mathbb{F}$ is any field extension generated by a single element $\gamma \in \mathbb{E}$ then $\gamma \in \mathbb{E}$ is called a *primitive element* for the extension. Later we will prove the so-called Primitive Element Theorem, which says that a primitive element exists when $\mathrm{char}(\mathbb{F}) = 0$ and $[\mathbb{E}/\mathbb{F}] < \infty$.
- In conclusion, the terms "primitive root" and "primitive element" are confusing and terrible. I prefer the term "Galois resolvent" instead of "primitive element", but the damage has already been done. The most I can do is warn you. ///

Theorem (Uniqueness of Finite Fields). Let $\mathbb{E}$ and $\mathbb{E}'$ be finite fields. Then

$$\#\mathbb{E} = \#\mathbb{E}' \quad \Longrightarrow \quad \mathbb{E} \cong \mathbb{E}'.$$

Proof. Let $\mathbb{F}_p \subseteq \mathbb{E}$ be the prime subfield and suppose that $[\mathbb{E}/\mathbb{F}_p] = k$, hence $\#\mathbb{E} = p^k$. From the Primitive Root Theorem there exists some $\alpha \in \mathbb{E}$ of multiplicative order $p^k - 1$. It follows from this that $\mathbb{E} = \{0, 1, \alpha, \alpha^2, \ldots, \alpha^{p^k-2}\} =$

$\mathbb{F}_p(\alpha)$ and hence

$$\mathbb{E} = \mathbb{F}_p(\alpha) \cong \frac{\mathbb{F}_p[x]}{\langle m_{\alpha/\mathbb{F}_p}(x)\rangle} \quad \text{with } \deg(m_{\alpha/\mathbb{F}_p}) = k.$$

Since $\alpha^{p^k-1} = 1$ we also know that

$$m_{\alpha/\mathbb{F}_p}(x)f(x) = \left(x^{p^k-1} - 1\right) \text{ for some } f(x) \in \mathbb{F}_p[x] \text{ of degree } p^k - 1 - k.$$

Now consider the field $\mathbb{E}'$, which also has size p^k. Since the group of units of $\mathbb{E}'$ has size $p^k - 1$ we conclude from Lagrange's Theorem that $\gamma^{p^k-1} = 1$ and hence

$$m_{\alpha/\mathbb{F}_p}(\gamma)f(\gamma) = \left(\gamma^{p^k-1} - 1\right) = 0 \quad \text{for all } 0 \neq \gamma \in \mathbb{E}'.$$

Since this holds for $p^k - 1$ distinct values of γ and since $\deg(f) < p^k - 1$ there must exist some $\gamma \in \mathbb{E}'$ such that $m_{\alpha/\mathbb{F}_p}(\gamma) = 0$ and hence $m_{\alpha/\mathbb{F}_p}(x) = m_{\gamma/\mathbb{F}_p}(x) \in \mathbb{F}_p[x]$. It follows that

$$\mathbb{E} = \mathbb{F}_p(\alpha) \cong \frac{\mathbb{F}_p[x]}{\langle m_{\alpha/\mathbb{F}_p}(x)\rangle} = \frac{\mathbb{F}_p[x]}{\langle m_{\gamma/\mathbb{F}_p}(x)\rangle} \cong \mathbb{F}_p(\gamma) \subseteq \mathbb{E}'.$$

Finally, since $\#\mathbb{F}_p(\gamma) = \#\mathbb{E} = \#\mathbb{E}'$ we conclude that

$$\mathbb{E} \cong \mathbb{F}_p(\gamma) = \mathbb{E}'.$$

□

[**Remark:** In fact, since $x^{p^k-1}-1$ splits in $\mathbb{E}'[x]$ we conclude that $m_{\alpha/\mathbb{F}_p}(x)$ also splits in $\mathbb{E}'[x]$.[147] Then from the Repeated Root Lemma below, this implies that $m_{\alpha/\mathbb{F}_p}(x)$ has k distinct roots $\gamma \in \mathbb{E}'$, leading to k distinct isomorphisms $\mathbb{E} \cong \mathbb{E}'$. We will see later that there can be no other isomorphisms between $\mathbb{E}$ and $\mathbb{E}'$.]

20.3 Existence of Finite Fields

So far we have proved that:

- Any finite field has size p^k for some prime p.
- Any two finite fields of the same size are isomorphic.

We have also see that irreducible polynomials in $\mathbb{F}_p[x]$ can be used to create finite fields. Indeed, if $f(x) \in \mathbb{F}_p[x]$ is irreducible of degree k then we obtain a field of size p^k:

$$\#\left(\frac{\mathbb{F}_p[x]}{\langle f(x)\rangle}\right) = p^k.$$

[147] Here were are using the fact that $\mathbb{E}'[x]$ is a UFD.

But it not obvious whether irreducible polynomials **exist**. Gauss gave a tricky proof for the existence of irreducible polynomials in the *Disquisitiones*. Galois was inspired by Gauss' work and he came up with an elegant direct proof for the existence of finite fields, which does not assume the existence of an irreducible polynomial, but obtains one as a corollary.

The proof requires two more lemmas, one of which you will prove on the homework.

Lemma (Repeated Roots). Let $\mathbb{F}$ be a field and let $D : \mathbb{F}[x] \to \mathbb{F}[x]$ be the "formal derivative" of polynomials. Given a polynomial $f(x) \in \mathbb{F}[x]$ we say that $\alpha \in \mathbb{F}[x]$ is a *repeated root* of $f(x)$ if $f(x) = (x-\alpha)^2 g(x)$ for some $g(x)$. Then I claim that

$$\alpha \text{ is a repeated root of } f(x) \qquad \iff \qquad f(\alpha) = 0 \quad \text{and} \quad Df(\alpha) = 0.$$

Proof. Homework. □

Lemma (The Frobenius Endomorphism).[148] Let R be any ring of **prime characteristic** p. Then the map $a \mapsto a^p$ defines a ring homomorphism $R \to R$, called the *Frobenius endomorphism* of R.

Proof. Note that $0^p = 0$ and $1^p = 1$, and for any $a, b \in R$ note that $(ab)^p = a^p b^p$. It only remains to show that $(a+b)^p = a^p + b^p$ for all $a, b \in R$. By definition we say that R has characteristic p when the unique ring homomorphism $\iota : \mathbb{Z} \to R$ has kernel $p\mathbb{Z}$. Then for all $n \in p\mathbb{Z}$ and $a \in R$ it follows that $\iota(n)a = 0$. Now recall the binomial theorem:

$$(a+b)^p = a^p + b^p + \sum_{k=1}^{p-1} \iota\left(\frac{p!}{k!(p-k)!}\right) a^k b^{p-k} \in R.$$

Now let $1 \leq k \leq p-1$ and consider the prime factorization of the binomial coefficient $p!/\left[k!(p-k)!\right] \in \mathbb{Z}$. Clearly p divides the numerator. But the denominator $k!(p-k)!$ is a product of integers, each of which is smaller than p.

[148] This result is sometimes called the "Freshman's Binomial Theorem", which I think is undignified. Apparently the name of Georg Frobenius was attached to this idea by Helmut Hasse in (1926–1930) because of a related and more difficult result proved by Frobenius. See the footnote on page 325 of *The Mathematics of Frobenius in Context* (2013), by Thomas Hawkins. Frobenius was a Berlin mathematician who nevertheless was influenced by Dedekind and the Göttingen school. He is known for proving many deep theorems that now form the backbone of linear algebra and representation theory. The history of linear algebra is much harder to trace than the history of abstract algebra because it is so ubiquitous in every area of mathematics.

Thus it follows from Euclid's lemma that p does **not** divide the denominator, and we conclude that

$$\frac{p!}{k!(p-k)!} \in p\mathbb{Z} \qquad \Longrightarrow \qquad \iota\left(\frac{p!}{k!(p-k)!}\right) = 0 \in R.$$

□

The following proof comes from Galois' paper *On the theory of numbers* (1830). This is the reason that finite fields are sometimes called "Galois fields".

Theorem (Existence of Finite Fields). For any integers $p, k \geq 1$ with p prime there exists a field of size p^k. In fact, I claim that any splitting field of $x^{p^k} - x \in \mathbb{F}_p[x]$ has size p^k.

Proof. The idea of this proof is due to Galois. Let $\mathbb{E} \supseteq \mathbb{F}_p$ be a splitting field for the polynomial $f(x) \coloneqq x^{p^k} - x \in \mathbb{F}_p[x]$. From the Repeated Root Lemma we know that if $\alpha \in \mathbb{E}$ is a repeated root of $f(x)$ then we must have $f(\alpha) = 0$ and $Df(\alpha) = 0$. But the derivative is $Df(x) = p^k x^{p^k - 1} - 1 = 0 - 1 = -1 \in \mathbb{F}_p[x]$, which has no roots at all. It follows that $f(x)$ has p^k distinct roots in $\mathbb{E}$. Let $\Omega \subseteq \mathbb{E}$ be the set of roots. We will show that in fact $\Omega \subseteq \mathbb{E}$ is a subfield, hence $\Omega = \mathbb{E}$ is our desired field of size p^k.

Indeed, Ω contains 0 and 1. Furthermore, if $f(\alpha) = 0$ and $f(\beta) = 0$ with $\alpha \neq 0$ then we have

$$(\alpha\beta)^{p^k} = \alpha^{p^k}\beta^{p^k} = \alpha\beta \quad \Longrightarrow \quad f(\alpha\beta) = 0$$

and

$$\left(\alpha^{-1}\right)^{p^k} = \left(\alpha^{p^k}\right)^{-1} = \alpha^{-1} \quad \Longrightarrow \quad f(\alpha^{-1}) = 0.$$

Finally, by applying the Frobenius endomorphism k times we obtain

$$(\alpha + \beta)^{p^k} = (\alpha^p + \beta^p)^{p^{k-1}} = (\alpha^{p^2} + \beta^{p^2})^{p^{k-2}} = \cdots = \alpha^{p^k} + \beta^{p^k} = \alpha + \beta,$$

and hence $f(\alpha + \beta) = 0$. □

Notation. We have seen that there exists a unique field of size p^k for any integers $p, k \geq 1$ with p prime. We will use the following notation for this field:

$$\boxed{\mathbb{F}_{p^k} \coloneqq \text{the unique field of size } p^k.}$$

Observe that this agrees with the earlier notation $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. ///

Corollary (Existence of Irreducible Polynomials). For any integers $p, k \geq 1$ with p prime there exists at least one irreducible polynomial $f(x) \in \mathbb{F}_p[x]$ of degree k.

Proof. Consider the field $\mathbb{F}_{p^k}$. By Lagrange's Theorem, the prime subfield (being an additive subgroup) must have size dividing p^k, hence the prime subfield is $\mathbb{F}_p \subseteq \mathbb{F}_{p^k}$. Observe that $[\mathbb{F}_{p^k}/\mathbb{F}_p] = k$ since for any finite-dimensional vector space V over $\mathbb{F}_p$ we have

$$\#V = p^{\dim(V)}.$$

Next, recall recall from the Primitive Root Theorem that there exists an element $\gamma \in \mathbb{F}_{p^k}$ such that $\mathbb{F}_{p^k} = \mathbb{F}_p(\gamma)$. Finally, we have

$$[\mathbb{F}_p(\gamma)/\mathbb{F}_p] = [\mathbb{F}_{p^k}/\mathbb{F}_p] = k,$$

which implies that the minimal polynomial $m_{\gamma/\mathbb{F}_p}(x) \in \mathbb{F}_p[x]$ has degree k. (Recall that minimal polynomials are always irreducible.) □

[**Remark:** It is not necessarily easy to find an irreducible polynomial of a given degree.]

Remarks:

- It follows from the uniqueness of finite fields that any two splitting fields of $x^{p^k} - x \in \mathbb{F}_p[x]$ are isomorphic. Next week we will prove that the same result holds for the splitting fields of **any** polynomial over **any** field.
- Conversely, let $\mathbb{E}$ be any field of size p^k with prime subfield $\mathbb{F}_p \subseteq \mathbb{E}$. From Lagrange's Theorem applied to the group of units, one can show that **every** element of $\mathbb{E}$ is a root of $x^{p^k-1} - x \in \mathbb{F}_p[x]$ and hence $\mathbb{E}$ is a splitting field for this polynomial. Thus the uniqueness of splitting fields will give a new proof for the uniqueness of finite fields.
- As I mentioned at the beginning of this lecture, Gauss proved in the *Disquisitiones* that there exist irreducible polynomials of all degrees in $\mathbb{F}_p[x]$. The way he did this was to first **count** the polynomials. To be specific, he first showed that the number of irreducible polynomials of degree k over $\mathbb{F}_p$ is given by

  $$\frac{1}{k}\sum_{d|k} \mu(k/d)p^d,$$

  where $\mu : \mathbb{Z} \to \{-1, 0, 1\}$ is the number-theoretic *möbius function*. Then he gave a tricky argument that this formula can never equal zero. This result can be viewed as the proof of existence for finite fields, but Gauss never discussed this in print. It turns out that Gauss privately developed a full theory of finite fields in parallel with Galois' theory, but this was only discovered after his death in 1855.[149]

[149] According to Günther Frei (2005) these results appear in an early unpublished section eight of the *Disquisitiones*, written by Gauss in 1797. Moreover, it seems that Gauss' treatment was more rigorous than that of Galois, since he treated the "Galois imaginagies" as cosets of polynomials. The "unpublished section eight" was first published by Dedekind in 1863 with a second printing in 1876.

Exercises

20.A Formal Derivation and Repeated Roots

If $\mathbb{F}$ is a field then we can think of the ring of polynomials $\mathbb{F}[x]$ as an infinite dimensional $\mathbb{F}$-vector space with basis $\{1, x, x^2, \ldots\}$. Let $D : \mathbb{F}[x] \to \mathbb{F}[x]$ be the unique $\mathbb{F}$-linear function defined by

$$D(x^n) = nx^{n-1} \text{ for all } n \geq 0.$$

(a) For all polynomials $f(x), g(x) \in \mathbb{F}[x]$ prove that the *product rule* holds:

$$D(fg) = f \cdot Dg + Df \cdot g.$$

[Hint: Show that each side is an $\mathbb{F}$-bilinear function of f and g. Thus it suffices to check the case when $f = x^m$ and $g = x^n$ are basis elements.]

(b) For all polynomials $f(x) \in \mathbb{F}[x]$ use part (a) and induction to prove the *power rule*:

$$D(f^n) = nf^{n-1} \cdot Df \text{ for all } n \geq 0.$$

(c) Consider a polynomial $f(x) \in \mathbb{F}[x]$ and a field extension $\mathbb{E} \supseteq \mathbb{F}$. We say that $\alpha \in \mathbb{E}$ is a *repeated root* of f when $f(x) = (x - \alpha)^2 g(x)$ for some polynomial $g(x) \in \mathbb{E}[x]$. Use part (a) to prove that

$$\alpha \text{ is a repeated root of } f \iff f(\alpha) = 0 \text{ and } Df(\alpha) = 0.$$

20.B The Primitive Root Theorem

If $\mathbb{F}$ is a finite field then the group of units $\mathbb{F}^\times$ is cyclic.

(a) Consider $m, n \in \mathbb{Z}$ with $\gcd(m, n) = 1$. If $m|nk$ prove that $m|k$. If $m|k$ and $n|k$ prove that $mn|k$. [Hint: Since $\gcd(m, n) = 1$ there exist $x, y \in \mathbb{Z}$ with $mx + ny = 1$.]

(b) Let A be an abelian group. If elements $a, b \in A$ have orders m, n with $\gcd(m, n) = 1$, prove that ab has order mn. [Hint: Show that $(ab)^k = \varepsilon$ implies $m|k$ and $n|k$.]

(c) Let A be an abelian group. If m is the **maximal order** of an element, prove that every element has order dividing m. [Hint: Let $a, b \in A$ have orders ℓ, m with $\ell \nmid m$. Then for some prime p we have $\ell = p^i \ell'$ and $m = p^j m'$ with $p \nmid \ell', m'$ and $i > j$. Use (b) to show that $a^{\ell'} b^{p^j}$ has order greater than m.]

(d) If $\alpha \in \mathbb{F}^\times$ is an element of **maximal order** m, prove that $\mathbb{F}^\times = \{1, \alpha, \ldots, \alpha^{m-1}\}$. [Hint: If not then the polynomial $x^m - 1 \in \mathbb{F}[x]$ has too many roots. Use (c).]

20.C Laplace's Proof of the FTA

The FTA is easily proved with complex analysis. However, it is still nice to have an elementary proof that is mostly algebraic. The following proof from Laplace (1795) builds on earlier ideas of Euler (1749) and Lagrange (1770). A logical gap in the proof was later filled by Kronecker's Theorem (1887). Specifically, we will prove that

every non-constant polynomial $f(x) \in \mathbb{R}[x]$ *has a root in* $\mathbb{C}$.

(a) Observe that every polynomial $f(x) \in \mathbb{R}[x]$ of odd degree has a root in $\mathbb{R}$.

(b) Now let $f(x) \in \mathbb{R}[x]$ have degree $n = 2^e m$ with $e \geq 1$ and m odd. Consider $f(x)$ as an element of $\mathbb{C}[x]$ and let $\mathbb{E} \supseteq \mathbb{C}$ be a splitting field:

$$f(x) = (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n) \in \mathbb{E}[x].$$

Now for any real number $\lambda \in \mathbb{R}$ we define the polynomial

$$g_\lambda(x) \coloneqq \prod_{1 \leq i < j \leq n} (x - \beta_{ij\lambda}) \in \mathbb{E}[x] \quad \text{with} \quad \beta_{ij\lambda} \coloneqq \alpha_i + \alpha_j + \lambda \alpha_i \alpha_j \in \mathbb{E}.$$

Prove that $g_\lambda(x) \in \mathbb{R}[x]$ and $\deg(g_\lambda) = 2^{e-1} m'$ with m' odd. [Hint: Newton.]

(c) By induction on e we can assume that $g_\lambda(x)$ has a complex root $\beta_{ij\lambda} \in \mathbb{C}$. If we apply this argument for more than $\binom{n}{2}$ different values of $\lambda \in \mathbb{R}$ then we will find specific indices $i < j$ and real numbers $\lambda \neq \mu$ such that $\beta_{ij\lambda}$ and $\beta_{ij\mu}$ are **both** in $\mathbb{C}$. In this case prove that α_i and α_j are in $\mathbb{C}$, hence $f(x)$ has a complex root.

Week 21

21.1 The Finiteness Theorem

After our detour through ring and field theory, we are finally ready to resume our discussion of Galois groups. The following definition is due to Dedekind, although he was only interested in the case when $\mathbb{E}$ is a subfield of $\mathbb{C}$.

Dedekind's Definition of Galois Groups. For any field extension $\mathbb{E} \supseteq \mathbb{F}$ we define

$$\mathrm{Gal}(\mathbb{E}/\mathbb{F}) \coloneqq \{\text{field automorphisms } \sigma : \mathbb{E} \to \mathbb{E} \text{ such that } \sigma(a) = a \text{ for all } a \in \mathbb{F}\}.$$

///

Our first goal is to show that the Galois group $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ is **finite** whenever $\mathbb{E}$ is a **finite-dimensional** vector space over $\mathbb{F}$. The proof will involve the notion of multi-variable polynomials. We skirted around this concept before, but now I will give you the official definition.

Definition of Multi-Variable Polynomials. Let R be a field and let $\{x_1, \ldots, x_n\}$ be a set of formal symbols, called "variables". We define the ring of polynomials by induction:

$$R[x_1, \ldots, x_n] \coloneqq R[x_1, \ldots, x_{n-1}][x_n].$$

Explicitly, each element of this ring has the form

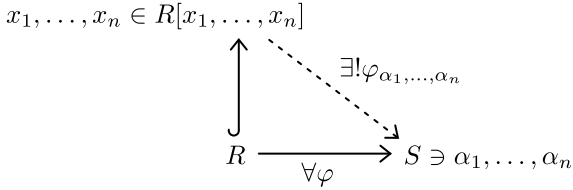
$$f(x_1, \ldots, x_n) = \sum a_{k_1,\ldots,k_n} x_1^{k_1} \cdots x_n^{k_n},$$

where the sum is over all n-tuples of natural numbers $(k_1, \ldots, k_n) \in \mathbb{N}^n$ and all but finitely many of the coefficients $a_{k_1,\ldots,k_n} \in R$ are equal to zero. This ring also satisfies a universal property, which is inherited from the one variable case.

Let $\varphi : R \to S$ be any ring homomorphism and let $\alpha_1, \ldots, \alpha_n \in S$ be any elements, not necessarily distinct. Then there exists a unique ring homomorphism $\varphi_{\alpha_1,\ldots,\alpha_n} : R[x_1, \ldots, x_n] \to S$ sending $x_i \mapsto \alpha_i$ for all i and acting on the coefficients by φ. Here is the explicit definition:

$$\varphi_{\alpha_1,\ldots,\alpha_n}\left(\sum a_{k_1,\ldots,k_n} x_1^{k_1} \cdots x_n^{k_n}\right) = \sum \varphi(a_{k_1,\ldots,k_n}) \alpha_1^{k_1} \cdots \alpha_n^{k_n}.$$

And here is the commutative diagram:



In modern terms we say that

$$R[x_1, \ldots, x_n] \text{ is the } \textit{free } R\textit{-algebra generated by the set } \{x_1, \ldots, x_n\}.$$

///

The definition of multi-variable polynomials is motivated by the following fact, which generalizes the case of one variable. For any ring extension $E \supseteq R$ and for any elements $\alpha_1, \ldots, \alpha_n \in E$, the smallest subring containing the set $R \cup \{\alpha_1, \ldots, \alpha_n\}$ is equal to the image of the evaluation:

$$R[\alpha_1, \ldots, \alpha_n] = \mathrm{im}(id_{\alpha_1,\ldots,\alpha_n}).$$

Proof. Since $\mathrm{im}(id_{\alpha_1,\ldots,\alpha_n}) \subseteq E$ is a subring containing the set $R \cup \{\alpha_1, \ldots, \alpha_n\}$, it must contain the smallest such subring. Conversely, since $R[\alpha_1, \ldots, \alpha_n] \subseteq E$ is a subring containing the set $R \cup \{\alpha_1, \ldots, \alpha_n\}$, it must contain every polynomial expression $f(\alpha_1, \ldots, \alpha_n)$. □

We are ready to prove our first theorem about Galois groups. Before reading the proof you may want to go back and remind yourself about the Orbit-Stabilizer Theorem for group actions.

The Finiteness Theorem. Let $[\mathbb{E}/\mathbb{F}] < \infty$ and $G = \mathrm{Gal}(\mathbb{E}/\mathbb{F})$. Then

(1) There exist elements $\alpha_1, \ldots, \alpha_n \in \mathbb{E}$ such that $\mathbb{E} = \mathbb{F}(\alpha_1, \ldots, \alpha_n)$.

(2) Every element of $\mathbb{E}$ is algebraic over $\mathbb{F}$, hence

$$\mathbb{F}(\alpha_1, \ldots, \alpha_n) = \mathbb{F}[\alpha_1, \ldots, \alpha_n].$$

(3) If $\sigma \in G$ satisfies $\sigma(\alpha_i) = \alpha_i$ for all i then we have $\sigma = id$.

(4) Let $m_i(x) \in \mathbb{F}[x]$ be the minimal polynomial of $\alpha_i/\mathbb{F}$. Then we have

$$\#G \leq \deg(m_1)\deg(m_2)\cdots\deg(m_n)$$

and hence G is finite.

///

Proof. (1) If $\mathbb{E} = \mathbb{F}$ then we are done. Otherwise, let $\alpha_1 \in \mathbb{E} - \mathbb{F}$ and consider the extension $\mathbb{F}(\alpha_1) \supseteq \mathbb{F}$. Since $[\mathbb{F}(\alpha_1)/\mathbb{F}] > 1$ we have $[\mathbb{E}/\mathbb{F}(\alpha_1)] < [\mathbb{E}/\mathbb{F}]$ and it follows by induction that there exist elements $\alpha_2, \ldots, \alpha_n \in \mathbb{E}$ such that

$$\mathbb{E} = \mathbb{F}(\alpha_1)(\alpha_2, \ldots, \alpha_n) = \mathbb{F}(\alpha_1, \alpha_2, \ldots, \alpha_n).$$

(2) Consider any element $\alpha \in \mathbb{E}$. Since $[\mathbb{E}/\mathbb{F}] < \infty$ we know that the set $\{1, \alpha, \alpha^2, \ldots\}$ is linearly dependent over $\mathbb{F}$, hence there exist some coefficients $a_0, \ldots, a_k \in \mathbb{F}$, not all zero, such that $a_0 + a_1\alpha + \cdots + a_k\alpha^k = 0$. In other words, α is algebraic over $\mathbb{E}$ over $\mathbb{F}$.

Now consider the elements $\alpha_1, \ldots, \alpha_n$ from part (1). Since α_1 is algebraic over $\mathbb{F}$ the Minimal Polynomial Theorem tells us that $\mathbb{F}[\alpha_1]$ is a field and hence $\mathbb{F}[\alpha_1] = \mathbb{F}(\alpha_1)$. Then since α_2 is algebraic over $\mathbb{F}$, hence also over $\mathbb{F}(\alpha_1)$, the Minimal Polynomial Theorem tells us that $\mathbb{F}[\alpha_1](\alpha_2) = \mathbb{F}[\alpha_1][\alpha_2] = \mathbb{F}[\alpha_1, \alpha_2]$. Continuing in this way gives the result.

(3) Consider any element $\sigma \in G$ such that $\sigma(\alpha_i) = \alpha_i$ for all i. From part (2) and the remarks before the theorem we know that every element of $\mathbb{E}$ can be expressed as $f(\alpha_1, \ldots, \alpha_n)$ for some polynomial $f(x_1, \ldots, x_n) \in \mathbb{F}[x_1, \ldots, x_n]$. But then since σ fixes $\mathbb{F}$ and preserves ring operations we have

$$\sigma\left(f(\alpha_1, \ldots, \alpha_n)\right) = f\left(\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\right) = f(\alpha_1, \ldots, \alpha_n),$$

and hence $\sigma = id$.

(4) From (2) we know that each generator $\alpha_i \in \mathbb{E}$ has a minimal polynomial $m_i(x) \in \mathbb{F}[x]$. For any $\sigma \in G$ we observe that $\sigma(\alpha_i)$ is a root of $m_i(x)$ because

$$m_i(\sigma(\alpha_i)) = \sigma(m_i(\alpha)) = \sigma(0) = 0.$$

In other words, the group G acts on the set $\Omega_i \subseteq \mathbb{E}$ of roots of $m_i(x)$ in the field $\mathbb{E}$. Moreover, G acts on the Cartesian product of sets:

$$G \curvearrowright (\Omega_1 \times \cdots \times \Omega_n).$$

Let $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n)$ be the only element of this set that we know,[150] and consider the G-orbit:

$$\mathrm{Orb}(\vec{\alpha}) = \{(\sigma(\alpha_1), \ldots, \sigma(\alpha_n)) : \sigma \in G\} \subseteq \Omega_1 \times \cdots \times \Omega_n.$$

[150] In fact, there may be no other elements.

In part (3) we proved that the stabilizer is trivial: $\mathrm{Stab}(\vec{\alpha}) = \{id\}$. Hence from the Orbit-Stabilizer Theorem we obtain a bijection

$$G \leftrightarrow \frac{G}{\{id\}} = \frac{G}{\mathrm{Stab}(\vec{\alpha})} \leftrightarrow \mathrm{Orb}(\vec{\alpha}) \subseteq \Omega_1 \times \cdots \times \Omega_n.$$

Since $m_i(x)$ has at most $\deg(m_i)$ roots in the field $\mathbb{E}$ we have $\#\Omega_i \leq \deg(m_i)$ and hence

$$\#G = \#\mathrm{Orb}(\vec{\alpha}) \leq \#(\Omega_1 \times \cdots \times \Omega_n) = \#\Omega_1 \times \cdots \times \#\Omega_n \leq \prod \deg(m_i) < \infty.$$

□

In fact, Dedekind proved a sharper bound:

$$\#\mathrm{Gal}(\mathbb{E}/\mathbb{F}) \leq [\mathbb{E}/\mathbb{F}].$$

However, his proof used very different methods.[151] For pedagogical reasons I chose to prove a weaker result using more relevant methods.

Below we will see that Galois theory is concerned with certain "nice" field extensions $\mathbb{E} \supseteq \mathbb{F}$ that achieve the upper bound: $\#\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = [\mathbb{E}/\mathbb{F}]$. From the proof of the Finiteness Theorem we can already see how this might happen. Suppose that an extension $\mathbb{E} \supseteq \mathbb{F}$ satisfies:

- $\mathbb{E} = \mathbb{F}(\gamma)$ for some element $\gamma \in \mathbb{E}$ with minimal polynomial $m(x) \in \mathbb{F}[x]$,
- the polynomial $m(x)$ splits in $\mathbb{E}[x]$,
- the polynomial $m(x)$ has no multiple roots in $\mathbb{E}$,
- the Galois group $G = \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ acts transitively on the roots of $m(x)$. In other words, for any two roots α, β there exists a group element $\sigma \in G$ with $\sigma(\alpha) = \beta$.

In this case let $\Omega \subseteq \mathbb{E}$ be the set of roots of $m(x)$. From the assumptions we have

$$\#\Omega = \deg(m) = [\mathbb{F}(\gamma)/\mathbb{F}] = [\mathbb{E}/\mathbb{F}].$$

Now if $\sigma \in G$ fixes γ then for every element $f(\gamma) \in \mathbb{F}[\gamma] = \mathbb{F}(\gamma) = \mathbb{E}$ we have $\sigma(f(\gamma)) = f(\sigma(\gamma)) = f(\gamma)$ and hence $\sigma = id$. Finally, since G acts transitively on Ω we obtain bijections

$$G \leftrightarrow \frac{G}{\{id\}} = \frac{G}{\mathrm{Stab}(\gamma)} \leftrightarrow \mathrm{Orb}(\gamma) = \Omega$$

and it follows that $\#G = \#\Omega = [\mathbb{E}/\mathbb{F}]$. ///

It may seem to you that the four properties above are rather special, but we will soon prove that these properties hold for a large and natural class of field extensions. To be specific, we will show that these four properties hold whenever:

[151] It uses the "linear independence of characters".

- $\mathbb{F}$ is finite or has characteristic zero,[152]
- $\mathbb{E}$ is a splitting field for some polynomial $f(x) \in \mathbb{F}[x]$.

21.2 Definition of Galois Extensions

From now on we will only consider finite-dimensional field extensions. Last time we proved that if $[\mathbb{E}/\mathbb{F}] < \infty$ then the group $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ is finite. I also mentioned (but did not prove) Dedekind's theorem, which says that

$$\#\mathrm{Gal}(\mathbb{E}/\mathbb{F}) \leq [\mathbb{E}/\mathbb{F}].$$

We have a special name for field extensions that achieve this bound.

Definition of Galois Extensions. Let $\mathbb{E} \supseteq \mathbb{F}$ be a finite-dimensional field extension. We say that $\mathbb{E}/\mathbb{F}$ is a *Galois extension*[153] if the following equality holds:

$$\#\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = [\mathbb{E}/\mathbb{F}].$$

///

Galois extensions are the field-theoretic version of "normal subgroups". (The Fundamental Theorem of Galois Theory will make this analogy precise.) And, just as with normal subgroups, there are several equivalent ways to state the definition. Before investigating this, let me show you some small examples.

Example: If $[\mathbb{E}/\mathbb{F}] = 1$ or 2 then $\mathbb{E}/\mathbb{F}$ is Galois.

Proof. If $[\mathbb{E}/\mathbb{F}] = 1$ then we have $\mathbb{E} = \mathbb{F}$ and it follows that $\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = \{id\}$. Now suppose that $[\mathbb{E}/\mathbb{F}] = 2$. On a previous homework you showed that this implies $\mathbb{E} = \mathbb{F}(\iota)$ for some element $\iota \in \mathbb{E}$ with $\iota \notin \mathbb{F}$ and $\iota^2 \in \mathbb{F}$. Let me briefly recall the proof.

Choose any $\alpha \in \mathbb{E} - \mathbb{F}$. Since $[\mathbb{E}/\mathbb{F}] = 2$ we know that the set $\{1, \alpha, \alpha^2\}$ is linearly dependent over $\mathbb{Q}$. Since $\alpha \notin \mathbb{F}$ it follows that $f(\alpha) = 0$ for some $f(x) \in \mathbb{Q}[x]$ of degree 2. Let $\beta \in \mathbb{E}$ be the other root of $f(x)$ and define $\iota := \alpha - \beta$. Then $\iota^2 = (\alpha - \beta)^2 \in \mathbb{F}$ is the discriminant of $f(x)$ and one can show that $\mathbb{E} = \mathbb{F}(\alpha) = \mathbb{F}(\iota)$.

If follows that $\{1, \iota\}$ is a basis for the vector space $\mathbb{E}/\mathbb{F}$. To compute the Galois group, let $\sigma \in \mathrm{Gal}(\mathbb{E}/\mathbb{F})$. Then for any element $a + b\iota \in \mathbb{E}$ we have

$$\sigma(a + b\iota) = a + b\sigma(\iota).$$

[152] This includes every field that you have ever seen, so it is barely a restriction.

[153] I don't know the origin of this terminology but it seems reasonable. In the literature you will see a "Galois extension" defined as "finite-dimensional, normal and separable". These last two terms have technical meanings that are only relevant for infinite fields of positive characteristic. I think it is appropriate to ignore such fields in a first course on the subject. (Also, I plan never to teach a second course.)

Furthermore, since $\iota^2 = a$ for some $a \in \mathbb{F}$ we must also have

$$\sigma(\iota)^2 = \sigma(\iota^2) = \sigma(a) = a.$$

Since the polynomial $x^2 - a \in \mathbb{F}[x]$ has at most two roots in $\mathbb{E}$, this implies that $\sigma(\iota) = \iota$ or $\sigma(\iota) = -\iota$. The first choice corresponds to the identity element and the second choice yields the following function:

$$\tau(a + b\iota) \coloneqq a - b\iota.$$

One can check by hand that this function is, indeed, a field automorphism and hence

$$\#\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = \#\{id, \tau\} = 2 = [\mathbb{E}/\mathbb{F}].$$

□

[**Remark:** This result is analogous to the fact that a subgroup $H \subseteq G$ satisfying $\#(G/H) = 2$ is necessarily normal. Again, The Fundamental Theorem of Galois Theory will make this analogy precise.]

Non-Example: The field extension $\mathbb{Q}(\sqrt[3]{2}) \supseteq \mathbb{Q}$ is not Galois.

Proof. Let $\alpha = \sqrt[3]{2} \in \mathbb{R}$ be the real cube root of 2. Since α is a root of the irreducible polynomial $x^3 - 2 \in \mathbb{Q}[x]$ we know from the Minimal Polynomial Theorem that

$$[\mathbb{Q}(\alpha)/\mathbb{Q}] = \deg(x^3 - 2) = 3.$$

On the other hand, for any $\sigma \in \mathrm{Gal}(\mathbb{Q}(\alpha)/\mathbb{Q})$ we must have

$$0 = \sigma(0) = \sigma(\alpha^3 - 2) = \sigma(\alpha)^3 - 2.$$

Then since $\mathbb{Q}(\alpha) \subseteq \mathbb{R}$ and since α is the only real root of $x^3 - 2$ we must have $\sigma(\alpha) = \alpha$. Finally, since $\alpha/\mathbb{Q}$ is algebraic we know that every element of $\mathbb{Q}(\alpha) = \mathbb{Q}[\alpha]$ has the form $f(\alpha)$ for some polynomial $f(x) \in \mathbb{Q}[x]$ and hence

$$\sigma\left(f(\alpha)\right) = f\left(\sigma(\alpha)\right) = f(\alpha).$$

If follows that $\#\mathrm{Gal}(\mathbb{Q}(\alpha)/\mathbb{Q}) = \#\{id\} = 1 < 3 = [\mathbb{Q}(\alpha)/\mathbb{Q}]$. □

[**Remark:** The ultimate problem with this example is that the field $\mathbb{Q}(\sqrt[3]{2})$ contains only one root of the irreducible polynomial $x^3 - 2 \in \mathbb{Q}[x]$. We can fix this problem by passing to the splitting field.]

Example: The splitting field of $x^3 - 2 \in \mathbb{Q}[x]$ is Galois.

Proof. The roots of $x^3 - 2 \in \mathbb{Q}[x]$ are the complex numbers $\alpha, \omega\alpha, \omega^2\alpha \in \mathbb{C}$ where $\alpha = \sqrt[3]{2} \in \mathbb{R}$ and $\omega = e^{2\pi i/3}$. It follows that the splitting field is

$$\mathbb{E} = \mathbb{Q}(\alpha, \omega\alpha, \omega^2\alpha) = \mathbb{Q}(\alpha, \omega).$$

Since x^3-2 is the minimal polynomial for $\alpha/\mathbb{Q}$ and since x^2+x+1 is the minimal polynomial for $\omega/\mathbb{Q}(\alpha)$ we conclude from the Minimal Polynomial Theorem and Dedekind's Tower Law that $[\mathbb{E}/\mathbb{Q}] = 6$ with basis $\{1, \alpha, \alpha^2, \omega, \omega\alpha, \omega\alpha^2\}$. To compute the Galois group, let $\sigma \in \mathrm{Gal}(\mathbb{E}/\mathbb{Q})$. Then for any $a, b, c, d, e, f \in \mathbb{Q}$ we have

$$\begin{aligned}\sigma(a + b\alpha + c\alpha^2 + d\omega + e\alpha\omega + f\alpha^2\omega)& \\ = a + b\sigma(\alpha) + c\sigma(\alpha)^2 + c\sigma(\omega) + e\sigma(\alpha)\sigma(\omega) + f\sigma(\alpha)^2\sigma(\omega).&\end{aligned}$$

It follows that the function σ is determined by the two numbers $\sigma(\alpha)$ and $\sigma(\omega)$. Furthermore, since $\sigma(\alpha)$ is a root of x^3-2 and since $\sigma(\omega)$ is a root of x^2+x+1 we must have

$$\sigma(\alpha) \in \{\alpha, \omega\alpha, \omega^2\alpha\} \qquad \text{and} \qquad \sigma(\omega) \in \{\omega, \omega^2\}.$$

Since all of these roots exist in $\mathbb{E}$ we obtain six different functions $\sigma : \mathbb{E} \to \mathbb{E}$. These functions are necessarily $\mathbb{F}$-linear, hence they fix $\mathbb{F}$ and preserve addition.

But how do we know that these functions preserve multiplication?

For the moment let me just say that one can check this by hand, or, better, with a computer. It follows that $\#\mathrm{Gal}(\mathbb{E}/\mathbb{Q}) = 6 = [\mathbb{E}/\mathbb{Q}]$ and hence $\mathbb{E}/\mathbb{Q}$ is Galois. □

This last example illustrates two points:

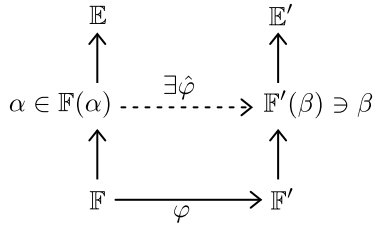
- Splitting fields are likely to be Galois. In fact, we will prove below that any splitting field $\mathbb{E}/\mathbb{F}$ is Galois as long as $\mathbb{F}$ is finite or has characteristic zero.
- The hard part of the proof is to show that certain functions defined on the roots can be **lifted** to automorphisms of the splitting field. Clearly the brute force method is not good enough. We will need a general theorem about this.

21.3 The Splitting Field Theorem

The Finiteness Theorem showed that Galois groups are small. Now we want to prove that the Galois group of a splitting field is big. The Splitting Field Theorem below is probably the most subtle theorem in this course. It is good to prepare for this theorem with a lemma.

The Lifting Lemma. Let $\mathbb{E} \supseteq \mathbb{F}$ and $\mathbb{E}' \supseteq \mathbb{F}'$ be field extensions and let $\varphi : \mathbb{F} \to \mathbb{F}'$ be an isomorphism. Let $\alpha \in \mathbb{E}$ be a root of an **irreducible** polynomial $f(x) \in \mathbb{F}[x]$ and let $\beta \in \mathbb{E}'$ be any root of $f^\varphi(x) \in \mathbb{E}'$. Then there

exists a field isomorphism $\hat{\varphi} : \mathbb{F}(\alpha) \to \mathbb{F}'(\beta)$ lifting $\varphi : \mathbb{F} \to \mathbb{F}'$ and sending $\alpha \mapsto \beta$. Here is a picture:



It is worth highlighting the special case when $\mathbb{F} = \mathbb{F}'$, $\mathbb{E} = \mathbb{E}'$ and $\varphi = id$. In this case if $\alpha, \beta \in \mathbb{E}$ are any two roots of an irreducible polynomial $f(x) \in \mathbb{F}$ then there exists an isomorphism $\mathbb{F}(\alpha) \to \mathbb{F}(\beta)$ sending $\alpha \mapsto \beta$ and fixing the elements of $\mathbb{F}$. ///

Proof. The proof is easy, but only because we have developed the right technology. Let $\varphi : \mathbb{F} \to \mathbb{F}'$ be a field isomorphism and let $\alpha \in \mathbb{E} \supseteq \mathbb{F}$ be a root of an **irreducible** polynomial $f(x) \in \mathbb{F}[x]$. Since $\varphi : \mathbb{F} \to \mathbb{F}'$ is a ring isomorphism we obtain an isomorphism of polynomial rings $\mathbb{F}[x] \to \mathbb{F}'[x]$ by letting φ act on the coefficients:

$$\begin{array}{ccc} \mathbb{F}[x] & \xrightarrow{\sim} & \mathbb{F}'[x] \\ f(x) & \mapsto & f^{\varphi}(x). \end{array}$$

Suppose that there exist a root $\beta \in \mathbb{E}'$ of the image polynomial $f^{\varphi}(x) \in \mathbb{F}[x]$. Since $f(x)$ and $f^{\varphi}(x)$ are both irreducible, it follows that these polynomial are (up to non-zero scalar multiples) the minimal polynomials for $\alpha/\mathbb{F}$ and $\beta/\mathbb{F}'$, respectively. Then from the isomorphism $\mathbb{F}[x] \cong \mathbb{F}'[x]$ and the Minimal Polynomial Theorem we obtain a sequence of three isomorphisms:

$$\begin{array}{ccccccc} \mathbb{F}(\alpha) & \cong & \mathbb{F}[x]/\langle f(x)\rangle & \cong & \mathbb{F}'[x]/\langle f^{\varphi}(x)\rangle & \cong & \mathbb{F}'(\beta) \\ \alpha & \leftrightarrow & x + \langle f(x)\rangle & \leftrightarrow & x + \langle f^{\varphi}(x)\rangle & \leftrightarrow & \beta. \end{array}$$

Composing these gives the desired isomorphism $\mathbb{F}(\alpha) \cong \mathbb{F}'(\beta)$. □

Application: Complex Conjugation. Let $\mathbb{E} = \mathbb{F}(\iota)$ with $\iota^2 = a \in \mathbb{F}$ and $\iota \notin \mathbb{F}$. Then $\pm\iota$ are roots of the irreducible polynomial $x^2 - a \in \mathbb{F}[x]$. It follows from the Lifting Lemma that there exists an isomorphism $\mathbb{E} = \mathbb{F}(\iota) \to \mathbb{F}(-\iota) = \mathbb{E}$ sending $\iota \mapsto -\iota$ and fixing $\mathbb{F}$. This proves that the "conjugation" map $a + b\iota \mapsto a - b\iota$ is a field automorphism, without doing any calculations. ///

That was a small time savings, but the next one is substantial.

Application: The Splitting Field of $x^3 - 2 \in \mathbb{Q}[x]$. Recall that there exist six functions $\sigma : \mathbb{E} \to \mathbb{E}$ defined by letting $\sigma(\alpha)$ and $\sigma(\omega)$ be any roots of

$x^3 - 2$ and $x^2 + x + 1$, respectively. Let's prove that these functions are field automorphisms.

First, let $\alpha' \in \{\alpha, \omega\alpha, \omega^2\alpha\}$ be any root of $x^3 - 2$. Since $x^3 - 2$ is irreducible there exists a field isomorphism $\varphi : \mathbb{Q}(\alpha) \to \mathbb{Q}(\alpha')$ sending $\alpha \mapsto \alpha'$ and fixing $\mathbb{Q}$. Next, observe that the polynomial $x^2 + x + 1$ is still irreducible over $\mathbb{Q}(\alpha)$ because $\mathbb{Q}(\alpha) \subseteq \mathbb{R}$ and $x^2 + x + 1$ has no real roots. Thus if $\omega' \in \{\omega, \omega^2\}$ is any root of $x^2 + x + 1$ then there exists an isomorphism $\hat{\varphi} : \mathbb{Q}(\alpha)(\omega) \to \mathbb{Q}(\alpha')(\omega')$ lifting φ and sending $\omega \mapsto \omega'$. In particular, this $\hat{\varphi}$ also sends $\alpha \mapsto \alpha'$ and fixes $\mathbb{Q}$. Finally, since

$$\mathbb{E} = \mathbb{Q}(\alpha, \omega) = \mathbb{Q}(\alpha)(\omega) \cong \mathbb{Q}(\alpha')(\omega') \subseteq \mathbb{E}$$

we conclude that $\hat{\varphi} : \mathbb{E} \to \mathbb{E}$ is a field automorphism. Thus we have proved the **existence** of the six desired elements of the Galois group $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$. ///

The following theorem simply generalizes this procedure. The theorem is strangely worded, but this is only for the purposes of the induction proof. Our real interest is the special case when $\varphi = id : \mathbb{F} \to \mathbb{F}$ is the identity. Before stating the theorem it is worth restating the definition of a splitting field.

Let $f(x) \in \mathbb{F}[x]$ be a polynomial. We say that $\mathbb{E} \supseteq \mathbb{F}$ is a *splitting field* for $f(x)$ if

- The polynomial $f(x)$ splits in $\mathbb{E}[x]$. That is, we have

$$f(x) = (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n) \text{ for some } \alpha_1, \ldots, \alpha_n \in \mathbb{E}.$$

- If $\mathbb{E} \supseteq \mathbb{K} \supseteq \mathbb{F}$ and if $f(x)$ splits in $\mathbb{K}[x]$ then $\mathbb{K} = \mathbb{E}$. Equivalently, we have

$$\mathbb{E} = \mathbb{F}(\alpha_1, \ldots, \alpha_n).$$

The Splitting Field Theorem (Existence of Automorphisms). Consider the following:

- Let $\varphi : \mathbb{F} \to \mathbb{F}'$ be an isomorphism of fields and let $f(x) \in \mathbb{F}[x]$ be any polynomial.
- Let $\mathbb{E} \supseteq \mathbb{F}$ and $\mathbb{E}' \supseteq \mathbb{F}'$ be splitting fields for $f(x) \in \mathbb{F}[x]$ and $f^{\varphi}(x) \in \mathbb{F}'[x]$.
- Let $p_i(x) | f(x)$ be a list of distinct[154] irreducible factors in $\mathbb{F}[x]$.

[154]Technically: We assume that the polynomials a pairwise non-associate. That is, for any $i \neq j$ and $\lambda \in \mathbb{F}$ we have $p_i(x) \neq \lambda p_j(x)$.

- For each i let $\alpha_i \in \mathbb{E}$ be a root of $p_i(x)$ and let $\beta_i \in \mathbb{E}'$ be a root of $p_i^{\varphi}(x)$. Note that such roots always exist because $\mathbb{E}$ and $\mathbb{E}'$ are splitting fields.[155]

Then there exists an isomorphism $\Phi : \mathbb{E} \to \mathbb{E}'$ lifting φ and sending $\alpha_i \mapsto \beta_i$ for all i. ///

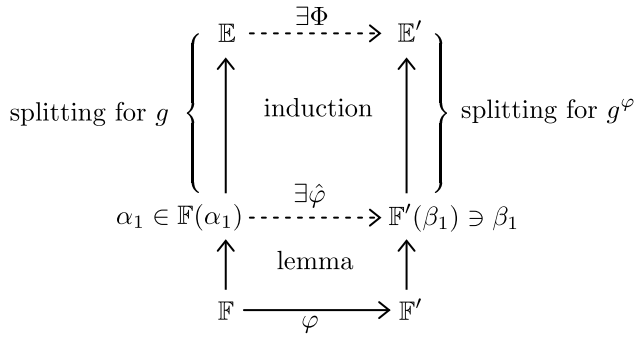
[**Remark:** There are two small issues in this proof that you will check on the homework. Namely, (a) any divisor of a split polynomial is also split, and (b) non-associate irreducible polynomials have no roots in common.]

Proof. We will use induction on $\deg(f)$. The result is vacuously true when $\deg(f) = 1$ so let $\deg(f) = n \geq 2$ and let $\mathbb{E} \supseteq \mathbb{F}$ be a splitting field for $f(x) \in \mathbb{F}[x]$. If $p_1(x)|f(x)$ is any irreducible factor then since $p_1(x)$ splits in $\mathbb{E}$ we know that $p_1(x)$ has a root, say $\alpha_1 \in \mathbb{E}$. Next, observe that $p_1^{\varphi}(x)|f^{\varphi}(x)$ in $\mathbb{F}'[x]$. Since $\mathbb{E}'$ is a splitting field for $f^{\varphi}(x)$ this implies that $p_1^{\varphi}(x)$ has some root, say $\beta_1 \in \mathbb{E}'$. Thus from the Lifting Lemma we obtain an isomorphism $\hat{\varphi} : \mathbb{F}(\alpha_1) \to \mathbb{F}'(\beta_1)$ lifting φ and sending $\alpha_1 \mapsto \beta_1$.

Next, by Descartes' Theorem there exists $g(x) \in \mathbb{F}(\alpha_1)[x]$ of degree $n-1$ such that $f(x) = (x - \alpha_1)g(x)$ and by applying the isomorphism $\hat{\varphi}$ we obtain $f^{\varphi}(x) = (x - \beta_1)g^{\varphi}(x)$ for some $g^{\varphi}(x) \in \mathbb{F}'[x]$. Observe that $\mathbb{E} \supseteq \mathbb{F}(\alpha)$ is a splitting field for $g(x)$ since if $g(x)$ splits over an intermediate field $\mathbb{E} \supseteq \mathbb{K} \supseteq \mathbb{F}(\alpha)$ then $f(x)$ also splits over $\mathbb{K}$. Since $\mathbb{E}$ is a splitting field for $f(x)$ this implies that $\mathbb{K} = \mathbb{E}$. Similarly, $\mathbb{E}'$ is a splitting field for $g^{\varphi}(x)$.

Furthermore, if $p_2(x)|f(x)$ is irreducible and not a scalar multiple of $p_1(x)$ then since $p_1(x), p_2(x)$ have no common root we must have $p_2(x) \nmid (x - \alpha_1)$ and hence $p_2(x)|g(x)$. Finally, let $\alpha_i \in \mathbb{E}$ and $\beta_i \in \mathbb{E}'$ be any roots of the polynomials $p_i(x) \in \mathbb{F}[x]$ and $p^{\varphi}(x) \in \mathbb{F}'[x]$ for $i \geq 2$. Since $\deg(g) < \deg(f)$ we conclude by induction that there exists an isomorphism $\Phi : \mathbb{E} \to \mathbb{E}'$ lifting $\hat{\varphi} : \mathbb{F}(\alpha_1) \to \mathbb{F}'(\beta_1)$ and sending $\alpha_i \mapsto \beta_i$ for all $i \geq 2$, hence Φ also lifts φ and sends $\alpha_1 \mapsto \beta_1$. Here is a picture:

[155] This is fairly obvious but it still needs a proof. You will provide a proof on the homework, using the fact that $\mathbb{E}[x]$ is a UFD.



□

To see how this applies to the previous example, let $\mathbb{E} \supseteq \mathbb{Q}$ be a splitting field for $x^3 - 2$ and observe that $\mathbb{E}$ is also a splitting field for $f(x) = (x^3 - 2)(x^2 + x + 1)$. Let $p_1(x) = x^3 - 2$ and $p_2(x) = x^2 + x + 1$. Then for any roots α, α' of $x^3 - 2$ and roots ω, ω' of $x^2 + x + 1$ there exists an isomorphism $\mathbb{E} \to \mathbb{E}$ sending $(\alpha, \omega) \mapsto (\alpha', \omega')$ and fixing $\mathbb{Q}$. This is a powerful theorem.

In addition to helping us compute Galois groups, the Splitting Field Theorem has an important theoretical corollary.

Corollary (Uniqueness of Splitting Fields). Let $\mathbb{E}, \mathbb{E}' \supseteq \mathbb{F}$ be splitting fields for the same polynomial $f(x) \in \mathbb{F}[x]$. Then there exists a (non-unique) isomorphism $\Phi : \mathbb{E} \to \mathbb{E}'$ fixing $\mathbb{F}$.

Proof. Take $\mathbb{F} = \mathbb{F}'$ and $\varphi = id$ in the theorem. Ignore the roots of $f(x)$. □

So far we have only defined Galois groups for extensions. This corollary allows us to define the Galois group of a polynomial. Note that this is the reverse of the historical development.

The Galois Group of a Polynomial. Let $f(x) \in \mathbb{F}[x]$ be any polynomial and let $\mathbb{E} \supseteq \mathbb{F}$ be any splitting field for $f(x)$. We define the *Galois group of f over $\mathbb{F}$* as follows:

$$\mathrm{Gal}(f/\mathbb{F}) \coloneqq \mathrm{Gal}(\mathbb{E}/\mathbb{F}).$$

I claim that this group is well-defined up to isomorphism.

Proof. Let $\mathbb{E}, \mathbb{E}' \supseteq \mathbb{F}$ be any two splitting fields and let $\Phi : \mathbb{E} \to \mathbb{E}'$ be an isomorphism fixing $\mathbb{F}$, which exists by the corollary. Then claim that the map $\sigma \mapsto \Phi \circ \sigma \circ \Phi^{-1}$ is a group isomorphism $\mathrm{Gal}(\mathbb{E}/\mathbb{F}) \to \mathrm{Gal}(\mathbb{E}'/\mathbb{F})$. Indeed, for any $\sigma \in \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ we observe that the function $\Phi \circ \sigma \circ \Phi^{-1} : \mathbb{E}' \to \mathbb{E}'$ is a field

automorphism that fixes $\mathbb{F}$. Then we observe that the function $\sigma \mapsto \Phi \circ \sigma \circ \Phi^{-1}$ is invertible and preserves composition. $\square$

Remarks:

- The uniqueness of splitting fields is not news for finite extensions of $\mathbb{Q}$. Indeed, if $[\mathbb{F}/\mathbb{Q}] < \infty$ then since every element of $\mathbb{F}$ is algebraic over $\mathbb{Q}$ we know from the FTA that $\mathbb{F} \subseteq \mathbb{C}$. Then for any polynomial $f(x) \in \mathbb{F}[x]$ we may view the splitting field as the intersection of all subfields of $\mathbb{C}$ that contain the roots of $f(x)$.
- However, for fields of characteristic p we get new information. For example, let $\mathbb{E}$ be finite of characteristic p. Then we have previously shown that $\mathbb{E}$ is a splitting field for the polynomial $x^{p^k} - x \in \mathbb{F}_p[x]$ for some k. It follows from the uniqueness of splitting fields that any two fields of size p^k are isomorphic. This new proof is a bit more elegant than our old proof because it avoids the Primitive Root Theorem.
- It is worth emphasizing one more consequence of the Splitting Field Theorem. If $\mathbb{E} \supseteq \mathbb{F}$ is a splitting field for $f(x) \in \mathbb{F}[x]$ and if $p(x)|f(x)$ is any irreducible factor, then for any two roots $\alpha, \beta \in \mathbb{E}$ of $p(x)$ there exists an automorphism $\sigma : \mathbb{E} \to \mathbb{E}$ fixing $\mathbb{F}$ and sending $\alpha \mapsto \beta$. In other words:

  The Galois group $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ acts **transitively** on the roots of $p(x)$.

  Note that this fact does **not** apply to reducible polynomials. For example, let $\mathbb{E} \supseteq \mathbb{Q}$ be a splitting field for $f(x) = (x^2 - 2)(x^2 - 3)$. Then we have $f(\sqrt{2}) = f(\sqrt{3}) = 0$, but there does **not** exist any group element $\sigma \in \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ sending $\sqrt{2} \mapsto \sqrt{3}$. (Why not?)

Exercises

21.A Divisor of a Split Polynomial

Let $\mathbb{E} \supseteq \mathbb{F}$ be and suppose that $g(x)|f(x)$ in $\mathbb{F}[x]$. If $f(x)$ splits in $\mathbb{E}[x]$ prove that $g(x)$ also splits in $\mathbb{E}[x]$. [Hint: Use the fact that $\mathbb{E}[x]$ is a UFD.]

21.B Repeated Roots, Part II

We say that a polynomial $f(x) \in \mathbb{F}[x]$ is *inseparable* if it has a repeated root in some field extension. Otherwise we say that $f(x)$ is *separable*. Prove that

$$f(x) \text{ is separable} \quad \iff \quad \gcd(f, Df) = 1.$$

21.C Finite Fields are Separable

Let $\mathbb{E}$ be finite field of characteristic p. For all polynomials $f(x) \in \mathbb{E}[x]$ we will show that

$$f(x) \text{ is irreducible} \quad \implies \quad f(x) \text{ is separable.}$$

(a) Let $f(x) \in \mathbb{E}[x]$ be irreducible and assume for contradiction that $f(x)$ is inseparable. Prove that the derivative $Df(x) \in \mathbb{E}[x]$ is the zero polynomial.

(b) Use part (a) to show that $f(x) = g(x^p)$ for some polynomial $g(x) \in \mathbb{E}[x]$.

(c) Finally, show that $g(x^p) = h(x)^p$ for some polynomial $h(x) \in \mathbb{E}[x]$. Contradiction. [Hint: You showed in a previous problem that the Frobenius map $\alpha \mapsto \alpha^p$ is surjective.]

Week 22

22.1 Perfect Fields

We will see that that the basic theorems of Galois theory hold for finite fields and for fields of characteristic zero. However, it is a sad fact that there exist certain infinite fields of characteristic p for which the theorems break down. Ernst Steinitz (1910) was the first person to deal uniformly with the good cases, while excluding the pathological cases. He did this with the following definition.

Steinitz' Definition of Perfect Fields. We say that a field $\mathbb{F}$ is *perfect* [vollkommene] if

- $\text{char}(\mathbb{F}) = 0$, or
- $\text{char}(\mathbb{F}) = p$ and the function $\mathbb{F} \to \mathbb{F}$ defined by $a \mapsto a^p$ is surjective. You will show on the homework that this case includes **all finite fields**.

///

Unfortunately, this definition is just a notational device because any theorem about perfect fields requires two separate proofs for the two cases.[156] Here are the properties that we need for Galois theory.

Nice Properties of Perfect Fields. Let $\mathbb{F}$ be a perfect field. Then:

(1) **Irreducible Polynomials are Separable.** If $f(x) \in \mathbb{F}[x]$ is irreducible then $f(x)$ has no repeated roots in any field extension.

(2) **Primitive Elements Exist.** If $\mathbb{E} \supseteq \mathbb{F}$ is finite-dimensional then there exists an element $\gamma \in \mathbb{E}$ such that $\mathbb{E} = \mathbb{F}(\gamma)$.

///

[156] Maybe there is a deep connection between the two cases that I don't know about?

As I mentioned, this theorem requires separate proofs for the cases $\mathrm{char}(\mathbb{F}) = 0$ and $\#\mathbb{F} < \infty$. I hope you don't mind that I relegated some of the steps to the homework. Also, we will ignore the case of infinite perfect fields of characteristic p.

Proof. (1) Let $f(x) \in \mathbb{F}[x]$ be irreducible and assume for contradiction that $f(x)$ has a repeated root in some field extension. On the homework you will show that this implies $g(x) = \gcd(f, Df)$ has degree ≥ 1, where $Df(x) \in \mathbb{F}[x]$ is the formal derivative. Since $f(x)$ is irreducible this implies that $g(x) = \lambda f(x)$ for some $\lambda \in \mathbb{F}$. But we also know that $g(x)|Df(x)$. If $\mathrm{char}(\mathbb{F}) = 0$ then this is a contradiction because $\deg(Df) = \deg(f) - 1$. If $\#\mathbb{F} < \infty$ then it could possibly be the case that $Df(x)$ is identically zero, but you will show on the homework that this also leads to a contradiction.

(2) If $\#\mathbb{F} < \infty$ and $[\mathbb{E}/\mathbb{F}] < \infty$ then we also have $\#\mathbb{E} < \infty$. You showed on a previous homework that the group $(\mathbb{E}^\times, \times, 1)$ is cyclic (we called this the **Primitive Root Theorem**). Say $\mathbb{E}^\times = \{\gamma^n : n \in \mathbb{Z}\}$ for some $\gamma \in \mathbb{E}$. Then clearly every element of $\mathbb{E}$ can be expressed in the form $f(\gamma)$ for some $f(x) \in \mathbb{F}[x]$, hence $\mathbb{E} = \mathbb{F}(\gamma)$.

(2) Next suppose that $\mathrm{char}(\mathbb{F}) = 0$ and $[\mathbb{E}/\mathbb{F}] < \infty$. This case is sometimes called the **Primitive Element Theorem**. The proof is due to Galois.

By the Finiteness Theorem we know that $\mathbb{E} = \mathbb{F}(\alpha_1, \ldots, \alpha_n)$ for some algebraic elements $\alpha_1, \ldots, \alpha_n \in \mathbb{E}$. Thus by induction it suffices to prove for all algebraic $\alpha, \beta \in \mathbb{E}$ that

$$\mathbb{F}(\alpha, \beta) = \mathbb{F}(\gamma) \quad \text{for some } \gamma \in \mathbb{E}.$$

Let $m_\alpha(x), m_\beta(x) \in \mathbb{F}[x]$ be the minimal polynomials of α, β over $\mathbb{F}$. Then since $\mathbb{F}$ is **infinite** we may choose a non-zero element $c \in \mathbb{F}$ such that

$$c \neq \frac{\alpha' - \alpha}{\beta - \beta'} \quad \text{for all roots } \alpha' \neq \alpha \text{ of } m_\alpha \text{ and } \beta' \neq \beta \text{ of } m_\beta.$$

In this case I claim that $\gamma \coloneqq \alpha + c\beta$ is a primitive element. Indeed, since $\gamma \in \mathbb{F}(\alpha, \beta)$ we have $\mathbb{F}(\gamma) \subseteq \mathbb{F}(\alpha, \beta)$. Conversely, we want to show that $\alpha, \beta \in \mathbb{F}(\gamma)$ and hence $\mathbb{F}(\alpha, \beta) \subseteq \mathbb{F}(\gamma)$, and for this it suffices to prove $\beta \in \mathbb{F}(\gamma)$ since then we also have $\alpha = \gamma - c\beta \in \mathbb{F}(\gamma)$.

We will show that $\beta \in \mathbb{F}(\gamma)$ by an indirect argument. That is, let $m'_\beta(x) \in \mathbb{F}(\gamma)[x]$ be the minimal polynomial of β over $\mathbb{F}(\gamma)$. We will prove that $\deg(m'_\beta) = 1$ and hence $\beta \in \mathbb{F}(\gamma)$. By thinking of $m_\beta(x)$ as an element of $\mathbb{F}(\gamma)[x]$ we clearly have $m'_\beta(x)|m_\beta(x)$. Now we need to get α in on the action. So (TRICK) define the polynomial

$$f(x) \coloneqq m_\alpha(\gamma - cx) \in \mathbb{F}(\gamma)[x].$$

By construction we have $f(\beta) = m_\alpha(\gamma - c\beta) = m_\alpha(\alpha) = 0$ which implies that $m'_\beta(x)|f(x)$. It follows that any root of $m'_\beta(x)$ is a common root of $m_\beta(x)$ and $f(x)$.

Finally, let $\mathbb{E}' \supseteq \mathbb{E}$ be a splitting field for the polynomial $m_\alpha(x)m_\beta(x)$. We know from part (1) that each of the polynomials $m_\alpha(x), f(x), m_\beta(x), m'_\beta(x)$ splits and has no repeated roots in $\mathbb{E}'$. It follows that the number of common roots of $m_\beta(x)$ and $f(x)$ in $\mathbb{E}'$ is equal to $\deg(m'_\beta)$. We will be done if we can show that β **is the only common root**. So assume for contradiction that we have $m_\beta(\beta') = f(\beta') = 0$ for some $\beta' \neq \beta$. By definition of $f(x)$ this means that $\alpha' \coloneqq \gamma - c\beta'$ is a root of $m_\alpha(x)$. But then we have

$$\begin{aligned} \alpha' &= \gamma - c\beta' \\ \alpha' &= (\alpha + c\beta) - c\beta' \\ c &= (\alpha' - \alpha)/(\beta - \beta'), \end{aligned}$$

which contradicts the definition c. □

Here is an application to our favorite example.

Example: A Primitive Element for the Splitting Field of x^3-2. Recall that the splitting field is $\mathbb{Q}(\alpha, \omega)$ where $\alpha = \sqrt[3]{2} \in \mathbb{R}$ and $\omega = e^{2\pi i/3}$. We are looking for an element of the form $\gamma = \alpha + c\omega$ with nonzero $c \in \mathbb{Q}$ such that $\mathbb{Q}(\alpha, \omega) = \mathbb{Q}(\gamma)$. By the proof of the previous theorem it suffices to choose c such that

$$c \neq \frac{\alpha' - \alpha}{\omega - \omega'} \quad \text{for all } \alpha' \in \{\omega\alpha, \omega^2\alpha\} \text{ and } \omega' \in \{\omega^2\}.$$

But note that $\omega - \omega^2$ is purely imaginary and $\alpha' - \alpha$ never is. Thus we may take **any nonzero element** $c \in \mathbb{Q}$. For example, $c = 1$. ///

Remarks:

- The proof of (2) for finite fields goes back to Gauss and the proof of (2) for characteristic zero fields is due to Galois. In fact, John Stillwell[157] says that this was the first substantial result in Galois' 1831 memoir. In modern language, Galois' version says that for any algebraic imaginaries $\alpha, \beta \in \mathbb{C}$ there exists an integer $c \in \mathbb{Z}$ such that $\mathbb{Q}(\alpha, \beta) = \mathbb{Q}(\alpha + c\beta)$. By induction it follows that any finite extension satisfies

  $$\mathbb{Q}(\alpha_1, \ldots, \alpha_n) = \mathbb{Q}(c_1\alpha_1 + \cdots + c_n\alpha_n) \text{ for some integers } c_1, \ldots, c_n \in \mathbb{Z}.$$

  Such an element $\gamma = c_1\alpha_1 + \cdots + c_n\alpha_n$ was called a *Galois resolvent*, but today it is usually called a *primitive element*.

- Sadly, there exist pathological examples of finite-dimensional field extensions which do not have a primitive element. For example, consider the field $\mathbb{F} = \mathbb{F}_p(x, y)$ consisting of fractions $f(x, y)/g(x, y)$ where $f, g \in \mathbb{F}_p[x, y]$ and $g \neq 0$. Let $\mathbb{E} = \mathbb{F}(\alpha, \beta)$ where $\alpha^p = x$ and $\beta^p = y$. Then one can show that $[\mathbb{E}/\mathbb{F}] = p^2 < \infty$ but has no primitive element.

[157] *Elements of Algebra*, page 160

- When Dedekind modernized Galois theory he continued to use primitive elements because his main concern was with finite extensions of $\mathbb{Q}$. However, after Steinitz included characteristic p in his 1910 memoir, other mathematicians such as Emmy Noether began to reject the use of primitive elements because they are not completely general. In a 1935 memorial address[158] after Emmy Noether's death, Hermann Weyl praised her "drive toward axiomatic purity", but he thought that it was not always appropriate:

  > *This can be carried too far, however, as when she disdained to employ a primitive element in the development of Galois theory.*

- I agree with with Weyl that the use of primitive elements leads to the cleanest development of Galois theory, at least for beginners. You will see this in the next lecture.

22.2 Characterization of Galois Extensions

Dedekind proved for any field extension $\mathbb{E} \supseteq \mathbb{F}$ that $\#\mathrm{Gal}(\mathbb{E}/\mathbb{F}) \leq [\mathbb{E}/\mathbb{F}]$.[159] Recall that finite-dimensional extensions satisfying $\#\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = [\mathbb{E}/\mathbb{F}]$ are called *Galois extensions*. As with normal subgroups, there are several equivalent ways to state the definition. Today we will prove a big characterization theorem for Galois extensions over perfect fields.

We isolate the following lemma for pedagogical reasons. Emil Artin proved this lemma for general fields,[160] using linear algebraic techniques inspired by Dedekind. We will only prove it for finite-dimensional extensions over perfect fields.

Artin's Fixed Field Lemma. Let $\mathbb{E}$ be any field and let $G \subseteq \mathrm{Aut}(\mathbb{E})$ be any **finite** group of automorphisms with fixed field $\mathrm{Fix}_{\mathbb{E}}(G) \subseteq \mathbb{E}$. Then we have

$$[\mathbb{E}/\mathrm{Fix}_{\mathbb{E}}(G)] = \#G.$$

///

Proof. As I said, we will only prove a special case of this. Let $\mathbb{F}$ be perfect and let $\mathbb{E} \supseteq \mathbb{F}$ be a finite-dimensional extension, so there exists a primitive element $\mathbb{E} = \mathbb{F}(\gamma)$. Now consider any (finite) subgroup $G \subseteq \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ and let $\mathrm{Orb}_G(\gamma) = \{\gamma_1, \ldots, \gamma_n\}$ be the G-orbit of γ. Since $\mathbb{E}$ is generated by γ over $\mathbb{F}$ we see that $\mathrm{Stab}_G(\gamma) = \{id\}$ and hence

$$n = \#\mathrm{Orb}_G(\gamma) = \#G/\#\,\mathrm{Stab}_G(\gamma) = \#G.$$

[158] Reprinted as an appendix in *Emmy Noether: 1882–1935* by Auguste Dick (1981).
[159] We didn't prove this, but we did prove a weaker version called the Finiteness Theorem.
[160] It is Theorem 14 in his *Galois Theory* (1942), reprinted by Dover (1998).

Now consider the following polynomial with degree n and no repeated roots:

$$f(x) = (x - \gamma_1) \cdots (x - \gamma_n) \in \mathbb{E}[x].$$

Since every element of G permutes the roots of $f(x)$ it also fixes the coefficients, hence $f(x) \in \mathrm{Fix}_{\mathbb{E}}(G)[x]$. I claim in fact that $f(x)$ is the minimal polynomial for γ over $\mathrm{Fix}_{\mathbb{E}}(G)$. Indeed, let $m(x) \in \mathrm{Fix}_{\mathbb{E}}(G)[x]$ be the minimal polynomial. Then since $f(\gamma) = 0$ we have $m(x)|f(x)$. Conversely, since every $\gamma_i \in \mathrm{Orb}_G(\gamma)$ has the form $\sigma(\gamma)$ for some $\sigma \in G$ we must have

$$m(\gamma_i) = m(\sigma(\gamma)) = \sigma(m(\gamma)) = \sigma(0) = 0.$$

Then it follows from Descartes' Theorem that $f(x)|m(x)$ and hence $f(x) = m(x)$. Finally, since $\mathrm{Fix}_{\mathbb{E}}(\gamma) = \mathbb{F}(\gamma) = \mathbb{E}$ we conclude from the Minimal Polynomial Theorem that

$$[\mathbb{E}/\mathrm{Fix}_{\mathbb{E}}(G)] = [\mathrm{Fix}_{\mathbb{E}}(G)(\gamma)/\mathrm{Fix}_{\mathbb{E}}(G)] = \deg(m) = \deg(f) = n = \#G.$$

□

Before stating today's theorem let me note that if $\mathbb{F}$ is perfect and if $\mathbb{E} \supseteq \mathbb{F}$ is finite-dimensional then $\mathbb{E}$ is also perfect. Indeed, for all $\mathbb{E} \supseteq \mathbb{F}$ we have $\mathrm{char}(\mathbb{F}) = 0 \Rightarrow \mathrm{char}(\mathbb{E}) = 0$ and for all $[\mathbb{E}/\mathbb{F}] < \infty$ we have $\#\mathbb{F} < \infty \Rightarrow \#\mathbb{E} < \infty$. Again, we don't care about the other cases.

I find the following theorem amazing. Galois is lucky to have this concept named after him.

Characterization Theorem for Galois Extensions (of Perfect Fields). Let $\mathbb{E} \supseteq \mathbb{F}$ be a finite-dimensional extension of perfect fields. Then the following five conditions are equivalent:

(GE1) $\#\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = [\mathbb{E}/\mathbb{F}]$

(GE2) $\mathrm{Fix}_{\mathbb{E}}(\mathrm{Gal}(\mathbb{E}/\mathbb{F})) = \mathbb{F}$

(GE3) $\mathbb{E}$ is a splitting field for some polynomial $f(x) \in \mathbb{F}[x]$.

(GE4) For any $\mathbb{E}' \supseteq \mathbb{E}$ and $\sigma \in \mathrm{Gal}(\mathbb{E}'/\mathbb{F})$ we have $\sigma(\mathbb{E}) \subseteq \mathbb{E}$.[161]

(GE5) If $m(x) \in \mathbb{F}[x]$ is irreducible and has a root in $\mathbb{E}$, then $m(x)$ splits in $\mathbb{E}[x]$.

[161] And hence $\sigma(\mathbb{E}) = \mathbb{E}$. Indeed, since $\sigma(1) = 1$ we know that $\ker \sigma \neq \mathbb{E}$. Then since a field has no non-trivial ideals we must have $\ker \sigma = \{0\}$. Finally, since $\sigma : \mathbb{E}/\mathbb{F} \to \mathbb{E}/\mathbb{F}$ is an injective endomorphism of a finite-dimensional $\mathbb{F}$-vector space we conclude from the Rank-Nullity Theorem that σ is also surjective.

///

Proof. To save space we will write $G = \mathrm{Gal}(\mathbb{E}/\mathbb{F})$.

(GE1)⇔(GE2): Since $\mathbb{F}$ is perfect and $[\mathbb{E}/\mathbb{F}] < \infty$ there exists a primitive element $\gamma \in \mathbb{E}$ with $\mathbb{E} = \mathbb{F}(\gamma)$. Since $\mathbb{E} \supseteq \mathrm{Fix}_{\mathbb{E}}(G) \supseteq \mathbb{F}$ this also implies $\mathbb{E} = \mathrm{Fix}_{\mathbb{E}}(G)(\gamma)$. Then from the Fixed Field Lemma and Dedekind's Tower Law we have

$$[\mathbb{E}/\mathbb{F}] = [\mathbb{E}/\mathrm{Fix}_{\mathbb{E}}(G)] \cdot [\mathrm{Fix}_{\mathbb{E}}(G)/\mathbb{F}] = \#G \cdot [\mathrm{Fix}_{\mathbb{E}}(G)/\mathbb{F}].$$

It follows that

$$\#G = [\mathbb{E}/\mathbb{F}] \iff [\mathrm{Fix}_{\mathbb{E}}(G)/\mathbb{F}] = 1 \iff \mathrm{Fix}_{\mathbb{E}}(G) = \mathbb{F}.$$

(GE2)⇒(GE3): Assume that $\mathrm{Fix}_{\mathbb{E}}(G) = \mathbb{F}$. Since $\mathbb{F}$ is perfect there exists a primitive element $\mathbb{E} = \mathbb{F}(\gamma)$ with minimal polynomial $m(x) \in \mathbb{F}[x]$ satisfying $\deg(m) = [\mathbb{F}(\gamma)/\mathbb{F}] = [\mathbb{E}/\mathrm{Fix}_{\mathbb{E}}(G)] = \#G$. From the proof of the Fixed Field Lemma we also know that $m(x)$ has $\#G$ distinct roots in $\mathbb{E}$. It follows that $m(x)$ splits in $\mathbb{E}[x]$ and hence $\mathbb{E} = \mathbb{F}(\gamma)$ is a splitting field for $m(x) \in \mathbb{F}[x]$.

(GE3)⇒(GE4): Assume that there exists some $f(x) \in \mathbb{F}[x]$ with $f(x) = (x - \alpha_1) \cdots (x - \alpha_n) \in \mathbb{E}[x]$ and $\mathbb{E} = \mathbb{F}(\alpha_1, \ldots, \alpha_n)$. From the Finiteness Theorem we know that $\mathbb{F}(\alpha_1, \ldots, \alpha_n) = \mathbb{F}[\alpha_1, \ldots, \alpha_n]$. In other words, $\mathbb{E}$ is the set of evaluations $g(\alpha_1, \ldots, \alpha_n)$ of polynomials $g \in \mathbb{F}[x_1, \ldots, x_n]$. Now consider any field extension $\mathbb{E}' \supseteq \mathbb{E}$ and any automorphism $\sigma \in \mathrm{Gal}(\mathbb{E}'/\mathbb{F})$. Since σ fixes $\mathbb{F}$ it necessarily permutes the roots of $f(x)$. Then for any element $g(\alpha_1, \ldots, \alpha_n) \in \mathbb{E}$ we have

$$\sigma\left(g(\alpha_1, \ldots, \alpha_n)\right) = g\left(\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\right) \in \mathbb{E},$$

since this last expression is also a polynomial evaluated at the roots of $f(x)$.

(GE4)⇒(GE5): Let $m(x) \in \mathbb{F}[x]$ be irreducible and let $\mathbb{E}' \supseteq \mathbb{E}$ be a splitting field for $m(x)$. Let $\Omega \subseteq \mathbb{E}'$ be the roots of $m(x)$ and assume that this set contains an element of $\mathbb{E}$, say $\alpha \in \Omega \cap \mathbb{E}$. Now consider the group $G' = \mathrm{Gal}(\mathbb{E}'/\mathbb{F})$. Since $\mathbb{E}'$ is a splitting field for the irreducible polynomial $m(x) \in \mathbb{F}[x]$ we know from the Splitting Field Theorem that G' acts transitively on Ω. In other words, we have $\mathrm{Orb}_{G'}(\alpha) = \Omega$. But by assumption we also know that G' sends $\mathbb{E}$ to $\mathbb{E}$. It follows that

$$\Omega = \mathrm{Orb}_{G'}(\alpha) \subseteq \mathbb{E},$$

and hence $m(x)$ splits in $\mathbb{E}[x]$.

(GE5)⇒(GE1): Since $\mathbb{F}$ is perfect there exists a primitive element $\mathbb{E} = \mathbb{F}(\gamma)$. If $m(x) \in \mathbb{F}[x]$ is the minimal polynomial for $\gamma/\mathbb{F}$ then by assumption we know that $m(x)$ splits in $\mathbb{E}[x]$. Let $\Omega \subseteq \mathbb{E}$ be the set of roots of $m(x)$ and consider the action of G on Ω. Since γ generates $\mathbb{E}$ over $\mathbb{F}$ we have $\mathrm{Stab}_G(\gamma) = \{id\}$ and since $\mathbb{E} = \mathbb{F}(\gamma)$ is a splitting field for $m(x)$ we know from the Splitting

Field Theorem that $\mathrm{Orb}_G(\gamma) = \Omega$. Finally, since $\mathbb{F}$ is perfect[162] we know that $m(x)$ has no repeated roots in $\mathbb{E}$ and it follows that

$$\#G = \frac{\#G}{\#\,\mathrm{Stab}_G(\gamma)} = \#\mathrm{Orb}_G(\gamma) = \#\Omega = \deg(m_\gamma) = [\mathbb{F}(\gamma)/\mathbb{F}] = [\mathbb{E}/\mathbb{F}].$$

[This is the argument that I previewed after the proof of the Finiteness Theorem.] □

Remarks:

- Normally I don't like TFAE[163] theorems, but I can't think of any pedagogically better way to state these results.
- Observe that this theorem contains $5 \cdot 4 = 20$ implications. I tried to make the whole proof as short as possible, which has the drawback that your favorite implication might not be optimized.
- Many of these equivalences break for extensions of non-perfect fields. If you want to know the details about that then you are reading the wrong book.

22.3 The Fundamental Theorem of Galois Theory

The previous lecture was the most difficult of the course. It's all downhill from here.

Consider any field extension $\mathbb{E} \supseteq \mathbb{F}$ with Galois group $G = \mathrm{Gal}(\mathbb{E}/\mathbb{F})$. Let $\mathscr{L}(\mathbb{E}, \mathbb{F})$ be the lattice of intermediate fields and let $\mathscr{L}(G)$ be the lattice of subgroups. Now recall from the introduction of Part II that we have an abstract Galois connection:

$$\mathrm{Gal}(\mathbb{E}/-) : \mathscr{L}(\mathbb{E}, \mathbb{F}) \rightleftarrows \mathscr{L}(G)^{\mathrm{op}} : \mathrm{Fix}_{\mathbb{E}}(-).$$

Technically, this means that for all subfields $\mathbb{E} \supseteq \mathbb{K} \supseteq \mathbb{F}$ and for all subgroups $H \subseteq G$ we have

$$\mathbb{K} \subseteq \mathrm{Fix}_{\mathbb{E}}(H) \iff \mathrm{Gal}(\mathbb{E}/\mathbb{K}) \supseteq H.$$

Recall that an abstract Galois connection always restricts to an isomorphism between certain subposets of "closed elements". In general, it follows from Artin's Fixed Field Lemma that every **finite** subgroup of G is "closed". If $\mathbb{E} \supseteq \mathbb{F}$ is a Galois extension of perfect fields then it turns out that every intermediate field is also "closed", and in this case we have an isomorphism of lattices $\mathscr{L}(\mathbb{E}, \mathbb{F}) \cong \mathscr{L}(G)^{\mathrm{op}}$.[164] Here is the full statement.

[162] Alternatively, we could argue again that $m(x) = \prod_i (x - \gamma_i)$ where $\mathrm{Orb}_G(\gamma) = \{\gamma_i\}_i$.
[163] "The following are equivalent".
[164] This result can be extended to certain "infinite Galois extensions" by replacing the lattice of subgroups with the lattice of "profinite subgroups". Never mind.

The Fundamental Theorem of Galois Theory. Let $\mathbb{E} \supseteq \mathbb{F}$ be a Galois extension of perfect fields and let $G = \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ be the Galois group. Then:

(1) The Galois connection $\mathrm{Gal}(\mathbb{E}/-) : \mathscr{L}(\mathbb{E}, \mathbb{F}) \leftrightarrows \mathscr{L}(G)^{\mathrm{op}} : \mathrm{Fix}_{\mathbb{E}}(-)$ is actually a **bijection**. That is, for all intermediate fields $\mathbb{E} \supseteq \mathbb{K} \supseteq \mathbb{F}$ and for all subgroups $H \subseteq G$ we have

$$\mathrm{Fix}_{\mathbb{E}}(\mathrm{Gal}(\mathbb{E}/\mathbb{K})) = \mathbb{K} \qquad \text{and} \qquad \mathrm{Gal}(\mathbb{E}/\mathrm{Fix}_{\mathbb{E}}(H)) = H.^{165}$$

(2) For any pair $\mathbb{K} = \mathrm{Fix}_{\mathbb{E}}(H)$ and $H = \mathrm{Gal}(\mathbb{E}/\mathbb{K})$ we have

$$\#\{\text{cosets of } H \text{ in } G\} = \#(G/H) = [\mathbb{K}/\mathbb{F}] = \dim(\mathbb{K} \text{ as a vector space over } \mathbb{F}).$$

(3) Furthermore, we have

$$\mathbb{K} \supseteq \mathbb{F} \text{ is a Galois field extension} \iff H \trianglelefteq G \text{ is a normal subgroup,}$$

in which case the quotient group is isomorphic to the Galois group:

$$\frac{G}{H} = \frac{\mathrm{Gal}(\mathbb{E}/\mathbb{F})}{\mathrm{Gal}(\mathbb{E}/\mathbb{K})} \cong \mathrm{Gal}(\mathbb{K}/\mathbb{F}).$$

///

Proof. The proof will refer to the Characterization Theorem for Galois extensions.

(1) Consider any intermediate field $\mathbb{E} \supseteq \mathbb{K} \supseteq \mathbb{F}$. Since $\mathbb{E}/\mathbb{F}$ is Galois we know from (GE3) that $\mathbb{E}$ is a splitting field for some polynomial $f(x) \in \mathbb{F}[x]$. But then $\mathbb{E}$ is also a splitting field for $f(x) \in \mathbb{K}[x]$ which implies that $\mathbb{E}/\mathbb{K}$ is Galois. We conclude from (GE2) that $\mathrm{Fix}_{\mathbb{E}}(\mathrm{Gal}(\mathbb{E}/\mathbb{K})) = \mathbb{K}$.

Now consider any subgroup $H \subseteq G$ and let $\mathbb{K} = \mathrm{Fix}_{\mathbb{E}}(G)$, so that $H \subseteq \mathrm{Gal}(\mathbb{E}/\mathbb{K})$. As above we know that $\mathbb{E}/\mathbb{K}$ is Galois, hence from (GE1) we have $\#\mathrm{Gal}(\mathbb{E}/\mathbb{K}) = [\mathbb{E}/\mathbb{K}]$. On the other hand, we know from the Fixed Field Lemma that $\#H = [\mathbb{E}/\mathbb{K}]$ and it follows that $H = \mathrm{Gal}(\mathbb{E}/\mathbb{K})$.

(2) Consider any pair $\mathbb{K} = \mathrm{Fix}_{\mathbb{E}}(H)$ and $H = \mathrm{Gal}(\mathbb{E}/\mathbb{K})$. Since $\mathbb{E}/\mathbb{F}$ and $\mathbb{E}/\mathbb{K}$ are both Galois, we know from Lagrange's Theorem, (GE1) and Dedekind's Tower Law that

$$\#(G/H) = \frac{\#G}{\#H} = \frac{\#\mathrm{Gal}(\mathbb{E}/\mathbb{F})}{\#\mathrm{Gal}(\mathbb{E}/\mathbb{K})} = \frac{[\mathbb{E}/\mathbb{F}]}{[\mathbb{E}/\mathbb{K}]} = [\mathbb{K}/\mathbb{F}].$$

[165] As I mentioned above, the equation $\mathrm{Gal}(\mathbb{E}/\mathrm{Fix}_{\mathbb{E}}(H)) = H$ holds for any field $\mathbb{E}$ and for any finite group of automorphisms $H \subseteq \mathrm{Aut}(\mathbb{E})$. The proof only depends on Artin's Fixed Field Lemma (which, however, we did not prove in full generality). The other equation $\mathrm{Fix}_{\mathbb{E}}(\mathrm{Gal}(\mathbb{E}/\mathbb{K})) = \mathbb{K}$ is more interesting.

(3) Furthermore, I claim that

$$\mathrm{Gal}(\mathbb{E}/\sigma(\mathbb{K})) = \sigma\,\mathrm{Gal}(\mathbb{E}/\mathbb{K})\,\sigma^{-1} = \sigma H \sigma^{-1} \quad \text{for all } \sigma \in G.$$

Indeed, this follows immediately from the definitions:

$$\begin{aligned}
\mu \in \mathrm{Gal}(\mathbb{E}/\sigma(\mathbb{K})) &\iff \mu(\sigma(a)) = \sigma(a) \text{ for all } a \in \mathbb{K}. \\
&\iff (\sigma^{-1}\mu\sigma)(a) = a \text{ for all } a \in \mathbb{K}. \\
&\iff \sigma^{-1}\mu\sigma \in H \\
&\iff \mu \in \sigma H \sigma^{-1}.
\end{aligned}$$

Now suppose that $\mathbb{K}/\mathbb{F}$ is Galois. Then from (GE4) we have $\sigma(\mathbb{K}) = \mathbb{K}$ and hence $\sigma H \sigma^{-1}$ for all $\sigma \in G$. In other words, $H \trianglelefteq G$ is normal. Conversely, suppose that $H \trianglelefteq G$ is normal. Then we have $\sigma H \sigma^{-1} = H$ and hence $\mathrm{Gal}(\mathbb{E}/\sigma(\mathbb{K})) = \mathrm{Gal}(\mathbb{E}/\mathbb{K})$ for all $\sigma \in G$. We conclude from the bijection in part **(1)** that $\sigma(\mathbb{K}) = \mathbb{K}$ for all $\sigma \in G$.

Now since each $\sigma \in G$ restricts to an element of $\mathrm{Gal}(\mathbb{K}/\mathbb{F})$ we obtain a "restriction homomorphism" $\varphi : \mathrm{Gal}(\mathbb{E}/\mathbb{F}) \to \mathrm{Gal}(\mathbb{K}/\mathbb{F})$ with kernel $\mathrm{Gal}(\mathbb{K}/\mathbb{F}) = H$. Furthermore, since $\mathbb{E}/\mathbb{K}$ is a splitting field we know from the Splitting Field Theorem that each automorphism $\sigma : \mathbb{K} \to \mathbb{K}$ fixing $\mathbb{F}$ lifts to an automorphism $\hat{\sigma} : \mathbb{E} \to \mathbb{E}$. Hence the restriction homomorphism is **surjective** and we conclude from the First Isomorphism Theorem that

$$\frac{\mathrm{Gal}(\mathbb{E}/\mathbb{F})}{\mathrm{Gal}(\mathbb{E}/\mathbb{K})} = \frac{G}{H} = \frac{G}{\ker\varphi} \cong \mathrm{im}\,\varphi = \mathrm{Gal}(\mathbb{K}/\mathbb{F}).$$

Finally, from part **(2)** we have $\#\mathrm{Gal}(\mathbb{K}/\mathbb{F}) = \#(G/H) = [\mathbb{K}/\mathbb{F}]$ and it follows from (GE1) that $\mathbb{K} \supseteq \mathbb{F}$ is a Galois extension. □

Mathematical Remarks:

- Note that this proof was quite short because already did the hard work. The details are spread over three previous results: the Splitting Field Theorem, the Fixed Field Lemma and the Characterization Theorem for Galois Extensions.
- One surprising corollary of this theorem is that any finite-dimensional extension $\mathbb{E} \supseteq \mathbb{F}$ of perfect fields has **finitely many intermediate fields**. Indeed, if $\mathbb{E} \supseteq \mathbb{F}$ is not Galois then let $\mathbb{E}' \supseteq \mathbb{E}$ be a splitting field for some polynomial $f(x) \in \mathbb{F}[x]$. Then it follows from the Fundamental Theorem that that $\mathscr{L}(\mathbb{E}', \mathbb{F}) \cong \mathscr{L}(G)^{\mathrm{op}}$ where $G = \mathrm{Gal}(\mathbb{E}'/\mathbb{F})$. Since G is a finite group this implies that the lattice $\mathscr{L}(\mathbb{E}', \mathbb{F})$ is finite. Finally, since $\mathscr{L}(\mathbb{E}, \mathbb{F})$ is a subposet of $\mathscr{L}(\mathbb{E}', \mathbb{F})$ we conclude that $\mathscr{L}(\mathbb{E}, \mathbb{F})$ is also finite.
- If the field $\mathbb{E}$ is infinite then one can prove from the finiteness of $\mathscr{L}(\mathbb{E}, \mathbb{F})$ that there exists a primitive element $\mathbb{E} = \mathbb{F}(\gamma)$. Indeed, suppose that

$\mathbb{E}$ has finitely many maximal subfields over $\mathbb{F}$. Since each of these is a proper $\mathbb{F}$-subspace of $\mathbb{E}$ we conclude that there exists some $\gamma \in \mathbb{E}$ that is not in any maximal subfield.[166] Then since $\mathbb{F}(\gamma)$ is not contained in any maximal subfield we must have $\mathbb{F}(\gamma) = \mathbb{E}$.

- In fact, Steinitz (1910) proved that the existence of a primitive element is **equivalent** to the existence of only finitely many intermediate fields. But this equivalence is useless for us because it doesn't help us to prove either statement.

Historical Remarks:

- The Fundamental Theorem is a theorem of *Galois Theory*, but it is not *Galois' Theorem*. The original version of the theorem appears in Dedekind's 11th supplement (1894) to Dirichlet's *Vorlesungen über Zahlentheorie* (Lectures on Number Theory). According to Walther Purkert (1976), Dedekind had lectured on this material at Göttingen as early as 1856.
- The modern statement of the theorem for abstract fields (i.e., not just for subfields of $\mathbb{C}$) is due to Emil Artin in his Notre Dame lectures (1942).
- So what did Galois actually do? Recall from the introduction that his main concern was the solvability of polynomial equations with rational or integer coefficients. Next week we will return to this subject and we will apply the Fundamental Theorem to finally prove Galois' Solvability Theorem (in modern language). ///

For now let me show you a "toy example" of the Fundamental Theorem.

Example: Galois Theory of Finite Fields. We have seen that any finite field has the form $\mathbb{E} = \mathbb{F}_{p^k}$ where $\mathbb{F}_{p^k}$ is the splitting field of the polynomial $x^{p^k} - x \in \mathbb{F}_p[x]$. It follows that $\mathbb{F}_{p^k} \supseteq \mathbb{F}_p$ is a Galois extension of perfect fields. Furthermore, you will show on the homework that the Galois group is **cyclic** and generated by the **Frobenius automorphism**:

$$\begin{array}{rccc} \varphi: & \mathbb{F}_{p^k} & \to & \mathbb{F}_{p^k} \\ & \alpha & \mapsto & \alpha^p. \end{array}$$

In other words, you will show that

$$\mathrm{Gal}(\mathbb{F}_{p^k}/\mathbb{F}_p) = \langle\varphi\rangle = \{id, \varphi, \varphi^2, \ldots, \varphi^{k-1}\}.$$

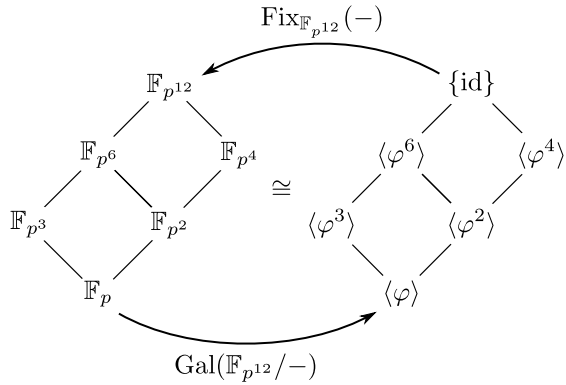
Then it follows from the Fundamental Theorems of Galois Theory and Cyclic Groups (which we proved early last semester) that the lattice of intermediate

[166] We are using the intuitively obvious fact that the complement of finitely many proper subspaces is not empty. I prefer not to prove this.

fields $\mathscr{L}(\mathbb{F}_{p^k}, \mathbb{F}_p)$ is isomorphic to the lattice of positive divisors $d|k$ of the integer k:

$$\begin{array}{ccccc} \mathscr{L}(\mathbb{F}_{p^k}, \mathbb{F}_p) & \cong & \mathscr{L}\langle\varphi\rangle^{\mathrm{op}} & \cong & \mathrm{Div}(k) \\ \mathbb{F}_{p^d} & \leftrightarrow & \langle\varphi^d\rangle & \leftrightarrow & d. \end{array}$$

Here is a picture for $k = 12$:



///

In hindsight, we see that the theory of finite fields is roughly as complicated as the theory of cyclic groups (i.e., not very). Galois studied finite fields in his paper *On the Theory of Numbers*, and this directly inspired his later work on the solvability of polynomial equations over $\mathbb{Q}$. The passage from finite fields to fields of characteristic zero is analogous to the passage from **cyclic groups** to **all finite groups**.[167] We should not expect it to be easy.

Exercises

22.A Cyclotomic Extensions are Abelian

Let $\mathbb{E} \supseteq \mathbb{F} \supseteq \mathbb{Q}$ and let $\omega \in \mathbb{E}$ be a primitive n-th root of unity. That is, assume that we have

$$x^n - 1 = (x-1)(x-\omega)\cdots(x-\omega^{n-1}) \text{ in } \mathbb{E}[x].$$

(a) Prove that $\mathbb{F}(\omega) \supseteq \mathbb{F}$ is a Galois extension.

(b) For all $\sigma \in \mathrm{Gal}(\mathbb{F}(\omega)/\mathbb{F})$ prove that $\sigma(\omega) = \omega^{k_\sigma}$ for some $\gcd(k_\sigma, n) = 1$.

(c) Prove that the map $\sigma \mapsto k_\sigma$ defines an injective group homomorphism

$$\mathrm{Gal}(\mathbb{F}(\omega)/\mathbb{F}) \hookrightarrow (\mathbb{Z}/n\mathbb{Z})^\times,$$

[167] The major open problem in Galois theory today is to establish whether or not every finite group G can be expressed in the form $G = \mathrm{Gal}(\mathbb{E}/\mathbb{Q})$. This is called the "inverse Galois problem". Shafarevich (1954) proved that every **solvable** group can be expressed in this way.

and hence $\mathrm{Gal}(\mathbb{F}(\omega)/\mathbb{F})$ is abelian.

(d) Let $\Phi_n(x) \in \mathbb{Q}[x]$ be the cyclotomic polynomial. Prove that

$$\mathrm{Gal}(\mathbb{F}(\omega)/\mathbb{F}) \cong (\mathbb{Z}/n\mathbb{Z})^\times \quad \Longleftrightarrow \quad \Phi_n(x) \text{ is irreducible in } \mathbb{F}[x].$$

22.B Radical Extensions are Abelian

Consider field extensions $\mathbb{E} \supseteq \mathbb{F}(\alpha) \supseteq \mathbb{F} \supseteq \mathbb{Q}$ where $\alpha^n \in \mathbb{F}$ for some $n \geq 2$ and suppose that $\mathbb{F}$ contains a primitive n-th root of unity.

(a) Prove that $\mathbb{F}(\alpha) \supseteq \mathbb{F}$ is a Galois extension.

(b) For any $\sigma \in \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ and $\beta \in \mathbb{F}(\alpha)$ prove that $\sigma(\beta) \in \mathbb{F}(\alpha)$.

(c) Prove that $\mathrm{Gal}(\mathbb{E}/\mathbb{F}(\alpha)) \subseteq \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ is a normal subgroup. [Hint: Use part (a) to define a group homomorphism $\mathrm{Gal}(\mathbb{E}/\mathbb{F}) \to \mathrm{Gal}(\mathbb{F}(\alpha)/\mathbb{F})$ with kernel $\mathrm{Gal}(\mathbb{E}/\mathbb{F}(\alpha))$.]

(d) Prove that the quotient group is abelian.

22.C Dedekind's Proof of the Irreducibility of $\Phi_n(x)$.

For any integer $n \geq 1$ recall that the *cyclotomic polynomial* is defined by

$$\Phi_n(x) = \prod_{\substack{0<k<n \\ \gcd(k,n)=1}} (x - \omega^k) \qquad \text{where } \omega = e^{2\pi i/n}.$$

You proved in Exercise 19.B that $\Phi_n(x) \in \mathbb{Z}[x]$. Now you will prove that $\Phi_n(x)$ is irreducible in the ring $\mathbb{Q}[x]$.[168] The following proof from van der Waerden's *Moderne Algebra* (1930) goes back to Dedekind.[169]

(a) Let $p \in \mathbb{Z}$ be prime and let $f(x) \mapsto f^\varphi(x)$ denote the ring homomorphism $\mathbb{Z}[x] \to \mathbb{Z}/p\mathbb{Z}[x]$ defined by reducing each coefficient mod p. Prove that for any $f(x) \in \mathbb{Z}[x]$ we have $f^\varphi(x^p) = f^\varphi(x)^p$. [Hint: The ring $\mathbb{Z}/p\mathbb{Z}[x]$ has characteristic p, hence it has a Frobenius endomorphism.]

(b) If $n \in \mathbb{Z}$ is not divisible by p, show that $x^n - 1$ has no repeated factor in $\mathbb{Z}/p\mathbb{Z}[x]$. [Hint: Any repeated factor is also a factor of the derivative.]

(c) Suppose that we can write $\Phi_n(x) = f(x)g(x)$ for some monic $f(x), g(x) \in \mathbb{Q}[x]$ with $f(x)$ irreducible. Use Gauss' Lemma to prove that $f(x)$ and $g(x)$ must have integer coefficients.

(d) Continuing from (c), suppose that we have $f(\omega^k) = 0 \Rightarrow f(\omega^{kp}) = 0$ for all $\gcd(k, n) = 1$ and for all primes $p \nmid n$. In this case prove that $f(x) = \Phi_n(x)$ and hence $\Phi_n(x)$ is irreducible. [Hint: Show that $f(\omega^\ell) = 0$ for all $\gcd(\ell, n) = 1$.]

---

[168] By Gauss' Lemma (Exercise 18.C) we also conclude that $\Phi_n(x)$ is irreducible in $\mathbb{Z}[x]$, but this fact is not very useful because $\mathbb{Z}[x]$ is not a PID.

[169] See Dedekind, *Beweis für die Irreductibilität der Kreisteilungs-Gleichungen* (1857). Gauss had proved the irreducibility of $\Phi_p(x)$ for prime p in the *Disquisitiones* (1801).

(e) Now we will show that the situation of part (d) must hold. To do this we assume for contradiction that there exists some integer $\gcd(k, n) = 1$ and prime $p \nmid n$ with $f(\omega^k) = 0$ and $g(\omega^{kp}) = 0$. In this case prove that $g(x^p) = f(x)h(x)$ for some $h(x) \in \mathbb{Z}[x]$. [Hint: Show that $f(x)$ is the minimal polynomial for ω^k over $\mathbb{Q}$.]

(f) Now it follows from (a) and (c) that $f^{\varphi}(x)h^{\varphi}(x) = g^{\varphi}(x)^p$ in $\mathbb{Z}/p\mathbb{Z}[x]$. Use the fact that $\mathbb{Z}/p\mathbb{Z}[x]$ is a UFD to prove that $f^{\varphi}(x)$ and $g^{\varphi}(x)$ have a common factor in $\mathbb{Z}/p\mathbb{Z}[x]$.

(g) Finally, use part (b) to obtain a contradiction.

Week 23: Epilogue

23.1 Radical Implies Solvable

We have come full circle. At the very beginning of this course I told you that Galois established a relationship between the "solvability of polynomial equations by radicals" and a certain structural property of abstract groups (which for this reason is called "solvability of groups"). Now we have (almost) all of the tools that we need to prove Galois' theorem.

However, let me warn you that you might find the result unsatisfying. To illustrate this, let's consider the case of Emil Artin, who — more than anyone — is responsible for the modern form of the subject. Here is a quote from a lecture he gave in 1950:

> *Since my mathematical youth I have been under the spell of the classical theory of Galois. This charm has forced me to return to it again and again, and to try to find new ways to prove its fundamental theorems.*[170]

However, in Artin's Notre Dame lectures (1942) which are considered his definitive statement on the subject, **he did not include a proof of the solvability theorem!** Instead, this theorem appears in an appendix[171] on "Applications", written by the American mathematician Arthur Milgram. It seems that in the preceding hundred years, the core of Galois theory had shifted from the "solvability theorem" to the "fundamental theorem", and that Milgram's appendix was included only as an accommodation to tradition.

It often happens in mathematics that the original motivation for a subject is discarded after we have discovered "what is really going on". But tradition still has pedagogical value.

So on to the Solvability Theorem. Let me recall the important definitions.

[170] Quoted in *The development of Galois Theory from Lagrange to Artin* (1971) by B. Melvin Kiernan.

[171] Technically, it is Part III.

Definition of Solvable Groups. We say that a finite group G is *solvable* if there exists a chain of subgroups

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_r = \{id\}$$

in which $G_i \trianglelefteq G_{i-1}$ is normal for all i and the quotient group G_{i-1}/G_i is abelian.

By inserting extra groups into the chain as necessary, we may assume without loss of generality that the that there does not exist any subgroup $G_{i-1} \supsetneq H \supsetneq G_i$ with $H \trianglelefteq G_{i-1}$ normal, which, by the Correspondence Theorem, is equivalent to assuming that the quotient groups G_{i-1}/G_i have no non-trivial normal subgroups. Finally, since each G_{i-1}/G_i is abelian, we may assume without loss of generality that $G_{i-1}/G_i \cong \mathbb{Z}/p_i\mathbb{Z}$ for some prime numbers $p_i \in \mathbb{Z}$. ///

Next let me recall Dedekind's algebraic version of "solvable by radicals".

Definition of Solvable Field Extensions. We say that a field extension $\mathbb{E} \supseteq \mathbb{F}$ is *solvable* if there exists a chain of field extensions

$$\mathbb{F} = \mathbb{F}_0 \subseteq \mathbb{F}_1 \subseteq \mathbb{F}_2 \subseteq \cdots \subseteq \mathbb{F}_r \supseteq \mathbb{E}$$

in which for all i we have $\mathbb{F}_i = \mathbb{F}_{i-1}(\alpha_i)$ for some element with $\alpha_i^{n_i} \in \mathbb{F}_{i-1}$. In the special case that $\mathbb{E}$ is the splitting field for a polynomial $f(x) \in \mathbb{F}[x]$ we say that that the equation $f(x) = 0$ is *solvable by radicals.* ///

Galois' Solvability Theorem. Consider a polynomial $f(x) \in \mathbb{F}[x]$ over a field $\mathbb{F}$ of characteristic zero, and let $\mathbb{E} \supseteq \mathbb{F}$ be a the splitting field. Then

$$f(x) = 0 \text{ is solvable by radicals} \iff \mathrm{Gal}(\mathbb{E}/\mathbb{F}) \text{ is a solvable group.}$$

///

Even though we have some rather powerful theorems at our disposal, the proof of this result is still trickier than one might guess. It is amazing how much effort is required to appreciate the the insights of an 18 year old who lived almost 200 years ago! Today we will prove that

$$f(x) = 0 \text{ is solvable by radicals} \implies \mathrm{Gal}(\mathbb{E}/\mathbb{F}) \text{ is a solvable group,}$$

and for this we still need a few lemmas.

Lemma (Quotient of a Solvable Group is Solvable). Let G be a solvable group and let $\varphi : G \to G'$ be a surjective group homomorphism. Then I claim that G' is solvable. It follows that any quotient group G/N is solvable since it is the image of the projection $G \to G/N$.

Proof. By assumption we have a chain of subgroups

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_r = \{id\}$$

where each quotient G_{i-1}/G_i exists and is abelian. Now apply φ to obtain a chain of subgroups

$$G' = G'_0 \supseteq G'_1 \supseteq G'_2 \supseteq \cdots \supseteq G'_r = \{id\},$$

where $G'_i \coloneqq \varphi[G_i]$ for all i. It remains to prove that each quotient G'_{i-1}/G'_i exists an is abelian. So consider any elements $\varphi(h) \in G'_i$ and $\varphi(g) \in G'_{i-1}$. Then since $G_i \trianglelefteq G_{i-1}$ is normal we have

$$\varphi(g)\varphi(h)\varphi(g)^{-1} = \varphi(ghg^{-1}) \in \varphi[G_i] = G'_i,$$

which implies that $G'_i \trianglelefteq G'_{i-1}$ is normal. Furthermore, I claim that the rule $\Phi(gG_i) \coloneqq \varphi(g)G'_i$ defines a (surjective) group homomorphism $\Phi : G_{i-1}/G_i \to G'_{i-1}/G'_i$. Indeed, we only need to check that this function is well-defined:

$$\begin{aligned} gG_i = hG_i &\implies h^{-1}g \in G_i \\ &\implies \varphi(h^{-1}g) \in G'_i \\ &\implies \varphi(h)^{-1}\varphi(g) \in G'_i \\ &\implies \varphi(g)G'_i = \varphi(h)G'_i. \end{aligned}$$

Finally, consider any two elements $\Phi(a), \Phi(b) \in G'_{i-1}/G'_i$. Since G_{i-1}/G_i is abelian we have

$$\Phi(a)\Phi(b) = \Phi(ab) = \Phi(ba) = \Phi(b)\Phi(a),$$

and hence G'_{i-1}/G'_i is abelian. □

The next two lemmas were proved by you on the previous homework. I will state them in exactly the form that we will use them.

Abelian Lemmas. Let $\mathbb{E} \supseteq \mathbb{F}$ be fields of chacteristic zero.

(1) For any root of unity $\omega \in \mathbb{E}$ the extension $\mathbb{F}(\omega)/\mathbb{F}$ is Galois with abelian Galois group.

(2) If $\mathbb{F}$ contains a primitive n-th root of unity and if $\alpha \in \mathbb{E}$ satisfies $\alpha^n \in \mathbb{F}$ then the extension $\mathbb{F}(\alpha)/\mathbb{F}$ is Galois with abelian Galois group.

Proof. Homework. □

Proof That Radical Implies Solvable. Let $\mathbb{E} \supseteq \mathbb{F} \supseteq \mathbb{Q}$ be the splitting field of a polynomial $f(x) \in \mathbb{F}[x]$ and assume that there exists a chain of radical extensions

$$\mathbb{F} = \mathbb{F}_0 \subseteq \mathbb{F}_1 \subseteq \mathbb{F}_2 \subseteq \cdots \subseteq \mathbb{F}_r \supseteq \mathbb{E}$$

where for each i we have $\mathbb{F}_i = \mathbb{F}_{i-1}(\alpha_i)$ for some element $\alpha_i \in \mathbb{F}_i$ with $\alpha_i^{n_i} = a_i \in \mathbb{F}_{i-1}$. Our goal is to construct a field $\mathbb{F}'_r \supseteq \mathbb{F}_r$ such that $\mathbb{F}'_r/\mathbb{F}$ is Galois and $\mathrm{Gal}(\mathbb{F}'_r/\mathbb{F})$ is a solvable group.[172] Then since $\mathbb{E} \supseteq \mathbb{F}$ (being a splitting field) is Galois we will conclude from the Fundamental Theorem and the Lemma on quotient groups that

$$\mathrm{Gal}(\mathbb{E}/\mathbb{F}) \cong \frac{\mathrm{Gal}(\mathbb{F}'_r/\mathbb{F})}{\mathrm{Gal}(\mathbb{F}'_r/\mathbb{E})} \quad \text{is also solvable.}$$

The difficulty has to do with the existence of enough roots of unity. I will follow Milgram's proof from the appendix of Artin's Notre Dame lectures. First let $\mathbb{F}'_0 \coloneqq \mathbb{F}_0 = \mathbb{F}$. Then let $\mathbb{F}'_1$ be the splitting field of the polynomial $f_1(x) \coloneqq x^{n_1} - a_1 \in \mathbb{F}[x]$ and observe that

- $\mathbb{F}'_1/\mathbb{F}$ is Galois,
- $\mathbb{F}'_1 \supseteq \mathbb{F}_1 = \mathbb{F}(\alpha_1)$,
- If ω_{n_1} is a primitive n_1-th root of unity then we observe that the splitting field contains α_1 and $\omega_{n_1}\alpha_1$, hence it also contains ω_{n_1}. Furthermore, we can get from $\mathbb{F} = \mathbb{F}'_0$ to $\mathbb{F}'_1$ by **first** adjoining ω_{n_1} and **then** adjoining α_1. From the Abelian Lemma we know that each of these extensions is Galois with abelian Galois group.

Next let $\mathbb{F}'_2$ be a splitting field for the following polynomial:

$$f_2(x) \coloneqq f_1(x) \cdot \prod_{\sigma \in \mathrm{Gal}(\mathbb{F}'_1/\mathbb{F})} (x^{n_2} - \sigma(a_2)) \in \mathbb{F}[x].^{173}$$

Observe that

- $\mathbb{F}'_2/\mathbb{F}$ is Galois,
- $\mathbb{F}'_2 \supseteq \mathbb{F}_2 = \mathbb{F}(\alpha_1, \alpha_2)$,
- Again we note that the splitting field contains a primitive n_2-th root of unity: $\omega_{n_2} \in \mathbb{F}'_2$. Then we can get from $\mathbb{F}'_1$ to $\mathbb{F}'_2$ by **first** adjoining ω_{n_2} and **then** adjoining (in any order) a primitive n_2-th root of each element $\sigma(a_2)$. Again we know from the Abelian Lemma that each of these extensions is Galois with abelian Galois group.

One more time. Let $\mathbb{F}'_3$ be the splitting field of

$$f_3(x) \coloneqq f_2(x) \cdot \prod_{\sigma \in \mathrm{Gal}(\mathbb{F}_3/\mathbb{F})} (x^{n_3} - \sigma(a_3)) \in \mathbb{F}[x].$$

[172] If $\mathbb{F}'_r/\mathbb{F}$ is Galois then to prove that $\mathrm{Gal}(\mathbb{F}'_r/\mathbb{F})$ is solvable it suffices by the Fundamental Theorem to show that we can get from $\mathbb{F}$ to $\mathbb{F}'_r$ by a sequence of Galois extensions, each of which has an abelian Galois group.

[173] This polynomial has coefficients in $\mathbb{F}$ because each coefficient is a symmetric polynomial in the elements $\{\sigma(a_2) : \sigma \in \mathrm{Gal}(\mathbb{F}'_1/\mathbb{F})\}$. But the elements of this set are permuted by the action of $\mathrm{Gal}(\mathbb{F}'_1/\mathbb{F})$, hence every coefficient is in the fixed field. Finally, since $\mathbb{F}'_1/\mathbb{F}$ is a Galois extension we know that the fixed field is $\mathbb{F}$.

For the same reasons as above we see that

- $\mathbb{F}_3'/\mathbb{F}$ is Galois,
- $\mathbb{F}_3' \supseteq \mathbb{F}_3 = \mathbb{F}(\alpha_1, \alpha_2, \alpha_3)$,
- We can get from $\mathbb{F}_2'$ to $\mathbb{F}_3'$ by **first** adjoining a primitive root ω_{n_3} and **then** adjoining (in any order) a primitive n_3-th root of each element $\sigma(a_3) \in \mathbb{F}_2'$. We know that each of these extensions is Galois with abelian Galois group.

By continuing in this way we will obtain a field extension $\mathbb{F}_r' \supseteq \mathbb{F}_r$ such that $\mathbb{F}_r'/\mathbb{F}$ is Galois and such that we can get from $\mathbb{F}$ to $\mathbb{F}_r'$ by a sequence of Galois extensions with abelian groups, hence the Galois group $\mathrm{Gal}(\mathbb{F}_r'/\mathbb{F})$ is solvable. □

Corollary. For $n \geq 5$ the general polynomial equation of degree n is not solvable by radicals.

Proof. We will prove below that the "general polynomial equation of degree n" has Galois group S_n. We proved last semester that this group is not solvable when $n \geq 5$. □

Of course, the unsolvabiity of the quintic was not an original discovery of Galois. It is generally attributed to Abel (1824) and Ruffini (1799), so is called the Abel-Ruffini Theorem. The original contribution of Galois was to explain precisely which equations are solvable and to provide a method by which one could (in principle, but not usually in practice) solve these equations. We will prove this next time.

23.2 Solvable Implies Radical

Today we will prove that any polynomial equation with a solvable Galois group is (in principle) solvable by radicals. For this we will need two more lemmas. The first is a straightforward translation of the Second Isomorphism Theorem for Groups into the language of field extensions. I will prove this at the maximum level of generality.

Lemma (The Second Isomorphism Theorem). Let $\mathbb{E} \supseteq \mathbb{F}$ be a finite-dimensional extension of perfect fields and consider any two intermediate fields $\mathbb{K}, \mathbb{L} \in \mathscr{L}(\mathbb{E}, \mathbb{F})$. If $\mathbb{L} \supseteq \mathbb{F}$ is Galois then $(\mathbb{KL}) \supseteq \mathbb{K}$ and $\mathbb{L} \supseteq (\mathbb{K} \cap \mathbb{L})$ are both Galois and we have

$$\mathrm{Gal}(\mathbb{KL}/\mathbb{K}) \cong \mathrm{Gal}(\mathbb{L}/\mathbb{K} \cap \mathbb{L}).$$

///

For the purpose of the proof we may assume that $\mathbb{E} \supseteq \mathbb{F}$ is a Galois extension, since otherwise we can enlarge $\mathbb{E}$ to a splitting field for some polynomial over $\mathbb{F}$. The proof will use the Fundamental Theorem of Galois Theory.

Proof. Let $G = \mathrm{Gal}(\mathbb{E}/\mathbb{F})$. Since $\mathbb{E} \supseteq \mathbb{F}$ is Galois we have the Galois correspondence:

$$\mathscr{L}(\mathbb{E}, \mathbb{F}) \cong \mathscr{L}(G)^{\mathrm{op}}.$$

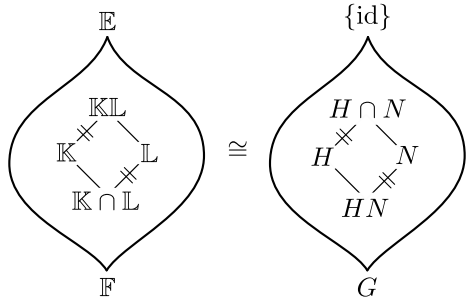
Now define $H = \mathrm{Gal}(\mathbb{E}/\mathbb{K})$ and $N = \mathrm{Gal}(\mathbb{E}/\mathbb{L})$. By assumption we know that $N \trianglelefteq G$ is normal. Since any isomorphism of posets preserves[174] meet and join we also have

$$H \cap N = \mathrm{Gal}(\mathbb{E}/\mathbb{K}\mathbb{L}) \quad \text{and} \quad HN = \mathrm{Gal}(\mathbb{E}/\mathbb{K} \cap \mathbb{L}).$$

Then since $(H \cap N) \trianglelefteq H$ and $N \trianglelefteq HN$ are normal subgroups we see that $\mathbb{K}\mathbb{L}/\mathbb{K}$ and $\mathbb{L}/(\mathbb{K} \cap \mathbb{L})$ are Galois extensions and it follows from the Second Isomorphism Theorem that

$$\mathrm{Gal}(\mathbb{K}\mathbb{L}/\mathbb{K}) \cong \frac{\mathrm{Gal}(\mathbb{E}/\mathbb{K})}{\mathrm{Gal}(\mathbb{E}/\mathbb{K}\mathbb{L})} = \frac{H}{H \cap N} \cong \frac{HN}{N} = \frac{\mathrm{Gal}(\mathbb{E}/\mathbb{K} \cap \mathbb{L})}{\mathrm{Gal}(\mathbb{E}/\mathbb{L})} \cong \mathrm{Gal}(\mathbb{L}/\mathbb{K} \cap \mathbb{L}).$$

Here is a picture:



□

The second lemma is similar in spirit to the Primitive Root Theorem and the Primitive Element Theorem. Today this result is regarded as part of "Kummer Theory", so we will call it "Kummer's Lemma".[175] However, the key idea of the proof goes back to Lagrange's 1770 work on algebraic equations.

Kummer's Lemma (Existence of Lagrange Resolvents). Let $\mathbb{E} \supseteq \mathbb{F}$ be a Galois extension of characteristic zero fields and let $[\mathbb{E}/\mathbb{F}] = p$ be prime. If

[174] Note that the meet and join in $\mathscr{L}(G)$ are flipped because we are using the opposite partial order.

[175] Ernst Eduard Kummer developed these ideas in the 1840s as part of his work on Fermat's Last Theorem.

$\mathbb{F}$ contains a primitive p-th root of unity $\omega \in \mathbb{F}$ then we can find some element $\alpha \in \mathbb{E} - \mathbb{F}$ such that $\alpha^p \in \mathbb{F}$ and $\mathbb{E} = \mathbb{F}(\alpha)$. We will call this element α a *Lagrange resolvent* for the extension $\mathbb{E}/\mathbb{F}$.

Proof. Let $G = \mathrm{Gal}(\mathbb{E}/\mathbb{F})$. Since $\#G = [\mathbb{E}/\mathbb{F}] = p$ is prime we know that $G = \{id, \sigma, \ldots, \sigma^{p-1}\}$ is cyclic. Furthermore, we know from Dedekind's Tower Law that $\mathbb{E} \supseteq \mathbb{F}$ has no nontrivial intermediate field. Our goal is to find some $\alpha \in \mathbb{E} - \mathbb{F}$ with $\sigma(\alpha^p) = \alpha^p$. Then $\alpha \notin \mathbb{F}$ implies that $\mathbb{E} = \mathbb{F}(\alpha)$ because there are no intermediate fields, and $\sigma(\alpha^p) = \alpha^p$ implies that $\alpha^p \in \mathbb{F}$ because σ generates G and because $\mathbb{F}$ is the fixed field of G. For fun, I will give two proofs: (1) an easy existence proof, (2) a tricky constructive proof.

(1) We have assumed that there exists a primitive p-th root of unity $\omega \in \mathbb{F}$. Thus we have $x^p - 1 = \prod_{k=0}^{p-1}(x - \omega^k)$ in $\mathbb{F}[x]$. Since powers of σ commute under composition we have an "evaluation homomorphism" from $\mathbb{F}[x]$ into the endomorphism ring $\mathrm{End}(\mathbb{E}/\mathbb{F})$[176] sending $x \mapsto \sigma$ and $1 \mapsto id$. Applying this to $x^p - 1$ gives $\prod_k^{p-1}(\sigma - \omega^k \cdot id) = \sigma^p - id = \mathbf{0}$, where the product on the left denotes composition of functions and $\mathbf{0}$ denotes the zero function. Since $\sigma \neq id$ there exists some $\beta \in \mathbb{E}$ with $\sigma(\beta) \neq \beta$ and hence $(\sigma - id)(\beta) \neq 0$. But note that

$$(\sigma - \omega^{p-1} \cdot id) \cdots (\sigma - \omega^2 \cdot id)(\sigma - \omega \cdot id)(\sigma - id)(\beta) = \mathbf{0}(\beta) = 0.$$

Let k be minimal such that $0 \neq (\sigma - \omega^k \cdot id) \cdots (\sigma - \omega \cdot id)(\sigma - id)(\beta)$ and call this nonzero element $\alpha \in \mathbb{E}$. By definition of k we have $(\sigma - \omega^{k+1} \cdot id)(\alpha) = 0$ and hence $\sigma(\alpha) = \omega^{k+1}\alpha \neq \alpha$. Since $\mathbb{F}$ is the fixed field of G this implies that $\alpha \notin \mathbb{F}$. Finally, note that

$$\sigma(\alpha^p) = \sigma(\alpha)^p = (\omega^{k+1})^p \alpha^p = (\omega^p)^{k+1}\alpha^p = \alpha^p.$$

(2) **Lagrange's Proof.** Choose any $\alpha \in \mathbb{E} - \mathbb{F}$ and for each $0 \leq j \leq p-1$ define the element

$$\alpha_j \coloneqq \sum_{i=0}^{p-1} \omega^{ij} \sigma^i(\alpha) \in \mathbb{E}.$$

Since $\omega \in \mathbb{F}$ we have for all $\sigma \in G$ that

$$\sigma(\alpha_j) = \sum_{i=0}^{p-1} \omega^{ij} \sigma^{i+1}(\alpha) = \omega^{-j} \sum_{i=0}^{p-1} \omega^{(i+1)j} \sigma^{i+1}(\alpha) = \omega^{-j}\alpha_j$$

which implies that $\sigma(\alpha_j^p) = \sigma(\alpha_j)^p = (\omega^{-j})^p \alpha_j^p = (\omega^p)^{-j}\alpha_j^p = \alpha_j^p$. It only remains to show that $\alpha_j \notin \mathbb{F}$ for some j. To prove this, we observe for all

[176] This is the **non-commutative** ring of $\mathbb{F}$-linear functions $\mathbb{E}/\mathbb{F} \to \mathbb{E}/\mathbb{F}$ under pointwise addition and composition. Note that we have a natural inclusion $\mathbb{F} \to \mathrm{End}(\mathbb{E}/\mathbb{F})$ defined by $a \mapsto a \cdot id$. Since the subring generated over $\mathbb{F}$ by a single element σ is **commutative**, the evaluation at σ is still a ring homomorphism.

$1 \leq i \leq p-1$ that ω^i is a primitive p-th root of unity and hence $1 + \omega^i + (\omega^i)^2 + \cdots + (\omega^i)^{p-1} = 0$. Then we have

$$\sum_{j=0}^{p-1} \alpha_j = \sum_{i,j=0}^{p-1} \omega^{ij} \sigma^i(\alpha) = \sum_{i=0}^{p-1} \sigma^i(\alpha) \sum_{j=0}^{p-1} (\omega^i)^j = \alpha^0(\alpha) \cdot p = p\alpha \notin \mathbb{F},$$

which implies that $\alpha_j \notin \mathbb{F}$ for some j. □

Proof that Solvable Implies Radical. Let $\mathbb{E} \supseteq \mathbb{F} \supseteq \mathbb{Q}$ be the splitting field for some polynomial $f(x) \in \mathbb{F}[x]$. Suppose that the Galois group $G = \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ is solvable. From the above definition this means that we have a chain of subgroups

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_r = \{id\}$$

where $G_i \trianglelefteq G_{i-1}$ is normal for all i and each quotient G_i/G_{i+1} is isomorphic to $\mathbb{Z}/p_i\mathbb{Z}$ for some prime number $p_i \in \mathbb{Z}$. Since $\mathbb{E} \supseteq \mathbb{F}$ is a Galois extension we can apply the Galois correspondence to obtain a chain of subfields

$$\mathbb{F} = \mathbb{F}_0 \subseteq \mathbb{F}_1 \subseteq \mathbb{F}_2 \subseteq \cdots \subseteq \mathbb{F}_r = \mathbb{E}$$

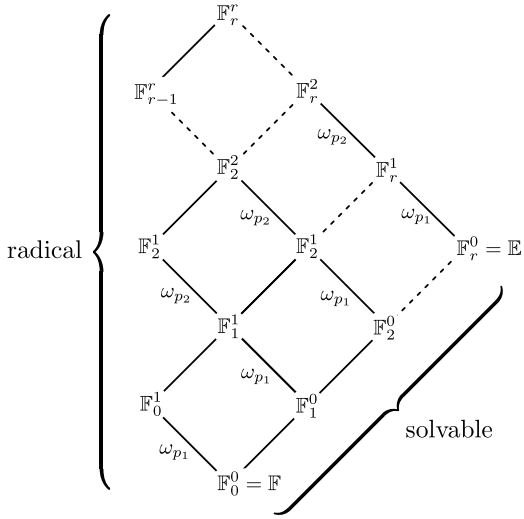
where $G_i = \mathrm{Gal}(\mathbb{E}/\mathbb{F}_i)$ for all i. Furthermore, since each subgroup $G_i \trianglelefteq G_{i-1}$ is normal we know from the Fundamental Theorem that each extension $\mathbb{F}_i \supseteq \mathbb{F}_{i-1}$ is Galois with

$$\mathrm{Gal}(\mathbb{F}_i/\mathbb{F}_{i-1}) \cong G_{i-1}/G_i \cong \mathbb{Z}/p_i\mathbb{Z}.$$

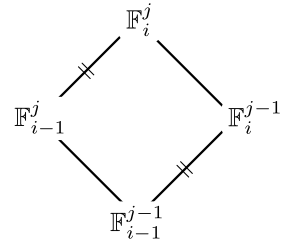
This chain of fields might not be radical, due to the fact that it might not contain enough roots of unity. To fix this situation I will describe a method to construct a "zig-zag chain" of radical field extensions containing $\mathbb{E}$:

$$\mathbb{F} = \mathbb{F}_0^0 \subseteq \mathbb{F}_0^1 \subseteq \mathbb{F}_1^1 \subseteq \mathbb{F}_1^2 \subseteq \mathbb{F}_2^2 \subseteq \cdots \subseteq \mathbb{F}_{r-1}^r \subseteq \mathbb{F}_r^r \supseteq \mathbb{E}$$

First let $\mathbb{F}_i^0 \coloneqq \mathbb{F}_i$ for all i. Then for all j let $\mathbb{F}_i^j \coloneqq \mathbb{F}_i^{j-1}(\omega_{p_j})$ where ω_{p_j} is a primitive p_j-th root of unity. Here is a picture:



By definition, each extension $\mathbb{F}_i^j \supseteq \mathbb{F}_i^{j-1}$ is either trivial or a radical extension generated by a root of unity $\omega_{p_j} \in \mathbb{F}_i^j - \mathbb{F}_i^{j-1}$ with $\omega_{p_j}^{p_j} = 1 \in \mathbb{F}_i^{j-1}$. Furthermore, for all indices i, j we observe that the following "diamond" satisfies the hypotheses of the Second Isomorphism Theorem:



Therefore we have $\mathrm{Gal}(\mathbb{F}_i^j/\mathbb{F}_{i-1}^j) \cong \mathrm{Gal}(\mathbb{F}_i^{j-1}/\mathbb{F}_{i-1}^{j-1})$ and by induction it follows that

$$\mathrm{Gal}(\mathbb{F}_i^i/\mathbb{F}_{i-1}^i) \cong \mathrm{Gal}(\mathbb{F}_i^0/\mathbb{F}_{i-1}^0) = \mathrm{Gal}(\mathbb{F}_i/\mathbb{F}_{i-1}) \cong \mathbb{Z}/p_i\mathbb{Z} \quad \text{for all } i.$$

Finally, since the field $\mathbb{F}_{i-1}^i = \mathbb{F}_{i-1}^{i-1}(\omega_{p_i})$ contains ω_{p_i} by construction, we conclude from Kummer's Lemma that the extension $\mathbb{F}_i^i \supseteq \mathbb{F}_{i-1}^i$ is radical. □

23.3 General Equations of Small Degree

That was it. To end the course I will show you how to apply Galois' theorem to the general polynomial equations of degrees $2, 3, 4$. But first, what is a "general polynomial equation"?

Definition/Theorem (The General Polynomial Equation). Let $\{x_1, \ldots, x_n\}$ be a set of variables representing the unknown roots of a general degree n polynomial over $\mathbb{Q}$. We will denote by $\mathbb{Q}(x_1, \ldots, x_n)$ the field of fractions of the ring of polynomials $\mathbb{Q}[x_1, \ldots, x_n]$ (which is an integral domain). To be explicit, we consider the set of formal fractions

$$\mathbb{E} = \mathbb{Q}(x_1, \ldots, x_n) \coloneqq \left\{ \frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)} : f, g \in \mathbb{Q}[x_1, \ldots, x_n] \text{ and } g \neq 0 \right\}.$$

with respect to the equivalence relation $f/g = f'/g' \Leftrightarrow fg' = f'g$. We know from a previous homework that this set is a field with respect to the obvious operations. Now consisider the *elementary symmetric polynomials* $e_1, e_2, \ldots, e_n \in \mathbb{Q}[x_1, x_2, \ldots, x_n]$ defined by

$$f(x) = x^n - e_1 x^{n-1} + e_2 x^{n-2} - \cdots + (-1)^n e_n = (x - x_1)(x - x_2) \cdots (x - x_n)$$

and let $\mathbb{F} \coloneqq \mathbb{Q}(e_1, \ldots, e_n) \subseteq \mathbb{E}$ be the smallest subfield containing these polynomials. Then clearly $\mathbb{E} \supseteq \mathbb{F}$ is a splitting field of $f(x) \in \mathbb{F}[x]$ and hence $\mathbb{E}/\mathbb{F}$ is a finite-dimensional Galois extension. Furthermore, since any element of the group $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$

- permutes the variables $x_1, \ldots, x_n$ (i.e., the roots of $f(x)$), and
- is determined by its action on the variables $x_1, \ldots, x_n$ (i.e., the generators of $\mathbb{E}/\mathbb{F}$),

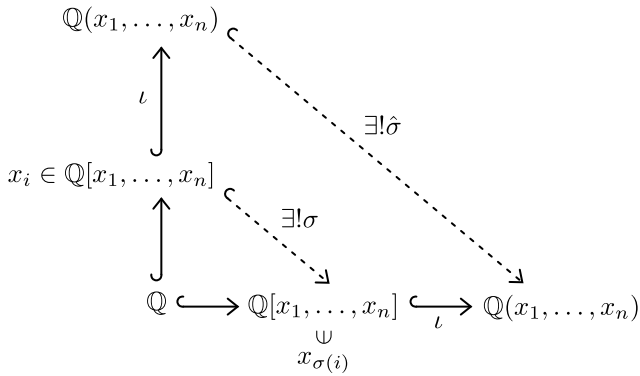
we obtain an injective group homomorphism $\mathrm{Gal}(\mathbb{E}/\mathbb{F}) \hookrightarrow S_n$ into the group of permutations of the variables. I claim that this homomorphism is also **surjective**, and hence

$$\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = \mathrm{Gal}(\mathbb{Q}(x_1, \ldots, x_n)/\mathbb{Q}(e_1, \ldots, e_n)) \cong S_n.$$

///

Proof. We need to show that every permutation $\sigma \in S_n$ of the variables $\{x_1, \ldots, x_n\}$ extends to a field automorphism $\hat{\sigma} : \mathbb{Q}(x_1, \ldots, x_n) \to \mathbb{Q}(x_1, \ldots, x_n)$ that fixes the subfield $\mathbb{Q}(e_1, \ldots, e_n)$.

First, we will prove the existence of $\hat{\sigma}$ by messing around with universal properties. For any permutation $\sigma \in S_n$ we know from the universal property of polynomials that the inclusion $\mathbb{Q} \hookrightarrow \mathbb{Q}[x_1, \ldots, x_n]$ extends to a unique ring homomorphism $\sigma : \mathbb{Q}[x_1, \ldots, x_n] \to \mathbb{Q}[x_1, \ldots, x_n]$ fixing $\mathbb{Q}$ and sending $x_i \mapsto x_{\sigma(i)}$ for all i. One can check that this homomorphism is **injective**.[177] Next, consider the inclusion $\iota : \mathbb{Q}[x_1, \ldots, x_n] \hookrightarrow \mathbb{Q}(x_1, \ldots, x_n)$ of the domain $\mathbb{Q}[x_1, \ldots, x_n]$ into its field of fractions. Then since $\iota \circ \sigma$ is an **injective** homomorphism from a domain to a field, we know from the universal property of fractions that there exists a unique extension $\hat{\sigma} : \mathbb{Q}(x_1, \ldots, x_n) \to \mathbb{Q}(x_1, \ldots, x_n)$ satisfying $\hat{\sigma} \circ \iota = \iota \circ \sigma$. Here is a picture:



We only need to show that the endomorphism $\hat{\sigma} : \mathbb{Q}(x_1, \ldots, x_n) \to \mathbb{Q}(x_1, \ldots, x_n)$ is invertible. To see this we will prove that if $\sigma, \mu \in S_n$ are inverse permutations then $\hat{\sigma}, \hat{\mu}$ are inverse endomorphisms, and hence automorphisms of $\mathbb{Q}(x_1, \ldots, x_n)$. Indeed, we have $\hat{\sigma} \circ \iota = \iota \circ \sigma$ and $\hat{\mu} \circ \iota = \iota \circ \mu$ by definition. But

[177] In other words, the variables $\{x_1, \ldots, x_n\}$ are *algebraically independent* over $\mathbb{Q}$.

then

$$(\hat{\sigma}\circ\hat{\mu})\circ\iota = \hat{\sigma}\circ(\hat{\mu}\circ\iota) = \hat{\sigma}\circ(\iota\circ\mu) = (\hat{\sigma}\circ\iota)\circ\mu = (\iota\circ\sigma)\circ\mu = \iota\circ(\sigma\circ\mu) = \iota\circ id$$

implies by uniqueness that $\hat{\sigma}\circ\hat{\mu} = \hat{id} = id$. For the same reason we have $\hat{\mu}\circ\hat{\sigma} = id$.

Next we need to show that each group element $\hat{\sigma}$ fixes the subfield $\mathbb{Q}(e_1,\ldots,e_n)$. Clearly we have $\hat{\sigma}(e_i) = e_i$ for each elementary symmetric polynomial, and hence $\hat{\sigma}(f(e_1,\ldots,e_n))$ for each polynomial $f(x_1,\ldots,x_n) \in \mathbb{Q}[x_1,\ldots,x_n]$. If the polynomials e_i were algebraic over $\mathbb{Q}$ then we would be done. Since they are not, we need one more step. We observe that

$$\mathbb{Q}(e_1,\ldots,e_n) = \left\{ \frac{f(e_1,\ldots,e_n)}{g(e_1,\ldots,e_n)} : f,g \in \mathbb{Q}[x_1,\ldots,x_n] \text{ and } g(e_1,\ldots,e_n) \neq 0 \right\}.^{178}$$

Indeed, the set on the right is a subfield of $\mathbb{Q}(x_1,\ldots,x_n)$ containing the elements $e_1,\ldots,e_n$, hence it contains the smallest such subfield. Conversely, since every element of the set on the right can be formed from the set $\mathbb{Q}\cup\{e_1,\ldots,e_n\}$ using field operations, we see that this set is contained in $\mathbb{Q}(e_1,\ldots,e_n)$. Finally, we conclude that every element of this field is fixed:

$$\hat{\sigma}\left(\frac{f(e_1,\ldots,e_n)}{g(e_1,\ldots,e_n)}\right) = \frac{\sigma(f(e_1,\ldots,e_n))}{\sigma(g(e_1,\ldots,e_n))} = \frac{f(e_1,\ldots,e_n)}{g(e_1,\ldots,e_n)}.$$

□

Remarks:

- This finally completes our proof that the general polynomial equation of degree $n \geq 5$ is not solvable by radicals.
- With a bit of extra work, one can use this result to give a non-constructive proof of Waring's Theorem. Here's a sketch: Since $\mathbb{E}/\mathbb{F}$ is a Galois extension with Galois group $S_n = \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ we know from condition (GE2) of the Characterization Theorem that $\mathbb{F} = \mathrm{Fix}_{\mathbb{E}}(S_n)$. Thus for any symmetric polynomial $f(x_1,\ldots,x_n) \in \mathrm{Fix}_{\mathbb{E}}(S_n)$ we must have $f(x_1,\ldots,x_n) \in \mathbb{F} = \mathbb{Q}(e_1,\ldots,e_n)$ and hence

  $$f(x_1,\ldots,x_n) = \frac{g(e_1,\ldots,e_n)}{h(e_1,\ldots,e_n)} \text{ for some } g,h \in \mathbb{Q}[x_1,\ldots,x_n].$$

  Finally, one can argue[179] that the denominator is constant, hence f is a polynomial in the elementary symmetric polynomials. ///

[178] In his second proof of the Fundamental Theorem of Algebra, Gauss proved that $g(x_1,\ldots,x_n) \neq 0$ implies $g(e_1,\ldots,e_n) \neq 0$. In other words, the elementary symmetric polynomials are *algebraically independent* over $\mathbb{Q}$.

[179] This is the hardest part. It involves the concept of "integral elements" of a ring extension, which is a generalization of "algebraic elements". This topic is more suitable for a graduate course.

Example: The General Quadratic. Let $\mathbb{E} = \mathbb{Q}(x_1, x_2)$ and $\mathbb{F} = \mathbb{Q}(e_1, e_2)$, so $\mathbb{E} \supseteq \mathbb{F}$ is the splitting field of the general quadratic polynomial

$$f(x) = x^2 - e_1 x + e_2 = (x - x_1)(x - x_2) \in \mathbb{F}[x].$$

Since $[\mathbb{E}/\mathbb{F}] = 2$ and since $\mathbb{F}$ contains a primitive 2-nd root of unity (namely, $-1 \in \mathbb{F}$) then we know from Kummer's Lemma that there exists an element $\gamma \in \mathbb{E} - \mathbb{F}$ with $\gamma^2 \in \mathbb{F}$ and $\mathbb{E} = \mathbb{F}(\gamma)$. Furthermore, note that $\sigma = (12)$ is a generator of $\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = S_2 = \{id, (12)\}$. Thus for any $\alpha \in \mathbb{E} - \mathbb{F}$ we know from Lagrange's proof that at least one of the following two elements is a resolvent:

$$\alpha_1 = \alpha + \sigma(\alpha),$$
$$\alpha_2 = \alpha - \sigma(\alpha).$$

In fact, we know that α_1 is **not** a resolvent because $\alpha(\alpha_1) = \alpha_1$ implies that α_1 is in the fixed field $\mathbb{F}$. Thus α_2 is **always** a resolvent. For simplicity, let's take $\alpha = x_1$ so that $\alpha_1 = x_1 + x_2$ and $\alpha_2 = x_1 - x_2$ is a resolvent. To be specific, we have

$$\alpha_2^2 = (x_1 - x_2)^2 = e_1^2 - 4e_2 \in \mathbb{F},$$

and then each of x_1 and x_2 is guaranteed to have the form $a + b\alpha_2 = a + b\sqrt{e_1^2 - 4e_2}$ for some $a, b \in \mathbb{F}$. With a bit of thought we find that

$$x_1 = (\alpha_1 + \alpha_2)/2 = (e_1 + \sqrt{e_1^2 - 4e_2})/2,$$
$$x_2 = (\alpha_1 - \alpha_2)/2 = (e_2 - \sqrt{e_1^2 - 4e_2})/2.$$

///

Example: The General Cubic. Let $\mathbb{E} = \mathbb{Q}(x_1, x_2, x_3)$ and $\mathbb{F} = \mathbb{Q}(e_1, e_2, e_3)$ so that $\mathbb{E} \supseteq \mathbb{F}$ is the splitting field of the general cubic polynomial

$$f(x) = x^3 - e_1 x^2 + e_2 x - e_3 = (x - x_1)(x - x_2)(x - x_3) \in \mathbb{F}[x].$$

From the above theorem we also have $\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = S_3$. Recall that S_3 is a **solvable** group with composition series

$$S_3 \supsetneq A_3 \supsetneq \{id\}.$$

Explicitly, $A_3 \subseteq S_3$ is the cyclic subgroup generated by the 3-cycle (123). Now apply the Galois correspondence to obtain a chain of field extensions

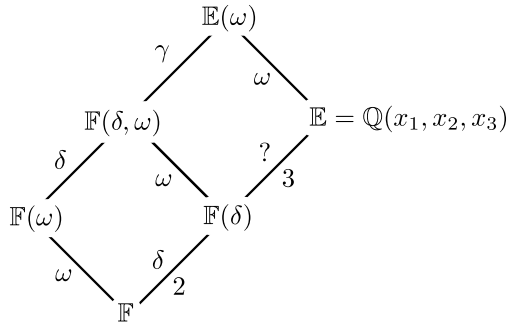
$$\mathbb{F} \subsetneq \mathbb{K} \subsetneq \mathbb{E},$$

with $[\mathbb{K}/\mathbb{F}] = 2$, $[\mathbb{E}/\mathbb{K}] = 3$ and $\mathrm{Gal}(\mathbb{E}/\mathbb{K}) = A_3$. Next I claim that $\mathbb{K} = \mathbb{F}(\delta)$, where

$$\delta := (x_1 - x_2)(x_1 - x_3)(x_2 - x_3).$$

Indeed, we have $\delta \in \mathbb{K} - \mathbb{F}$ because δ is fixed by the alternating group A_3 but not by the full symmetric group S_3. And we have $\delta^2 \in \mathbb{F}$ because δ^2 is fixed by S_3.[180]

Unfortunately, the extension $\mathbb{E} \supseteq \mathbb{K} = \mathbb{F}(\delta)$ is **not** radical. To fix this, let $\omega^2 + \omega + 1 = 0$ be a primitive third root of unity and adjoin ω to every field in the chain:



Now since $\mathbb{E}(\omega) \supseteq \mathbb{F}(\delta,\omega)$ is a Galois extension of (prime) degree 3 which contains a primitive 3-rd root of unity, Kummer's Lemma guarantees that there exists a Lagrange resolvent $\gamma \in \mathbb{E}(\omega)$ with $\mathbb{E}(\omega) = \mathbb{F}(\delta,\omega,\gamma)$ and $\gamma^3 \in \mathbb{F}(\delta,\omega)$. To be explicit, consider the generator $\sigma = (123)$ of the Galois group $A_3 = \mathrm{Gal}(\mathbb{E}(\omega)/\mathbb{F}(\delta,\omega))$. Then for any element $\alpha \in \mathbb{E}(\omega) - \mathbb{F}(\delta,\omega)$ we know that at least one of the following elements[181] is a Lagrange resolvent:

$$\alpha_2 = \alpha + \omega\sigma(\alpha) + \omega^2\sigma^2(\alpha),$$
$$\alpha_3 = \alpha + \omega^2\sigma(\alpha) + \omega\sigma^2(\alpha).$$

To simplify things, let's take $\alpha = x_1$ so the two potential Lagrange resolvents become

$$\alpha_2 = x_1 + \omega x_2 + \omega^2 x_3 \quad \text{and} \quad \alpha_3 = x_1 + \omega^2 x_2 + \omega x_3.$$

Since α_2^3 and α_3^3 are elements of $\mathbb{F}(\delta,\omega)$ and since $\mathbb{F}(\delta,\omega) \supseteq \mathbb{F}(\omega)$ has degree 2, we are guaranteed that each of α_2^3 and α_3^3 is a root of a quadratic equation with coefficients in $\mathbb{F}(\omega)$. In fact, the choice $\alpha = x_1$ is particularly nice because it turns out that α_2^3 and α_3^3 are both roots of a certain quadratic polynomial with coefficients in $\mathbb{F}$. The rest of the details are called "Cardano's Formula", which we discussed at the beginning of last semester. ///

Example: The General Quartic. Let $\mathbb{E} = \mathbb{Q}(x_1,x_2,x_3,x_4)$ and $\mathbb{F} =$

[180] Recall that δ^2 is called the *discriminant* of the polynomial $f(x)$. On a previous homework you showed that

$$\delta^2 = e_1^2e_2^2 - 4e_2^3 - 4e_1^3e_3 + 18e_1e_2e_3 - 27e_3^2.$$

[181] Again, the element $\alpha_1 = \alpha + \sigma(\alpha) + \sigma^2(\alpha)$ is **not** a resolvent because $\sigma(\alpha_1) = \alpha_1$ implies that α_1 is in the fixed field $\mathbb{F}(\delta,\omega)$.

$\mathbb{Q}(e_1, e_2, e_3, e_4)$ so that $\mathbb{E} \supseteq \mathbb{F}$ is the splitting field of the general quartic polynomial

$$f(x) = x^4 - e_1x^3 + e_2x^2 - e_3x + e_4 = (x - x_1)(x - x_2)(x - x_3)(x - x_4) \in \mathbb{F}[x]$$

with Galois group $S_4 = \mathrm{Gal}(\mathbb{E}/\mathbb{F})$. Since $4! = 24$ is still a small number, it is a lucky accident that the group S_4 is solvable. To be explicit, we have the following composition series:

$$S_4 \supsetneq A_4 \supsetneq V_4 \supsetneq \langle (12)(34) \rangle \supsetneq \{id\}.$$

Here V_4 is the *Kleinsche Vierergruppe*:[182]

$$V_4 = \{id, (12)(34), (13)(24), (14)(23)\}.$$

For the same reason as above, the fixed field of the subgroup A_4 is $\mathrm{Fix}_{\mathbb{E}}(A_4) = \mathbb{F}(\delta)$, where

$$\delta := (x_1 - x_2)(x_1 - x_3)(x_1 - x_4)(x_2 - x_3)(x_2 - x_4)(x_3 - x_4)$$

is "the square root of the discriminant" $\delta^2 \in \mathbb{F}$.[183] Now apply the Galois correspondence to obtain a chain of fields

$$\mathbb{F} \subsetneq \mathbb{F}(\delta) \subsetneq \mathbb{K} \subsetneq \mathbb{L} \subsetneq \mathbb{E},$$

where $\mathbb{K} = \mathrm{Fix}_{\mathbb{E}}(V_4)$ and $\mathbb{L} = \mathrm{Fix}_{\mathbb{E}}(\langle (12)(34) \rangle)$. With a bit of thought, one can show that

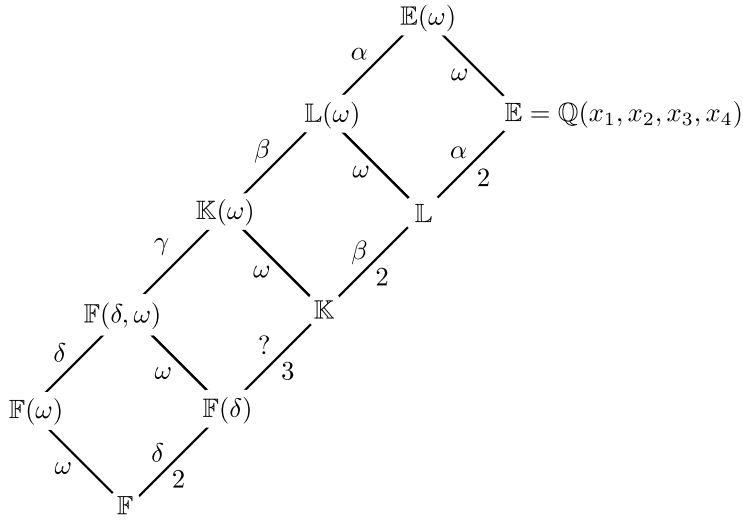
$$\mathbb{L} = \mathbb{Q}(x_1 + x_2, x_1x_2, x_3 + x_4, x_3x_4),$$
$$\mathbb{K} = \mathbb{Q}(x_1x_2 + x_3x_4, x_1x_3 + x_2x_4, x_1x_4 + x_2x_3).$$

Since $[\mathbb{E}/\mathbb{L}] = 2$ and $[\mathbb{L}/\mathbb{K}] = 2$, we are guaranteed that each of these extensions is radical. However, since $[\mathbb{K}/\mathbb{F}(\delta)] = 3$ and since $\mathbb{F}(\delta)$ does not contain a primitive 3rd root of unity,[184] this extension is **not** radical. Thus we should adjoin a primitive root $\omega^2 + \omega + 1 = 0$ to obtain the following diagram:

[182] *Klein's four-group* is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, and hence is the smallest non-cyclic group. G. A. Miller tells us but according to Miller (*Group theory in the history of mathematics*, 1938) the term was bestowed by "various German writers" in honor of Felix Klein, who used the term "Vierergruppe" in his work. I think *Ruffini's four-group* is a better name, since Paolo Ruffini (1799) was the first to apply the group to the solvability of the quartic.

[183] Believe me, you do not want to see the explicit formula for δ^2 in terms of the coefficients e_1, e_2, e_3, e_4.

[184] for reasons of degree



The rest of the solution follows from the quadratic and cubic cases. First choose any element of $\mathbb{E}(\omega) - \mathbb{L}(\omega)$; for example x_1. Then since $\sigma = (12)(34)$ generates the group $\mathrm{Gal}(\mathbb{E}(\omega)/\mathbb{L}(\omega))$ we know that $\alpha \coloneqq x_1 - \sigma(x_1) = x_1 - x_2$ is a resolvent. Next choose any element of $\mathbb{L}(\omega) - \mathbb{K}(\omega)$; for example $x_1 + x_2$. Then since (the coset of) $\sigma = (13)(24)$ generates the group $\mathrm{Gal}(\mathbb{L}(\omega)/\mathbb{K}(\omega))$ we know that $\beta \coloneqq x_1 + x_2 - \sigma(x_1 + x_2) = x_1 + x_2 - x_3 - x_4$ is a resolvent. Finally, we need to choose an element of $\mathbb{K}(\omega) - \mathbb{F}(\delta, \omega)$; for example $x_1x_2 + x_3x_4$. Then since (the coset of) $\sigma = (123)$ generates the group $\mathrm{Gal}(\mathbb{K}(\omega)/\mathbb{F}(\delta, \omega) = A_4/V_4$ we know that

$$\begin{aligned}\gamma &= (x_1x_2 + x_3x_4) + \omega\sigma(x_1x_2 + x_3x_4) + \omega^2\sigma^2(x_1x_2 + x_3x_4) \\ &= (x_1x_2 + x_3x_4) + \omega(x_1x_4 + x_2x_3) + \omega^2(x_1x_3 + x_2x_3)\end{aligned}$$

is a resolvent. From this recipe it is possible to find explicit radical formulas for the roots x_1, x_2, x_3, x_4 in terms of the coefficients e_1, e_2, e_3, e_4, but what would be the point? The full solution will certainly not fit on a page.[185] ///

Galois knew that he had achieved a complete conceptual understanding of the solvability of polynomial equations. But he also knew that this understanding was mostly useless because the solutions are too complicated to write down. I will end this course by quoting Galois on this issue. The following excerpt is from the preface to a planned pair of manuscripts. Galois wrote this in prison in December 1832. He was released in April 1832, and died in May. The corrections and modifications are copied from the handwritten original:[186]

[185] Here I chose only the most obvious resolvents. The history of the quartic equation is filled with more elegant choices. But even the most beautiful version of the "quartic formula" will still not fit on a page.

[186] Quoted from Dossier 11 in *The mathematical writings of Évartiste Galois* (2011) by Peter M. Neumann.

Long algebraic calculations were at first hardly necessary for progress in Mathematics; the very simple theorems hardly gained from being translated into the language of analysis. It is only since Euler that this briefer language has become indispensable to the new extensions which this great geometer has given to science. Since Euler calculations have become more and more necessary but more and more ~~complicated~~ difficult, at least insofar as they are applied to the most advanced objects of science. Since the beginning of this century algorithmics had attained such a degree of complication that any progress had become impossible by these means, ~~except~~ without the elegance with which new modern geometers have believed they should imprint their research, and by means of which the mind promptly and with a single glance grasps a large number of operations.

It is clear that such vaunted elegance, and so properly claimed, has no other goal. From the well established fact that the efforts of the most advanced geometers have elegance as their object, ~~it follows that~~ ~~we have come to~~ ~~science has come to~~ ~~on this point~~ one may therefore ~~deduce~~ conclude with certainty ~~that the further the research of~~ ~~one advances, the more it is~~ that it becomes more and more necessary to embrace several operations ~~at a single glance~~ at once ~~in other words~~ because ~~the less~~ the mind does not have the time any more to stop ~~at each~~ at details.

Thus I believe that the simplifications produced by elegance of calculations (intellectual simplifications, of course; there are no material ones) have their limits; I believe that the time will come when the ~~calculations~~ algebraic transformations foreseen by the speculations of analysts will find neither the time nor the place for their realisation; at which point one will have to be content with having foreseen them.

~~*That is, according to me, the mission of future geometers; that is the path that I have entered.*~~ *I would not wish to say that there is nothing new for analysis without this rescue; but I believe that without this one day all will run out.* ~~*Embrace*~~ *Jump with both feet on calculations.* ~~*embrace*~~ *put operations into groups,* ~~*distinguish*~~ *class them according to their difficulty and not according to their form; that is, according to me, the mission of future geometers, that is the path that I have entered in this work.*

References

Abel, Niels Henrik
(1826) "Démonstration de l'impossibilité de la résolution algébrique des équations générales qui passent le quatrième degré", in the first volume of *Crelle's Journal*.

Artin, Emil
(1942) *Galois Theory*.

Artin, Michael
(1991) *Algebra*.

Asghari, Amir (remove)
(2018) "Equivalence: an attempt at a history of the idea", *Synthese*.

Bergman, George
(1998) *An Invitation to General Algebra and Universal Constructions*.

Bézout, Etienne
(1779) *Théorie générale des équations algébriques*.

Brook, Melissa and Macfarlane, J.A.
(2020) "Radical Solutions", in *Damn Interesting*:
`https://www.damninteresting.com/radical-solutions/`

Cardano, Gerolamo
(1545) *Ars Magna*.

Carmichael, Robert Daniel
(1921) Review of the lecture "Some famous problems of the theory of numbers and in particular Waring's problem", given by G. H. Hardy. *Bulletin of the American Mathematical Society*.

Cauchy, Augustin-Louis
(1815) "Sur le nombre des valeurs qu'une fonction peut acquérir lorsqu'on y permute de toutes les manières possibles les quantités qu'elle renferme", *Journal de l'École polytechnique*.
(1844) "Mémoire sur les arrangements que l'on peut former avec des lettres données et sur les permutations ou substitutions à l'aide desquelles on passe d'un arrangement à un autre", *Exercises d'analyse et de physique mathétique*.

Cayley, Arthur
(1854) "On the theory of groups, as depending on the symbolic equation $\theta^n = 1$", *Philosophical Magazine.*

Corry, Leo
(2004) *Modern algebra and the rise of mathematical structures.*

Dedekind, Richard
(1894) "Über die Theorie der ganzen algebraischen Zahlen", in Dirichlet (1894).
(1932) *Gesammelte mathematische Werke, Dritter Band.* The fragment "Äquivalenz von Gruppen", (page 440) shows that Dedekind understood the First Isomorphism Theorem in the 1850s.

de Moivre, Abraham
(1707) "Aequationum Quarundam Potestatis Tertiae, Quintae, Septimae, Nonae, & Superiorum, ad Infinitum Usque Pergendo, in Terminis Finitis, ad Instar Regularum pro Cubicis Quae Vocantur Cardani, Resolutio Analytica", *Philosophical Transactions.*

Dirichlet, Gustav Lejeune
(1894) *Vorlesungen über Zahlentheorie* (4th edition), edited and with supplements by Richard Dedekind.

Euler, Leonhard
(1748) *Introductio in analysin infinitorum.*

Ferreirós, José (remove)
(2007) *Laryrinth of Thought: A History of Set Theory and Its Role in Modern Mathematics.*

Frobenius, Georg
(1887) *Neuer Beweis des Sylowschen Satzes.*

Gauss, Carl Friedrich
(1801) *Disquisitiones Arithmeticae.* Written in 1798 when he was 21.

Gray, Jeremy
(2018) *A History of Abstract Algebra: From Algebraic Equations to Modern Algebra.*

Hasse, Helmut (remove)
(1826) *Höhere Algebra.*

Hölder, Otto
(1889) "Zurückfuhrung einer belieben algebraischen Gleichung auf eine Kette von Gleichungen", *Mathematische Annalen.*

Huntington, Edward Vermilye
(1902) "Simplified definition of a group", *Bulletin of the American Mathematical Society.*

Jordan, Camille
(1870) *Traité des substitutions et des équations algébriques.*

(1873) "Sur la limite de transitivité des groupes non alternés", *Bulletin de la Société Mathématique de France.*

al-Khwarizmi
(∼820) *The Compendious Book on Calculation by Completion and Balancing.*

Kiernan, B. Melvin
(1971) "The development of Galois theory from Lagrange to Artin", *Archive for History of Exact Sciences.*

Lagrange, Joseph-Louis
(1770) "Réflexions sur la résolution algébrique des équations", *Œuvres complètes.*

Laubenbacher, Reinhard and David Pengelley
(1999) *Mathematical Expeditions: Chronicles by the Explorers.*

Liouville, Joseph
(1846) *Oeuvres Mathématiques d'Evariste Galois.*

Lützen, Jesper
(1990) *Joseph Liouville 1809–1882: Master of Pure and Applied Mathematics.*

Mac Lane, Saunders
(1971) *Categories for the Working Mathematician.*

Neumann, Olaf
(2010) "The *Disquisitiones Arithmeticae* and the Theory of Equations", in *The Shaping of Arithmetic after C. F. Gauss's Disquisitiones Arithmeticae*, edited by Goldstein et al.

Nicholson, Julia
(1993) "The Development and Understanding of the Concept of Quotient Group", *Historia Mathematica.*

Noether, Emmy
(1929) "Hypercomplexe Größen und Darstellungstheorien", *Mathematische Zeitschrift*

Ore, Oystein
(1944) "Galois Connexions", *Transactions of the American Mathematical Society.*

Pontryagin, Lev Semyonovich
(1931) "Über den algebraischen Inhalt der topologischen Dualitätssätze", *Mathematische Annalen.*

Schröder, Ernst
(1890) *Vorlesungen über die Algebra der Logic.*

Serret, Joseph
(1849) *Cours d'algèbre supérieure.*

Steinitz, Ernst
(1910) "Algebraische Theorie der Körper", *Crelle's Journal.*

(1913) "Bedingt konvergente Reihen und konvexe Systeme", *Crelle's Journal*. Contains the "Steinitz Exchange Lemma".

Stillwell, John
(2001) *Elements of Algebra: Geometry, Numbers, Equations.*
(2008) *Naive Lie Theory.*

Sylow, Ludwig
(1872) *Théorèmes sur les groupes de substitutions.*

Tignol, Jean-Pierre
(2001) *Galois' Theory of Algebraic Equations.*

van der Waerden, Bartel Leendert
(1930) *Moderne Algebra.*
(1975) "On the sources of my book *Moderne Algebra*", *Historia Mathematica.*

Weber, Heinrich
(1895) *Lehrbuch der Algebra.*

Weil, André
(1984) *Number Theory: An approach through history from Hammurapi to Legendre.*

Weyl, Hermann
(1939) *The Classical Groups: Their Invariants and Representations.*

Wielandt, Helmut
(1959) *Ein Beweis für die Existenz der Sylowgruppen.*

Wussing, Hans
(1984) *The Genesis of the Abstract Group Concept: A Contribution to the History of the Origin of Abstract Group Theory.*

Ziegler, Francois
(2017) `https://mathoverflow.net/questions/194377/when-was-the-arrow-notation-for-functions-first-introduced`
(2019) `https://mathoverflow.net/questions/330638/history-of-the-kernel-of-a-homomorphism`

Index