Lecture notes on Abstract Algebra

Uli Walther

Version of Spring 2021

Contents

Expected progress:

- Week 1: Archimedes, Factorization
- Week 2: Symmetries, Groups, Subgroups, Order, Aut(G)
- Week 3: $\mathbb{Z}/n\mathbb{Z}$, Products, $U(n)$
- Week 5: Cosets, Morphisms
- Week 4: Symmetric and Free Group
- Week 6: Normal Subgroups, Quotients, Automorphism Theorem
- Week 7: Finitely Generated Abelian Groups
- Week 8: Review, Group Actions
- Week 9: Reading day, Intro to Rings
- Week 10: Midterm, Ideals, Morphisms
- Week 11: Euclidean Algorithm, PID, UFD
- Week 12: Fields, Eisenstein, Extensions,
- Week 13: Degrees and Splitting Fields
- Week 14: Minimal Polynomials, Finite Fields
- Week 15: Galois Outlook, Review

Basic notions

0.1. How to use these notes. These notes contain all I say in class, plus on occasion a lot more. If there are exercises in this text, you may do them but there is no credit, and you need not turn them in. All exercises that are due are specifically listed on gradescope.

This initial chapter is here so we have acommon understanding of the basic symbols and words. This should be known from MA375 (at least if I teach it).

Chapter 1 is still more than we did in week 1, but almost all of it should be familiar, and the rest (the open problems on primes) is for your entertainment. Future chapters correspond to actual weeks of classes and are much less verbose than the Basic Notions.

The chapter "Stuff for later" can be ignored. It will be worked in when the time comes.

REMARK .1. There are typos in these notes. If you find some, please inform me.

0.2. Set lingo. The mathematical word *set* denotes what in colloquial life would be called a collection of things. The "things" are in mathematics referred to as *elements* of the set. If S is a set and s is an element of S then one writes $s \in S$.

A sub-collection S' of elements of S is a *subset* and one writes $S' \subseteq S$, allowing for the possibility of S' being all of S, or to have no element. There is, strangely, a set that contains nothing. It's called the *e*mpty set (denoted $\varnothing$) and, despite its humble content, one of the most important sets. One uses the notation $S = \{s_1, \ldots, s_n, \ldots\}$ to indicate that S consists of exactly the elements s_i. (In many cases, one must allow the index set to be different from $\mathbb{N}$. In other words, not all sets can be "numbered", a fact we explore a bit below).

A *function* $\phi\colon A \to B$ from the set A to the set B is an assignment (think: a black box) that turns elements of A into elements of B. The crucial conditions are: the assignment works for every single input $a \in A$ (so the black box does not choke on any input from A) and for each input there is exactly one output specified (no more, no less). Graphically, functions are often depicted by the help of arrows (starting at the various elements of A and ending at the value $\phi(a)$ for each input a). (For an example, suppose $\phi\colon \mathbb{Z} \to \mathbb{Z}$ is the process of doubling. Then one could write $\phi(3) = 6$, or $3 \mapsto 6$). The set A is usually referred to as *source*, the set B as *target*.

DEFINITION .2. The function $\phi\colon A \to B$ is

(1) *surjective* ("onto") if every element b appears as output of ϕ;
(2) *injective* ("into") if the equality of outputs $\phi(a_1) = \phi(a_2)$ occurs exactly when the inputs a_1 and a_2 were equal;
(3) *bijective* if it is injective and surjective.

An injective map is indicated as $A \hookrightarrow B$, a surjective one by $A \twoheadrightarrow B$.

For example, the function $\phi(x) = x^3$ is a bijection from $A = \mathbb{R}$ to $B = \mathbb{R}$ (all real numbers have exactly one cubic root); the function $\phi(x) = x^2$ is neither injective (since always $\phi(x) = \phi(-x)$) nor surjective (since negative numbers have no real roots).

We will often say "map" instead of "function".

0.3. Size of sets. We wish to attach to each set S a size denoted $|S|$. In order to make sense of this, we need to compare sets by size.

DEFINITION .3. We write $|S| \leq |S'|$ if there is an injective map $\phi\colon S \hookrightarrow S'$.

Do not confuse the symbols $\leq$ and $\subseteq$. The following examples illustrate the nature of the relation $\leq$.

EXAMPLE .4.
$|\mathbb{N}| \leq |\mathbb{Z}|$ since each natural number is an integer. ◇

EXERCISE .5. Show that $|\mathbb{Z}| \leq |\mathbb{N}|$. ◇

EXAMPLE .6.
- $|\mathbb{Z}| \leq |\mathbb{Q}|$ since each integer is a rational number.
- $|\mathbb{Q}| \leq |\mathbb{R}|$ since each rational number is also real.
- Somewhat shockingly, $|\mathbb{Q}| \leq |\mathbb{Z}|$. To see this, it will be sufficient to prove that there is a way of labeling the rational numbers with integer labels. (One can then make an injective map that sends each rational to its label). How does one label? Imagine sorting the rational positive numbers into a two-way infinite table, as follows:

p \ q	1	2	3	4	$\cdots$
1	1/1	1/2	1/3	1/4	$\cdots$
2	2/1	2/2	2/3	2/4	$\cdots$
3	3/1	3/2	3/3	3/4	$\cdots$
4	4/1	4/2	4/3	4/4	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Clearly all positive rationals appear (multiple times) in the table. Now suppose you are moving through the table "on diagonals" where $p + q$ is constant: start at $1/1$, the only square on its diagonal (where $p + q = 2$). Next go on the diagonal $p + q = 3$, starting on the square with $1/2$ and then moving down and to the left. Next walk along the diagonal $p + q = 4$ starting on $1/3$ and moving down and left. It is clear that this process allows you to label each field: $1/1$ is number 1, $1/2$ is number 2, $2/1$ is number 3, and so on. So, the set of all squares is in bijection with the set $\mathbb{N}$. Since all positive rationals are sorted into the various fields, it follows that $|\mathbb{Q}| \leq |\{\text{all squares}\}| \leq |\mathbb{N}|$. A similar idea can be used on negative numbers, and this shows that $|\mathbb{Q}| \leq |\mathbb{Z}|$.

- In contrast, the statement $|\mathbb{R}| \leq |\mathbb{Q}|$ is false. The idea is due Cantor, and goes like this. If you believe that you can inject $\mathbb{R}$ into $\mathbb{Q}$ then you can also inject $\mathbb{R}$ into $\mathbb{Z}$ because $|\mathbb{Q}| \leq |\mathbb{Z}|$. Since $|\mathbb{Z}| \leq |\mathbb{N}|$, this also implies that you can inject $\mathbb{R}$ into $\mathbb{N}$. To inject $\mathbb{R}$ into $\mathbb{N}$ means to label the real numbers by using only natural (non-repeated) indices. In particular, this can be done to the reals between 0 and 1.

Suppose we have an exhaustive enumeration $(0,1) = \{r_0, r_1, r_2, \ldots\}$ of all real numbers in the unit interval. Let $r_{i,j}$ be the j-th digit in the decimal expansion of r_i. So, r_i is the real number with expansion $0.r_{i,1}r_{i,2}r_{i,3}\ldots$. Now write the real numbers into a two-way infinite table:

i \ j	1	2	3	4	$\cdots$
1	$r_{1,1}$	$r_{1,2}$	$r_{1,3}$	$r_{1,4}$	$\cdots$
2	$r_{2,1}$	$r_{2,2}$	$r_{2,3}$	$r_{2,4}$	$\cdots$
3	$r_{3,1}$	$r_{3,2}$	$r_{3,3}$	$r_{3,4}$	$\cdots$
4	$r_{4,1}$	$r_{4,2}$	$r_{4,3}$	$r_{4,4}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

We construct now the real number ρ whose decimal expansion is determined as follows: the i-th decimal of ρ is the i-th decimal of r_i. So ρ is "the diagonal". Finally, concoct a new real number σ whose i-th decimal is: 1 if $\rho_i = 3$; 3 if $\rho_i \neq 3$.

The point is that by looking at the i-th decimal, it is clear that σ is not r_i (as they don't agree in that position). So, σ is not on our list. So, one cannot make a list (indexed by $\mathbb{N}$) that contains all real numbers. In particular, there are seriously more reals than rationals or integers.

One can (and for example Cantor did) try to determine whether there are sets S such that $|\mathbb{Q}| \leq |S| \leq |\mathbb{R}|$ but neither $|S| \leq |\mathbb{Q}|$ nor $|\mathbb{R}| \leq |S|$. So the question is whether there is a set that is between $\mathbb{R}$ and $\mathbb{Q}$ but one cannot inject $\mathbb{R}$ into S and also not inject S into $\mathbb{Q}$. That there is no such set is called the *continuum hypothesis*. As it has turned out through fundamental work of Gödel and Cohen, this question cannot be answered within the framework of the axioms of Zermelo and Fraenkel.[1] (Gödel proved that no matter what system of axioms you take, it is either self-contradictory or allows unanswerable questions. Cohen showed that, in particular, the continuum hypothesis cannot be decided with Zermelo–Fraenkel's axioms). ◇

0.4. Finite vs infinite. Some sets S allow injections into themselves that are not surjective. For example, one can make a function $\phi\colon \mathbb{N} \to \mathbb{N}$ that sends x to $x+1$ and so is clearly injective but not onto. Such sets are called *infinite*. A set for which every injection $\phi\colon S \to S$ has to be also surjective is *finite*.

Finite sets allow to attach a familiar quantity to S, by answering the question "what is the smallest n such that $|S| \leq |\{1, 2, \ldots, n\}|$. One writes $|S| = n$ in that case and calls it the *cardinality*, although we will still call it the *size* of S. For infinite sets, one needs new symbols since the size of such set will not be a natural number. One writes $|\mathbb{N}| = \aleph_0$ (this thing is pronounced "aleph" and denotes the size of the smallest set that is not finite) and $|\mathbb{R}| = \aleph_1$. While we can't answer the question whether there is something between $\aleph_0$ and $\aleph_1$, it is known that there is no upper limit to sizes because of the following construction.

EXAMPLE .7. The *power set* of a set S is the collection of all subsets of S, denoted 2^S. This power set includes the empty set $\varnothing$ and the whole set S as special cases. By the exercise below, if S is finite, then the size of the power set is given

[1]All mathematical sentences we use are built from basic axioms laid down by Zermelo and Fraenkel. For a somewhat digestible description of the axioms and the surrounding issues, see `http://en.wikipedia.org/wiki/Zermelo-Fraenkel_set_theory`

by $|2^S| = 2^{|S|}$. If S is infinite, such equation makes no sense. But in any event, 2^S is strictly larger than S in the sense that there is no injection $2^S \hookrightarrow S$. The idea of the proof is the same as the Cantor diagonal trick for $S = \mathbb{N}$ we saw above. ⋄

EXERCISE .8. If the set S is finite, prove that $|2^S| = 2^{|S|}$. (Hint: an element of 2^S is a subset of S. What question do you need to answer for each element of S when you form a subset of S? How many possible answers can you get?) ⋄

EXERCISE .9. Let S be a finite set of size n. Determine (in terms of n) the number of pairs of sets (A, B) where both A and B are subsets of S, and where no element of S is both in A and B. Prove the formula you find.

So, for example, if S has one element called s, then the options for (A, B) are: $(\varnothing, \varnothing)$, $(\varnothing, \{s\})$ and $(\{s\}, \varnothing)$. ⋄

EXERCISE .10. Let S be a finite set of size n as in the previous exercise. We consider all pairs of sets $C \subseteq D$ where $D \subseteq S$. Show that the number of such pairs is the same as the numbers of pairs (A, B) from the previous exercise. ⋄

0.5. Inclusion/Exclusion.

NOTATION .11. Given two sets A and B, their *union* is the set $A \cup B$ that contains any element in A, any element in B, and no other. On the other hand, the *intersection* $A \cap B$ is the set that contains exactly those elements of A that are also in B, and no other.

For a list of sets $A_1, \ldots, A_k$ their common intersection is denoted $\bigcap_{i=1}^k A_i$ and their union $\bigcup_{i=1}^k A_i$.

Suppose A and B are two finite sets; we want to know the size of their union $A \cup B$. A first order approximation would be $|A| + |B|$, but this is likely to be off because A and B might have overlap and elements in the overlap $A \cap B$ would be counted twice, once in A and once in B. So, we must correct the count by removing one copy of each element in the overlap:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

How about three sets? In that case, there are three intersections: $A \cap B$, $B \cap C$ and $A \cap C$, whose sizes should presumably all be removed from $|A| + |B| + |C|$. This is the right idea but doesn't quite capture it. For example, if $A = \{1, 2, 3\}$, $B = \{3, 4\}$ and $C = \{2, 3, 5\}$ then $|A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C|$ is $3 + 2 + 3 - 1 - 1 - 2 = 4$ while the union is the set $\{1, 2, 3, 4, 5\}$. To understand what happened, look at each element separately. The expression above counts each of $1, 2, 4, 5$ a total of once. But the element 3 is counted three times, and then removed three times. So, the count is off by one. Inspection shows that this error will always happen if the intersection $A \cap B \cap C$ is not empty, and the count will be off by as many elements as this intersection contains. So, we conclude:

$$\begin{aligned} |A \cup B \cup C| &= |A| + |B| + |C| \\ &\quad -|A \cap B| - |A \cap C| - |B \cap C| \\ &\quad +|A \cap B \cap C|. \end{aligned}$$

It is clear then what the general pattern is:

THEOREM .12 (Inclusion/Exclusion Formula). *For any n finite sets* $A_1, \ldots, A_n$,

$$\begin{aligned} |\bigcup_{1\le i\le n} A_i| \quad = \quad & \sum_{i=1}^{n} |A_i| \\ & - \sum_{1\le i_1<i_2\le n} |A_{i_1} \cap A_{i_2}| \\ & + \sum_{1\le i_1<i_2<i_3\le n} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| \\ & \cdots \\ & +(-1)^{n+1}|A_1 \cap A_2 \cap \ldots \cap A_n|. \end{aligned}$$

REMARK .13. In the special case where the sets A_i are pairwise disjoint (all pairwise intersections $A_i \cap A_j$ are empty) then the formula just says: the size of the union of disjoint sets is the sum of the separate sizes. ◇

EXERCISE .14. Let $S = \{1, 2, \ldots, 666\}$. Determine the number of elements of S that are

(1) divisible by 3
(2) divisible by 7 (Achtung!)
(3) divisible by 3 or 2 or 37 ("or" means "divisible by at least one of them")
(4) divisible by 6 and 4
(5) divisible by 6 or 4
(6) divisible by 3 and 37 but not by 2
(7) divisible by 6 or 4 but not by 9.

◇

CHAPTER I

Week 1: Introduction

NOTATION. The following symbols will be used throughout to denote number sets:

- the *natural numbers* $0, 1, 2, 3, \ldots$ are denoted by $\mathbb{N}$;
- the *integer numbers* $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$ are denoted by $\mathbb{Z}$;
- the *rational numbers* (all p/q with p, q in $\mathbb{Z}$ and $q \neq 0$) are denoted by $\mathbb{Q}$;
- the *real numbers* are denoted by $\mathbb{R}$;
- the *complex numbers* are denoted $\mathbb{C}$.

If S is a *set* (compare Section 0.2 and the surrounding discussion), and s is an element of S then we shorthand this to $s \in S$. If a small set S is contained in a big set B we write $S \subseteq B$. If we want to stress that S does not fill all of B we write $S \subsetneq B$. If a set S is not actually contained in another set B we write $S \not\subseteq B$. (Note the logical and notational difference of the last two!)

We denote by $|S|$ the *size* of the set S, which is explained in Section 0.3.

If n is a natural number we will have need to differentiate between a set $\{a_1, \ldots, a_n\}$ and an *ordered n-tuple* $(a_1, \ldots, a_n)$. The difference is that when we use round brackets, it is important in what order the numbers come. It is also possible that entries are repeated in an ordered n-tuple. The ordered n-tuples are in one-to-one correspondence with the points in n-space. There is also a hybrid: a *family* $a_1, \ldots, a_n$ has no emphasis on order, but it can have repeated entries.

If $S = \{a_s\}_{s \in S}$ is a set whose elements are numbers, we write $\sum_{s \in S} a_s$ for the sum of all elements in S, and $\prod_{s \in S} a_s$ for their product. In the extreme case where S is empty, the sum over S is (by agreement) equal to zero, and the product equal to 1.

1. Induction

Suppose you are faced with the task of proving that, for all natural numbers n, the sum $1 + 2 + \ldots + n$ equals $n(n+1)/2$. A few tests show that the formula is probably right, but no matter how many checks you do, there are infinitely many others yet to run. It seems like a hopeless proposition. Mathematical induction is a tool that allows you to complete precisely this sort of job.

1.1. Setup. Suppose that, for each $n \in \mathbb{N}$, there is given a statement $P(n)$ that involves the symbol n more or less explicitly. (For example, $P(n)$ could be the statement "the sum $1 + 2 + \ldots + n$ equals $n(n+1)/2$" from above).

The task at hand is to prove that *all* statements $P(0), P(1), \ldots$ are *true*.

1.2. The idea. Imagine you are standing in front of a ladder that starts at your feet (level 0) and goes up indefinitely. Your job is to convince your friend that you are capable of climbing up to any step of the ladder. How might you do that?

One approach is to check that you can indeed make it to the lowest rung of the ladder (the "base case") and then to exhibit a kind of cranking mechanism that allows for any position on the ladder (no matter which exact one), say rung $n+1$, to find another rung that is *lower*, such that you can move to rung $n+1$ from the lower one.

If you can do these two things then clearly you can make it to any desired level. This is what induction does: imagine that the n-th step of the ladder symbolizes proving statement $P(n)$. The "base case" means that you should check explicitly the lowest n for which the statement $P(n)$ makes sense, and the "crank" requires you to provide a logical argument that says "If $P(k)$ is true for all $k \leq n$ then $P(n+1)$ is also true". This "crank" is called the *inductive step* where the part "If $P(k)$ is true for all $k \leq n$" is known as the *inductive hypothesis*. The "base case" is the *induction basis*.

REMARK I.1. In many cases, you will only use $P(n)$ in order to prove $P(n+1)$, but there are exceptions where using only $P(n)$ is not convenient. Some people call usage of all $P(i)$ with $i \leq n$ "strong induction". But there is nothing strong about this sort of induction: one can show that what can be proved with strong induction can also be proved if you just assume $P(n)$ for the sake of proving $P(n+1)$. ⋄

EXAMPLE I.2. We consider the question from the start of the section: show that $0+1+\ldots+n = n(n+1)/2$. So, for $n \in \mathbb{N}$ we let the statement $P(n)$ be "$0+1+\ldots+n = n(n+1)/2$".

The base case would be $n = 0$ or $n = 1$, depending on your taste. In either case the given statement is correct: if $n = 0$ then the sum on the left is the empty sum (nothing is being added) and that means (by default) that the sum is zero. Of course, so is $0(0+1)/2$. One might be more sympathetic towards the case $n = 1$ in which the purported identity becomes $1 = 1(2)/2$, clearly correct.

For the crank, one needs a way to convince other people that if one believes in the equation

$$P(n): \qquad 1+2+\ldots+n = n(n+1)/2$$

then one should also believe in the equation

$$P(n+1): \qquad 1+2+\ldots+n+(n+1) = (n+1)(n+1+1)/2.$$

In induction proofs for equational statements like this it is usually best to compare the left hand side (LHS) of the presumed and the desired equality and to show that their difference (or quotient, as the case may be) is the same as those of the right hand sides (RHS). In other words, one tries to manufacture the new equation from the old.

In the case at hand, the difference of the LHSs is visibly $n+1$. The RHS difference is $(n+1)(n+2)/2 - n(n+1)/2 = (n+2-n)(n+1)/2 = 2(n+1)/2 = n+1$. So, if one believes in the equation given by $P(n)$ then, upon adding $n+1$ on both sides, one is forced to admit that equation $P(n+1)$ must also be true. This completes the crank and the principle of induction asserts now that all statements $P(n)$ are true, simply because $P(0)$ is and because one can move from any $P(n)$ to the next "higher" one via the crank. ⋄

REMARK I.3. For the functionality of induction it is imperative that both the base case and the crank are in order. (It's clear that without crank there is not

much hope, but the checking of the base case is equally important, even if the crank has already been established!)

Consider for example the following attempt of proving that $1+2+\ldots+n = n(n+1)/2+6$. Let's write $P'(n)$ to be the statement "$1+2+\ldots+n = n(n+1)/2+6$". Now argue as follows: suppose that for some $n \in \mathbb{N}$, $P'(n)$ is true: $1+2+\ldots+n = n(n+1)/2+6$. Add $n+1$ on both sides to obtain $1+2+\ldots+n+(n+1) = n(n+1)/2+6+n+1 = [n(n+1)+2(n+1)]/2+6 = (n+1)(n+2)/2+6$. So, truth of $P'(n)$ implies truth of $P'(n+1)$.

Of course, if you believe that we did the right thing in Example I.2 above, then $P'(n)$ can't hold ever (unless you postulate $6=0$). The problem with climbing the P'-ladder is that while we have a crank that would move us from any step to the next step up, we never ever actually *are* on any step: the base case failed! ⋄

REMARK I.4. The usual principle of induction only works with collections of statements that are labeled by the natural numbers. If your statements involve labels that are not natural numbers then, typically, induction cannot be used indiscriminately.

One can make various errors in induction proofs. Indicated here are two, by way of an incorrect proof.

(1) "Theorem": all horses have the same color.

Proof by induction: let $P(n)$ ($n \in \mathbb{N}$) be the statement "within any group of n horses, all horses have the same color". The base case $P(0)$ is void (there is no horse to talk about, so $P(0)$ is true) and $P(1)$ is clearly true as well.

Now suppose $P(n)$ is true and we prove $P(n+1)$ from that. It means that we must show that in any group of $n+1$ horses all horses have the same color. So let S be a group of $n+1$ horses, which we name $H_1, H_2, \ldots, H_{n+1}$. Let T_1 stand for the size n group of the first n horses, $T_1 = \{H_1, \ldots, H_n\}$. Let T_2 stand for the last n horses, $T_2 = \{H_2, \ldots, H_{n+1}\}$. Since T_1 has n horses in it, statement $P(n)$ kicks in and says that all horses inside T_1 have the same color, which we denote by c_1. Similarly, all horses in group T_2 (of size n) have all one color, called c_2. However, the horses $H_2, \ldots, H_n$ appear in both sets, and so have colors c_1 and c_2 simultaneously. We conclude $c_1 = c_2$ and so all horses in S had the same color!

(2) "Theorem": Let a be any positive real number. Then, for all $n \in \mathbb{N}$, one has $a^n = 1$.

Proof by induction: let $P(n)$ be "$a^n = 1$". The base case is $n = 0$. In that case, $a^0 = a^{1-1} = a^1/a^1 = 1$.

Now assume that $P(i)$ is true for all $0, \ldots, n$.. We want to show $a^{n+1} = 1$. Rewrite: $a^{n+1} = a^n \frac{a^n}{a^{n-1}}$. Both "$a^n = 1$" (statement $P(n)$) and "$a^{n-1} = 1$" (statement $P(n-1)$) are covered by the inductive hypothesis and so $P(n+1)$ must be true.

Both proofs imply wrong results, so they can't really be proofs. What's the problem? The errors are not of the same type, although similar. In the first case, we use the collection $H_2, \ldots, H_n$ of horses that are simultaneously in both T_1 and T_2. The problem is that if $n = 1$ then there aren't any such horses: T_1 is just $\{H_1\}$ and T_2 is just $\{H_2\}$. So there is actually no horse that can be used to compare colors in group T_1 with those in group T_2, and so c_1 and c_2 have no reason to be equal.

In the second proof, you were sneakily made to believe that "$a^{n-1} = 1$" is covered by the inductive hypothesis, by calling it "$P(n-1)$". But if $n = 0$ then $n-1$ is negative, and we are not entitled to use negative indices on our statement!!! One must be very careful not to feed values of n outside the set of naturals into an inductive argument. ◇

1.3. Archimedean property and well-order. Here is a fundamental property of the natural numbers:

$\mathbb{N}$ is well-ordered.

What that means is this: every subset of $\mathbb{N}$, as soon as it has any element at all, will have a *minimal* element. So in particular, our set B of bad numbers has a smallest element: there is a first n for which $P(n)$ is false. Call this smallest bad guy b.

This is a notable property because S might have infinitely many elements. Any *finite* set has a minimum sor sure, because you can test one pair at a time. But for infinite sets this is not an option. And not all infinite sets have a minimum, just look at the open interval $(0, 1)$.

Remark I.5. The second example in the emark underscores an important point: induction only works well for the index set $\mathbb{N}$. What's so special about the naturals? Let's go back to the drawing board of induction. The idea is (rephrased): if $P(n)$ ever failed, let B be the bad indices: n is in B exactly if $P(n)$ is false.

Question: what could this b be? Answer: surely not $b = 0$ since the base case requires us explicitly to check that $P(0)$ is true. So, the minimal bad b is positive. Since it's positive, $b-1$ is natural (not just integer, but actually not negative. Since b was the minimal bad guy, $P(0), P(1), \ldots, P(b-1)$ can't be false, so they are all true. And now comes the kicker: since we do have a crank, $P(b)$ must also be true! It follows, that we were mistaken: the little bad b never existed, and the claims $P(n)$ are all true.

Could one hope for proofs of induction when the index set is something different from $\mathbb{N}$? Not so much. Some thought reveals that making inductive proofs is the same as the index set being well-ordered. But not many sets *are* well-ordered. For example, $\mathbb{Z}$ is not (it has no smallest element—one could not meaningfully speak of a lowest rung on the $\mathbb{Z}$-ladder). Also, the set of real numbers in the closed interval $[0, 1]$ is not well-ordered (for example, the subset given by the half-open interval $(0, 1]$ doesn't have a smallest element—this says that there is no real number that "follows" 0). So, induction with index set $\mathbb{Z}$ or $\mathbb{R}$ or $[0, 1]$ is not on the table. ◇

Remark I.6. One can formally turn induction "upside down". The purpose would be to prove that all statements in an infinite sequence $P(n)$ are false. The idea is: check that $P(0)$ is false explicitly; then provide a crank that shows: if some statement $P(n)$ is true then there must already be an earlier statement $P(i)$ with $i < n$ that is true.

This reverse method is called *infinite descent* and illustrated in the next example. ◇

Well-order of the natural numbers is a very basic property and closely related to the following "obvious" result:

THEOREM I.7 (Archimedean property). *Choose $a, b \in \mathbb{N}$ with $0 < a$. Then the sequence $a, a+a, a+a+a, \ldots$ contains an element that exceeds b, so that $b - (a + \ldots + a) < 0$. In other words, $\exists k \in \mathbb{N}$ with $ka > b$.* □

I call this a theorem because if one wrote down the axiomatc definition of $\mathbb{N}$ then this property is one that one needs to prove from the axioms. This axiomatic definiton, translated into English, says roughly that there is a natural number 0, and another natural number 1, and for each natural number a there is another one called $a+1$, and there aren't any other natural numbers than those you can make this way. And one says $a < b$ if b can be made from a by iterating the procedure $a \mapsto a+1$.

It is not always true that collecting lots of small things (like a) gives you something big (like b). For example, adding lots of polynomials of degree 3 does not give a polynomial of degree 5.

REMARK I.8. The Archimedean property implies well-ordering; one can see that as follows. Suppose $S \subseteq \mathbb{N}$ is not empty. Pick some $s \in S$. Then the sequence $s-1, s-2, s-3 \ldots$ eventually becomes negative. So only finitely many elements of S other than s could be the minimum of S. Compare them to one another and take the smallest one.

EXAMPLE I.9. We shall prove the following theorem: $\sqrt{2}$ is not a rational number. (We are going to assume that we have some reasonable understanding what "2" means. The square root of 2 must then be a number whose square equals 2. The point of the problem is to show that this root is not an easy number to determine.)

This statement seems not much related to induction, because it is not of the $P(n)$-type. However, consider the following variant: let $P(n)$ be the statement "there is a fraction m/n that equals $\sqrt{2}$, where $m \in \mathbb{N}$". If we can show that all $P(n)$ are false, $\sqrt{2}$ cannot be represented by a rational number.

The base cases $n = 0$ and $n = 1$ have $P(0)$ and $P(1)$ false. For $n = 0$ this is because 0 may not be a denominator, while for $n = 1$ it follows from the fact that 0^2 and 1^2 are less than 2 while $m^2 > 2$ for $m > 1$.

Now suppose, in the spirit of infinite descent, that $P(n)$ is true for some $n \in \mathbb{N}$ and try to show that $P(i)$ must then also be true for some natural $i < n$. To say "$P(n)$ is true" is to say that $2 = m^2/n^2$ for some natural m. In particular, $2n^2 = m^2$ so that m must be even (we are borrowing here a bit from the next chapter), $m = 2m'$. Now feed this info back into the equation: we have $2n^2 = 4m'^2$. The same reasoning shows now that n must be even, $n = 2n'$ for some $n' \in \mathbb{N}$. This now leads to the equation $2n'^2 = m'^2$, and it seems we made no progress. However, stepping back, we realize that $m/n = 2m'/2n' = m'/n'$ which would suggest that $\sqrt{2} = m'/n'$. But this is a representation of $\sqrt{2}$ with a denominator only half the size of n. So we have shown that if $P(n)$ holds then n is even and $P(n/2)$ also holds.

In concrete terms, if the first n for which $P(n)$ holds is called b then b must be even and $P(b/2)$ is also true. Of course, in most cases $b/2$ is less than b (so b wouldn't actually be the first), and the only natural number for which this does not cause a problem is $b = 0$. So if there is any n with $P(n)$ true, then $P(0)$ should also be true. But, as we checked, it isn't. So we deduce that $P(n)$ is always false and $\sqrt{2}$ must be irrational. ◇

REMARK I.10. The following reformulation of induction exists: Suppose there is a statement $P(n)$ for every natural number n, and suppose further we have checked that $P(0)$ is correct. Then $P(n)$ is correct for any $n \in \mathbb{N}$ provided that: for every $k \in \mathbb{N}$ we have a crank that says

If $P(0), P(1), \ldots, P(k)$ are ALL true, then $P(k+1)$ is also true.

This is usually referred to in textbooks as "strong induction", but is not stronger than usual induction. But the formulation has its advantages as we will show soon.

EXERCISE I.11.
(1) Show that $1 = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \cdots$. (Hint: find a guess for the partial sums and then use induction.)
(2) Show that $\frac{1}{1\cdot 2\cdot 3} + \cdots + \frac{1}{n\cdot(n+1)\cdot(n+2)} = \frac{n(n+3)}{4\cdot(n+1)\cdot(n+2)}$ and determine the limit of the corresponding series.
(3) Show that 7 divides $11^n - 4^n$ for all $n \in \mathbb{N}$.
(4) Show that 3 divides $4^n + 5^n$ for odd $n \in \mathbb{N}$.
(5) If one defines a number sequence by $f_0 = 1, f_1 = 1$, and $f_{i+1} = f_i + f_{i-1}$ for $i \geq 1$ then show that $f_i \leq 2^i$.
(6) Show the *Bernoulli inequality*: if $h \geq -1$ then $1 + nh \leq (1+h)^n$ for all $n \in \mathbb{N}$.
(7) Recall that $1 + \ldots + 1 = n$ and $1 + 2 + \ldots + n = \frac{n(n+1)}{2}$. Now show that $1 \cdot 2 + 2 \cdot 3 + \ldots + n \cdot (n+1) = \frac{n(n+1)(n+2)}{3}$.
(8) Show that $1 \cdot 2 \cdot 3 + 2 \cdot 3 \cdot 4 + \ldots + n \cdot (n+1) \cdot (n+2) = \frac{n(n+1)(n+2)(n+3)}{4}$.
(9) Generalize the two previous exercises to products of any length.
(10) Show that $1 + 3 + 5 + \ldots + (2n-1) = n^2$ both by induction and by a picture that needs no words.
(11) Show that 5 divides $n^5 - n$ for all $n \in \mathbb{N}$.
(12) Show that $\sum_{\emptyset \neq S \subseteq \{1,\ldots,n\}} \frac{1}{\prod_{\sigma \in S} \sigma} = n$. (For example, if $n = 2$ then the possible sets S are $\{1\}$, $\{2\}$ and $\{1,2\}$ and then the sum is $\frac{1}{1} + \frac{1}{2} + \frac{1}{1\cdot 2}$, which equals 2 as the formula predicts.) Hint: for $P(n+1)$, split the set of possible sets S into those which do and those which do not contain the number $n+1$.
(13) The list of numbers $1, 2, 3, \ldots, 2N$ is written on a sheet of paper. Someone chooses $N+1$ of these numbers. Prove, by induction, that of those numbers that were picked, at least one divides another one. (Hint: this is not easy. Consider cases: 1. what if all chosen numbers are at most $2N-2$? 2. What if at least N of the chosen numbers are at most $2N-2$? 3. If both $2N-1$ and $2N$ are chosen, ask whether N was chosen. If yes, something nice happens. If not, focus on the chosen numbers that are at most $2N-2$, and pretend that N was also chosen—even though it was not. Now use the inductive hypothesis. How do you deal with the fact that N was not really chosen? Recall that you DO have $2N-1$ and especially $2N$.)

◇

2. Arithmetic

We must begin with some algebra. We officially meet the definition of a ring only in week $X > 1$, but I state it here already. The idea is to list all the important properties of the set of integers.

DEFINITION I.12. A (commutative) *ring* R is a collection of things that have properties like the integers, namely

(1) there is an operation called *addition* (and usually written with a plus-sign) on R such that
   - $r + s = s + r$ for all r, s in R ("addition is commutative");
   - $r + (s + t) = (r + s) + t$ for all r, s, t in R ("addition is associative");
   - there is a *neutral additive element* (usually called "zero" and written 0_R such that $r + 0_R = r = 0_R + r$;
   - for each r there is an *additive opposite* number (usually called the "negative", and written $-r$) with $r + (-r) = 0_R$;
(2) there is an operation called *multiplication* on R (and usually written with a dot such that
   - $r \cdot s = s \cdot r$ for all r, s in R ("multiplication is commutative");
   - $r \cdot (s \cdot t) = (r \cdot s) \cdot t$ for all r, s, t in R ("multiplication is associative");
   - there is an *neutral multiplicative element* 1_R (usually called the *identity*) such that $1_R \cdot r = r = r \cdot 1_R$ for each $r \in R$;
(3) the law of distribution applies: $r \cdot (s + t) = r \cdot t + s \cdot t$ for all r, s, t in R.

Note that no assumption is made on being able to divide (although subtraction is guaranteed, because each ring element has a negative).

REMARK I.13. (1) In some cases one may want to consider rings where the existence of 1_R is not certain, or one may allow $r \cdot s$ and $s \cdot r$ to differ. There is a place and time for such less pleasant rings, but not here.

(2) We will usually drop the subscripts in 0_R and 1_R if it is clear what ring we mean.

(3) We often skip the dot and write ab for $a \cdot b$.

◇

EXAMPLE I.14. We list some examples of rings. If nothing is said, addition and multiplication is what you think.

- The integers, $\mathbb{Z}$. (The case after which the definition is modeled).
- The set of real numbers (or the complex numbers, or the rational numbers). These three rings are special, because in them all nonzero numbers even have inverses (one can divide by them). Such things are called "fields".
- The collection $\mathbb{R}[x]$ of all polynomials in the variable x with real coefficients.
- A weird one: look at all expressions of the form $a + b\sqrt{-5}$ where a and b are integers. It is clear that adding such things gives other such things. It is slightly less obvious that multiplying has the same property (check it!). This ring is denoted $\mathbb{Z}[\sqrt{-5}]$.

◇

EXERCISE I.15. Show that the set of natural numbers $\mathbb{N}$ is not a ring. Find another set that is not a ring and point out why it isn't. ◇

2.1. The Euclidean algorithm. The Archimedean property allows to formulate *division with remainder*:

> For all $a, b \in \mathbb{Z}$ there are $q, r \in \mathbb{Z}$ such that $a = bq + r$ and $0 \leq r \leq |b| - 1$.

The number r is the *remainder of a under division by b.*

This property has pleasant consequences. In order to get concrete, recall that the greatest common divisor and the least common multiple of two integer numbers are defined as follows.

DEFINITION I.16. Let $a, b \in \mathbb{Z}$. Then $\gcd(a,b) = \max\{d \in \mathbb{N} \text{ with } d|a, d|b\}$ and $\operatorname{lcm}(a,b) = \min\{m \in \mathbb{N} \text{ with } a|m \text{ and } b|m\}$. In concrete terms, factorize them into prime powers:

$$\begin{aligned} a &= 2^{a_2} \cdot 3^{a_3} \cdots p^{a_p}, \\ b &= 2^{b_2} \cdot 3^{b_3} \cdots p^{b_p}. \end{aligned}$$

Then

$$\begin{aligned} \gcd(a,b) &= 2^{\min(a_2,b_2)} \cdot 3^{\min(a_3,b_3)} \cdots p^{\min(a_p,b_p)}, \\ \operatorname{lcm}(a,b) &= 2^{\max(a_2,b_2)} \cdot 3^{\max(a_3,b_3)} \cdots p^{\max(a_p,b_p)}. \end{aligned}$$

Consider the equation $a = qb + r$ derived from integers $a > b$ through the Archimedean property. If a number $d \in \mathbb{Z}$ divides a and b then it clearly also divides $r = a - qb$. Conversely, a common divisor of b, r also divides a. So, the set of numbers dividing a, b equals the set of numbers dividing b, r and in particular $\gcd(a,b) = \gcd(b,r)$. We now exploit this to make an algorithm to find $\gcd(a,b)$.

EXAMPLE I.17 (Euclid's Algorithm).
Input: $a, b \in \mathbb{Z}$ with $b \neq 0$.

`Initialize:`

- $c_0 = a$, $c_1 = b$, $i = 1$.

`Iterate:`

- Write $c_{i-1} = q_i c_i + r_i$ where $q_i, r_i \in \mathbb{Z}$ and $0 \leq r_i \leq |c_i| - 1$.
- Set $c_{i+1} = r_i$.
- Replace i by $i + 1$.

`Until:`

- $c_{i+1} = 0$.

Output: $\gcd(a,b) = c_i$. ◇

From what we said above, $\gcd(c_j, c_{j+1}) = \gcd(c_{j-1}, c_j)$ at all stages of the algorithm. In particular, $\gcd(a,b) = \gcd(c_i, c_{i+1}) = \gcd(c_i, 0)$ by our choice of aborting the loop. The gcd of any number and zero is that "any number", so $\gcd(a,b)$ is really the last nonzero remainder c_i we found.

There is another aspect to the Euclidean algorithm, which is the following. The last equation says how to write c_i in terms of the previous two: $c_i = r_{i-1} = c_{i-2} - q_{i-1}c_{i-1}$. The second to last equation can be used to express c_{i-1} in terms of c_{i-2} and c_{i-3}. Substituting, we can write c_i in terms of c_{i-2} and c_{i-3}. Iterating this backwards, one arrives at a linear combination of the form $\gcd(a,b) = \alpha a + \beta b$ for suitable *integers* α, β. This is a fact to remember:

PROPOSITION I.18. *Working backwards from the end of Euclid's algorithm determines a $\mathbb{Z}$-linear combination*

$$\gcd(a,b) = \alpha a + \beta b.$$

EXAMPLE I.19. Let $a = 56 = c_0$, $b = 35 = c_1$. We find $56 = 1 \cdot 35 + 21$, so $c_2 = 21$. Then $35 = 1 \cdot 21 + 14$, so $c_3 = 14$. Next, $21 = 1 \cdot 14 + 7$ and so $c_4 = 7$. In the next iteration we get to the end: $14 = 2 \cdot 7 + 0$ so that $c_5 = 0$. This certifies $c_4 = 7 = \gcd(35, 56)$. Working backwards, $7 = 21 - 14 = 21 - (35 - 21) = 2 \cdot 21 - 35 = 2(56 - 35) - 35 = 2 \cdot 56 - 3 \cdot 35$. ◇

EXERCISE I.20. Find $\gcd(a, b)$ as a $\mathbb{Z}$-linear combination of a, b in the following cases:

(1) $(a, b) = (192, 108)$;
(2) $(a, b) = (3626, 111)$;
(3) $(a, b) = (34, 13)$.

◇

2.2. Primes and irreducibles. We now inspect prime numbers. Feel free to substitute "integer" for "ring element".

DEFINITION I.21. A *unit* of a ring R is a number to which there exists an inverse in the given ring. (Note that this is a relative notion: 2 is a unit in $\mathbb{R}$ with inverse $1/2$, but not a unit in $\mathbb{Z}$ since the only candidate for an inverse, $1/2$, fails to be an integer. The only units in $\mathbb{Z}$ are ± 1).

DEFINITION I.22. For ring elements a, b in R we shall write $a|b$ when the number a divides the number b in R (which just means that there is some element r of R such that $ar = b$). If divisibility fails, we write $a \nmid b$. The example in the previous paragraph indicates that one needs to know the ring in order to answer divisibility questions. ($2|1$ in the ring $\mathbb{R}$ since $1/2 \in \mathbb{R}$, but $2 \nmid 1$ in the ring $\mathbb{Z}$.)

The element p in the ring R is *prime* if $p|ab$ implies that either $p|a$ or $p|b$. On the other hand, p is *irreducible* if $p = ab$ implies that one of a and b must be a unit.

Note that if in some ring the element p is prime or irreducible, then the same is true for $-p$, its additive opposite. One of the fundamental properties of the integers is that *being prime is the same as being irreducible in* $\mathbb{Z}$:

THEOREM I.23. *For any $0 \neq n \in \mathbb{Z}$ the statements "n is prime" and "n is irreducible" are equivalent.*

PROOF. Choose $0 \neq n \in \mathbb{Z}$. An integer n is prime if and only $-n$ is, and it is irreducible if and only if $-n$ is. So we can actually assume that $n \in \mathbb{N}$.

Suppose $n \in \mathbb{Z}$ is prime. We want to show that it is irreducible, which means that whenever $n = ab$ appears as a product of natural numbers, then a or b is a unit. But that is automatic (not specific to the integers) from the definition of "prime" and "irreducible": if $n = ab$ then n divides ab and so it must (as a prime) divide one factor. Say, n divides a so that $a = nq$ with $q \in \mathbb{N}$. Now we have $n = ab = nqb$ and so $1 = qb$ upon division. It follows that b, as a divisor of 1, is a unit. Again, this had nothing to do with the integers beyond the fact that we could "cancel "n" in the product above.

Now suppose n is irreducible, and we try to show that it is prime. So, suppose n divides a product ab with $a, b \in \mathbb{N}$; we need to show that n divides a or b. Let g be the $\gcd(a, n)$. If $g > 1$ then $n = gq$ with $q \in \mathbb{N}$ implies that q is a unit an hence $q = 1$ and so $n = g$ divides a. On the other hand, of $g = 1$ then the Euclidean algorithm says that we can write $1 = g = \alpha a + \beta n$ with $\alpha, \beta \in \mathbb{N}$. Recall that we started with $n \mid ab$, so $cn = ab$ for some $c \in \mathbb{N}$. Multiplying $1 = \alpha a + \beta n$ by c

we get $c = \alpha ca + \beta cn = \alpha ac + \beta ab = a(\alpha c + \beta b)$. In particular, a divides c. But then, the equation $cn = ab$ becomes $(c/a)an = ab$ and cancellation of a shows that n divides b. Note that this part used the Euclidean algorithm, and is not true for all rings. □

THEOREM I.24. *Integers enjoy* unique factorization. *This means first off that for all* $0 \neq n \in \mathbb{Z}$ *there is an* prime factorization*, which is an equation*

$$n = c \cdot p_1 \cdots p_k$$

where each p_i *is a prime number and where* c *is a unit (which in* $\mathbb{Z}$ *implies* $c = \pm 1$*).*

It means secondly that any two such factorizations are almost the same: if

$$d \cdot q_1 \cdots q_\ell = n = c \cdot p_1 \cdots p_k$$

are two such factorizations (c, d *units and* p_i, q_j *prime) then* $k = \ell$ *and (up to sign and suitable reordering)* $p_i = q_i$

For example, $14 = 1 \cdot 2 \cdot 7 = (-1) \cdot 2 \cdot (-7)$ are two different but essentially equivalent prime factorizations of 14.

PROOF. What we need to show comes in two stages: given a natural number n, we need to show it factors at all into prime factors. And then we need to show that any two such factorizations agree, up to reordering. (Note that we can focus on $n > 0$, since $-n = (-1) \cdot n$ and so a factorization of n corresponds to one of $-n$).

We use strong induction. The base case is clear: 1 and 2 are surely factorizable into units and primes: $1 = 1$ and $2 = 2$. So we focus on the crank. So let $2 \leq n \in \mathbb{N}$ and assume that the numbers $1, 2, \ldots, n$ all have a factorization into positive prime numbers. (We don't need to show that we can factor stuff into positive prime numbers, but it is convenient when the number to be fctored is already positive). We consider now $n + 1$. There are two cases: either $n + 1$ is prime, in which case we can write $n+1 = 1 \cdot (n+1)$ as prime factorization. Or, $n+1$ is not prime. Then $n + 1$ is also not irreducible (since we will show later that prime = irreducible, not prime = not irreducible), and so it factors as $n + 1 = 1 \cdot a \cdot b$ with a, b not units. Since $n + 1$ was positive, we can arrange a, b to be positive, and so they both fall into the set of numbers $1, 2, \ldots, n$ about which we already know that they can all be factored. So, factor $a = 1 \cdot a_1 \cdots a_k$ and $b = 1 \cdot b_1 \ldots b_\ell$ into primes, so that $n + 1 = a \cdot b = 1 \cdot a_1 \cdots a_k \cdot b_1 \ldots b_\ell$ has a factorization. So, all natural numbers do have prime factorizations.

Now we need to show that these factorizations are unique. Take any natural number n with two prime factorizations $c_1 \cdot a_1 \cdots a_k = n = c_2 \cdot b_1 \ldots b_\ell$ where c_1, c_2 are units and each a_i, b_j is a prime number. If any a_i or b_j is negativem we can turn their signs by moving the signs into c_1 and c_2. So all a_i, b_j can be assumed to be positive.

Since a_1 is prime, and since it divides the product $c_2 \cdot b_1 \cdots b_\ell$, a_1 must divide one of the factors of this product. It cannot divide c_1 since $c_1 = \pm 1$ and a_1 as a prime has absolute value 2 or more. So, a_1 divides some b_t, so $b_t = a_1 q_1$ for some integer q_1. But b_t was supposed to be prime, hence irreducible, so q_1 is a unit. But a_1 and b_t are positive, so $q = 1$ and $a_1 = b_t$.

Divide out $a_1 = b_t$ to get $c_1 \cdot a_2 \ldots a_k = n/a_1 = c_2 \cdot b_1 \cdots \hat{b}_t \cdot b_\ell$, where the hat indicates that b_t has disappeared from the product. So these are two prime

factorizations for n/a_1. If we now set up a (strong) induction process, we can assume that we already know that n/a_1 has unique (up to reordering and shuffling of units) prime factorization. But then, up to units, the a_i for $i > 1$ are the b_j with $j \neq t$. Since $a_1 = b_t$, follows that n also has unique prime factorization. □

What is left in this subsection will be discussed in the distant future. It is only here to amuse.

EXAMPLE I.25. In $\mathbb{Z}[\sqrt{-5}]$ one can write $6 = 2 \cdot 3 = (1 + \sqrt{-5}) \cdot (1 - \sqrt{-5})$. It turns out that $2, 3, 1 \pm \sqrt{-5}$ are all irreducible (see the exercise below). So $\mathbb{Z}[\sqrt{-5}]$ does not have unique factorization. ◇

EXERCISE I.26.

(1) On $R = \mathbb{Z}[\sqrt{-5}]$ define a *norm* function
$$N \colon a + b\sqrt{-5} \mapsto N(a + b\sqrt{-5}) := a^2 + 5b^2 \in \mathbb{Z}.$$
Convince yourself that the norm of a number is the square of its (complex) absolute value (if you read the number as a complex number).
(2) Show that the norm is multiplicative: $N((a + b\sqrt{-5}) \cdot (c + d\sqrt{-5})) = N(a + b\sqrt{-5}) \cdot N(c + d\sqrt{-5})$.
(3) Find all elements of our ring R that have norm 1.
(4) Show that no element has norm 2 or 3.
(5) Show that $2, 3, 1 \pm \sqrt{-5}$ are all irreducible by inspecting the ways of factoring $N(2), N(3)$ and $N(1 \pm \sqrt{-5})$.

◇

2.3. Some famous theorems and open problems on prime numbers. The proofs of several theorems here is rather beyond us, but if interested you might look at [**?**] for further pointers.

The most basic, famous, and memorable was (together with the proof given here) already known to Euclid:

THEOREM I.27. *There are infinitely many prime numbers.*

PROOF. Suppose $p_1, \ldots, p_k$ are prime numbers. Then $M = p_1 \cdot \ldots \cdot p_k + 1$ is not divisible by any of them. It might be the case that M is prime, but that does not need to be so. However, M *does* have a prime factorization, $M = c \cdot q_1 \cdots q_t$ with c a unit and all q_i prime. Since no p_i divides M, none of q_i is on the list of primes $p_1, \ldots, p_k$. In other words, any finite list of primes is missing at least one other prime. □

Recall now the *harmonic series*

$$\underbrace{\frac{1}{1}}_{a_0} + \underbrace{\frac{1}{2}}_{a_1} + \underbrace{\frac{1}{3} + \frac{1}{4}}_{a_2} + \underbrace{\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}}_{a_3} + \underbrace{\frac{1}{9} + \frac{1}{10} + \frac{1}{11} + \frac{1}{12} + \frac{1}{13} + \frac{1}{14} + \frac{1}{15} + \frac{1}{16}}_{a_4} + \cdots$$

EXERCISE I.28. Show that the harmonic series diverges (has no finite value). (Hint: what can you say about each a_i?). ◇

Suppose you only took the terms that are inverses of prime numbers,

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \frac{1}{13} + \frac{1}{17} + \frac{1}{19} + \frac{1}{23} + \frac{1}{29} + \frac{1}{31} + \cdots,$$

how does this series behave? We will need to use the following fact.

EXERCISE I.29. Show:

(1) For any real or complex number x and any integer n, $\frac{1-x^{n+1}}{1-x} = 1 + x + x^2 + \cdots + x^n$.
(2) As long as $|x| < 1$, $\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$.

$\diamond$

THEOREM I.30. *The sum of the reciprocals of all the prime numbers is still divergent.*

This of course implies also that there are infinitely many primes. However, its proof is far more delicate than Euclid's proof above. We give here the idea behind the proof; that the steps can be made rigorous is somewhat involved.

PROOF. Any positive integer n is uniquely the product of positive prime numbers. The emphasis is on "unique", if you multiply together different sets of numbers you get different end products, pun intended.

Consider now the product

$$(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots) \cdot (1 + \frac{1}{3} + \frac{1}{3^2} + \ldots) \cdot (1 + \frac{1}{5} + \frac{1}{5^2} + \ldots) \cdots (1 + \frac{1}{p} + \frac{1}{p^2} + \ldots),$$

where p is some prime number.

If you actually multiply this out, the resulting mess contains, for each choice of finitely many primes $p_1, \ldots, p_k$ bounded by p, the quantity $1/(p_1 \cdots p_k)$. So, the mess actually contains exactly one copy of the inverse of every natural number that has a prime factorization in which only powers of primes occur that are bounded by p. Taking the limit $p \to \infty$, one might try to believe in an equation

$$\sum_{n=1}^{\infty} \frac{1}{n} = \prod_{p \text{ prime}} (1 + 1/p + 1/p^2 + 1/p^3 + \ldots),$$

and conclude that the right hand side is, like the harmonic series on the left, infinite. The art (which we omit) is to make this argument and all that builds on it, so that it become mathematically sound.

Using the geometric series, this suggests $\sum_{n=1}^{\infty} \frac{1}{n} = \prod_{p \text{ prime}} \frac{1}{1-1/p}$. Now take logs on both sides. The RHS turns into $\sum_{p \text{ prime}} \ln(\frac{1}{1-1/p}) = -\sum_{p \text{ prime}} \ln(1 - 1/p)$. Looking at the graph of the log-function, one sees that $\ln(1-x) \leq x$, so $\sum_{p \text{ prime}} \ln(\frac{1}{1-1/p}) \leq \sum_{p \text{ prime}} \frac{1}{p}$. But the left hand side was already infinite, so therefore the sum of the prime reciprocals must be too. $\square$

This theorem says that there are still quite a bit of primes, namely enough to make the sum diverge. (Remember: if you just looked at the subsum given by powers of your favorite prime, this would be geometric and hence convergent). How many primes are there in comparison to all numbers? This is best asked in the context of prime density.

THEOREM I.31 (Prime Number Theorem). *Let p_k be the k-th prime number. Then the fraction $\frac{p_k}{k \cdot \ln(k)}$ approaches 1 as $k \to \infty$.*

Equivalently, if you pick randomly a number n near the number N then the chance that n is prime is about $\ln(N)/N$.

Another set of questions one could ask is about primes in arithmetic progressions (rather than in all numbers). That means: let

$$\mathcal{A}(a, n) = \{a, a + n, a + 2n, \ldots\}$$

be the arithmetic progression starting with $a \in \mathbb{N}$ and with shift $n \in \mathbb{N}$.

DEFINITION I.32. The Euler ϕ-function attaches to each natural number n the number of natural numbers less than n that are relatively prime to n.

For example, $\phi(12) = 4$ because of $1, 5, 7, 11$, and $\phi(7) = 6$ because of $1, 2, 3, 4, 5, 6$.

THEOREM I.33. *The set $\mathcal{A}(a, n)$ contains approximately $x/(\ln(x) \cdot \phi(n))$ prime numbers of size less than x.*

If you agree that $\phi(1) = 1$ then this theorem specializes to the Prime Number Theorem above.

A *prime twin* is a pair $\{p, p + 2\}$ of primes (such as $\{101, 103\}$).

CONJECTURE I.34. *There are infinitely many twin primes.*

Not much is known except that if you looked at the subsum of the harmonic series that is comprised of the terms that are twin primes only then this sum *does* converge. So, there are rather fewer twin primes than prime numbers. However, you might relax your twin focus a little and ask "how many pairs of primes are there that are no further apart than 70 million"? In 2014, Yitang Zhang who has a mathematics PhD from Purdue, proved that the answer is now "yes"; he was awarded a very prestigious MacArthur Grant for this, see `http://www.macfound.org/fellows/927/`.

There are two more famous conjectures. The first is easy to state:

CONJECTURE I.35.
- *(Goldbach, strong form) All even integers $n > 2$ can be written as the sum of two prime numbers.*
- *(Goldbach, weak form) All integers $n > 1$ can be written as the sum of at most three primes.*

The other requires a bit of preparation.

DEFINITION I.36. The *Riemann zeta function* is

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Here, the input s is to be taken a complex rather than a real number. This sum will converge (absolutely) if the real value of s is greater than 1 because of things we know about the geometric series from Exercise I.29. On the other hand, the harmonic series teaches that at $s = 1$ the value of ζ is infinite. In the places where s has real part less than 1, one can use a graduate-level technique called "analytic continuation" to make sense of the series (even though it probably diverges). The result is a nice function in s that can have poles every now and then (such as in $s = 1$). Values of the zeta function appear in physics (how odd!) and chemistry (no more even!), and they have a tendency to involve the number π. At negative even integers, $\zeta(s)$ is zero for reasons that come from a "functional equation" that ζ satisfies:

$$\zeta(s) = 2^s \pi^{s-1} \sin(\pi s/2) \Gamma(1 - s) \zeta(1 - s).$$

Here, π is the usual π from trigonometry, and Γ is a version of "factorial" for non-integer input. (If you believe this equation, you must believe that $\zeta(-2n) = 0$ by looking at the contribution of the sine).

CONJECTURE I.37 (Riemann hypothesis). *Apart from negative even integers, all other values s where $\zeta(s) = 0$ satisfy: s has real part $1/2$.*

This one is one of the seven Clay *Millennium problems*, the complete list of which can be found under `http://www.claymath.org/millennium-problems`. It would earn you $1.000.000 to crack it. It also featured (with the Goldbach Conjecture) as one of *Hilbert's Problems*. This list, see `http://en.wikipedia.org/wiki/Hilbert's_problems`, was compiled by perhaps the last person that understood all of mathematics as it existed at the life-time of that person. The list was presented (in part) at the International Congress of Mathematicians in 1900. It has hugely influenced mathematics and mathematicians during the 20th century (although lots of mathematics was made that didn't related directly to the list), and the list of Clay Millennium Problems can be viewed as its descendant. Some solutions to Hilbert's problems have been awarded with a Fields medal `http://en.wikipedia.org/wiki/Fields_Medal`.

3. Modular arithmetic

We are perfectly used to claim that 4 hours after it was 11 o'clock it will be 3 o'clock. In effect, we equate 12 with zero in these calculations. In this section we learn how to calculate on more general "clocks" and even solve equations on "clocks".

3.1. Computing "modulo": $\mathbb{Z}/n\mathbb{Z}$.

DEFINITION I.38. For any integer n, write $n\mathbb{Z}$ for the set of all integer multiples of n, so $n\mathbb{Z}$ stands for $\{\ldots, -3n, -2n, -n, 0, n, 2n, \ldots\}$.

For an integer a write then $a + n\mathbb{Z}$ for the collection of all integers who leave remainder a when divided by n, so $a+n\mathbb{Z} = \{\ldots, a-2n, a-n, a, a+n, a+2n, \ldots\}$. Note that $a + n\mathbb{Z} = (a + n) + n\mathbb{Z}$. Such sets we call *cosets modulo* n, while the various integers that float around in a given coset are called *representatives*. They are also sometimes written as "a mod n". If the value of n is understood from the context, we may write $\overline{a}$ for $a + n\mathbb{Z}$. (If $n = 12$, and if $a = 3$, then $\overline{3} = 3 \bmod 12 = 3 + 12\mathbb{Z} = \{\ldots, -21, -9, 3, 15, 27, \ldots\}$. This is the set of all times on an absolute clock at which a usual clock shows "3o'clock").

Finally, write $\mathbb{Z}/n\mathbb{Z}$ ("zee modulo enn zee") for the collection of all cosets modulo n. (If $n = 12$, this is the set of all possible full hours, clustered by what a 12-hour-clock would show).

Here is an example on a very small "clock".

EXAMPLE I.39. Let $n = 4$. there are four cosets modulo 4, namely

$$0 + 4\mathbb{Z} = \{\ldots, -8, -4, 0, 4, 8, \ldots\}, \quad 1 + 4\mathbb{Z} = \{\ldots, -7, -3, 1, 5, 9, \ldots\}, \tag{1}$$

$$2 + 4\mathbb{Z} = \{\ldots, -6, -2, 2, 6, 10, \ldots\}, \quad 3 + 4\mathbb{Z} = \{\ldots, -5, -1, 3, 7, 11, \ldots\}. \tag{2}$$

So, $\mathbb{Z}/4\mathbb{Z} = \{\overline{0}, \overline{1}, \overline{2}, \overline{3}\}$. Representatives of $\overline{3}$ include $3, 7, -133$ among others. ◇

REMARK I.40. While this "modular arithmetic" may seem a bit esoteric, be assured that it is far more important than you can imagine. For example, all computers on this planet calculate in $\mathbb{Z}/2\mathbb{Z}$ or a slightly more complicated scheme. Without modular arithmetic, there would be no twitter, no email, no instagram. Not even a digital watch. ◇

Amusingly, one can calculate with these cosets as if they were numbers. Regarding addition, we always knew that 4 hours after it was 11 o'clock it will be 3 o'clock because the coset of 11 plus the coset of 4 gives the coset of 15, which is to say, of 3. That one can also multiply is a new idea:

$$\begin{aligned}(a+n\mathbb{Z})+(b+n\mathbb{Z}) &:= (a+b)+n\mathbb{Z};\\ (a+n\mathbb{Z})-(b+n\mathbb{Z}) &:= (a-b)+n\mathbb{Z};\\ (a+n\mathbb{Z})\cdot(b+n\mathbb{Z}) &:= (a\cdot b)+n\mathbb{Z}.\end{aligned}$$

The amusing part is that this works well on any choice of representatives. For example, in order to compute $(2+7\mathbb{Z})+(3+7\mathbb{Z})$ you could say: pick representative $-5=2+(-1)\cdot 7$ for $\overline{2}$ and representative $24=3+3\cdot 7$ for $\overline{3}$. Add them to get 19 and so $(2+7\mathbb{Z})+(3+7\mathbb{Z})=19+7\mathbb{Z}$. Of course, you probably would have chosen $2=2+0\cdot 7$ and $3=3+0\cdot 7$ as representatives, resulting in the coset of 5. The point is that $\overline{5}$ and $\overline{19}$ are actually *the same*. In order to prove that this is always ok and not just in our explicit example you should carry out

EXERCISE I.41. Show that for all choices of a, a', b, b', n with $n|(a-a')$ and $n|(b-b')$ one has:

- n divides $(a+b)-(a'+b')$; (this says that the cosets of $a+b$ and of $a'+b'$ always agree);
- n divides $(a-b)-(a'-b')$; (this says that the cosets of $a-b$ and of $a'-b'$ always agree);
- n divides $ab-a'b'$; (this says that the cosets of ab and of $a'b'$ always agree).

◇

3.2. Divisibility tests. Suppose you are asked "is 1234567 divisible by 3?". You could sit down and calculate, or ask a friend with a computer, but you could also think. Such as: $n=1234567$ comes to me as a decimally expanded number, $n=10^k\cdot a_k+\cdots+10\cdot a_1+a_0$ where $k=6$, $a_6=1$, $a_5=2$, $a_4=3$, $a_3=4$, $a_2=5$, $a_1=6$ and $a_0=7$. In order to test divisibility of n by 3, I'd like to know whether n mod 3 is zero or not. But, $n \bmod 3=(10^k\cdot a_k+\cdots+10\cdot a_1+a_0) \bmod 3$, and " mod " goes well over addition and multiplication:

$$\begin{aligned}n \bmod 3 &= \sum (a_i \bmod 3)\cdot(10 \bmod 3)^i\\ &= \sum (a_i \bmod 3)\cdot(1 \bmod 3)^i\\ &= \sum a_i \bmod 3.\end{aligned}$$

It follows that n is a multiple of 3 if and only if the sum of its digits is a multiple of 3. Of course, if you want, you can reapply this idea to the output:

$$\begin{aligned}1234567 \bmod 3 &= (1+2+3+4+5+6+7) \bmod 3 = 28 \bmod 3\\ &= (2+8) \bmod 3 = 10 \bmod 3\\ &= (1+0) \bmod 3 = 1 \bmod 3.\end{aligned}$$

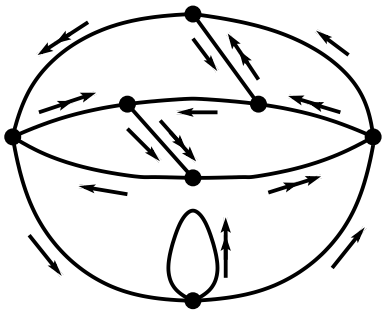
Hence, 1234567 leaves rest 1 when divided by 3.

Obviously, a similar thought works for 9 instead of 3 since any power of 10 leaves rest 1 when divided by 9.

EXERCISE I.42. Prove that 11 divides n if and only if it divides $a_0-a_1+a_2-a_3+\cdots+(-1)^k a_k$. ◇

EXAMPLE I.43. Here is a test on divisibility by 7 by way of a picture. Take the digital representation of your number. Start a the bottom node in the picture. Starting with the front digit, do the following for each digit: go as many steps along simple arrows as the current digit says; then go one step along a double arrow.

If you end up at the bottom node, n is a multiple of 7. (In general, the node index is the remainder of n divided by 7).



Question: Following the single arrows just counts how big the current digit is. What is the function of the double arrows?

CHAPTER II

Week 2: Groups

1. Symmetries

EXAMPLE II.1. Let T be an equilateral triangle  . We imagine its vertices to carry labels A, B, C, but these are not visibly written on the vertices.

We want to discuss ways to move around the triangle so that it looks after the movement the same way as before. There are the rotations by $0° = e, 120° =: \ell, 240° =: r$. Then there are 3 reflections, a, b, c. Here, a leaves A fixed and interchanges B with C, and so on. One checks there are no other symmetries (3! is an upper bound). So $\mathrm{Sym}($  $) = \{e, r, \ell, a, b, c\}$.

Symmetries can be composed. For example, $rr = \ell$, $rrr = e$. The 36 products are as follows (the entry in row r and column a, for example, is the composition ar):

	e	r	ℓ	a	b	c
e	e	r	ℓ	a	b	c
r	r	ℓ	e	b	c	a
ℓ	ℓ	e	r	c	a	b
a	a	c	b	e	ℓ	r
b	b	a	c	r	e	ℓ
c	c	b	a	ℓ	r	e

Note: if you write ra for example, you mean "first a, then r, in the same way as $f(g(x))$ evaluates first $g(x)$ and then stuffs this into f. So, we imagine ra really means "r applied to (the result of a applied to triangle)".

DEFINITION II.2. The full set of symmetries of a regular n-gon is denoted D_n and called the n-th *dihedral group*.

Note that D_n consists of n rotations and n reflections (for $n \geq 3$). In contrast, D_2 is the symmetries of a line segment, which is just $\{e, f\}$ where f is the flip exchanging the ends. D_1 is the symmetries of a point, so just $\{e\}$. The two composition tables are

	e	f
e	e	f
f	f	e

and

	e
e	e

.

REMARK II.3. It is clear that the row labeled e and the column labeled e always agree with the top row and column that simply list the the symmetries. We will henceforth skip this row and column and place e in the upper left corner. So the table for D_2 would just be

e	f
f	e

.

EXAMPLE II.4. Let OS be the oriented square  . Its symmetry group has fewer elements than that of the square, namely only the rotations $\{e, \ell, \ell^2, \ell^3\}$ with

a composition table

e	ℓ	ℓ^2	ℓ^3
ℓ	ℓ^2	ℓ^3	e
ℓ^2	ℓ^3	e	ℓ
ℓ^3	e	ℓ	ℓ^2

since $\ell^4 = e$ and there is no other relation. This group of symmetries is called the *cyclic group* C_4, since you only need to know one element of it (such as ℓ) and everyone else is a power of it. The "4" comes from the fact that $\ell^4 = e$ and no lower positive power will do (or, because there are 4 elements in this cyclic group. Itis like a clock with 4 hours).

EXAMPLE II.5. Now we look at the symmetries of the letter H. It has 4 elements, the identity e, the rotation $\curvearrowleft$ by 180° the left-right flip $\leftrightarrow$ and the up-down flip $\updownarrow$. The table is

e	$\leftrightarrow$	$\updownarrow$	$\curvearrowleft$
$\leftrightarrow$	e	$\curvearrowleft$	$\updownarrow$
$\updownarrow$	$\curvearrowleft$	e	$\leftrightarrow$
$\curvearrowleft$	$\updownarrow$	$\leftrightarrow$	e

.

(You should actually check a few of the products listed here).

This set of symmetries is called the *Klein 4-goup* and denoted KV_4. Felix Klein was the superstar of symmetry investigations. Note that Sym(H) $\subseteq$ Sym($\square$) since drawing a box $\boxed{\text{H}}$ around the H does not change the symmetries.

Note also that the tables for KV_4 and C_4 are seriously different since e shows up on the diagonal with different multiplicity. (The element e is special and can be recognized even if you use different letter as it is the one element for which $ex = x$ for every symmetry x).

2. Groups

We are now ready to define what a group is. It generalizes the symmetry studies above.

DEFINITION II.6. A *group* is a set G with an operation $\cdot$ that takes ordered pairs (g, g') from $G \times G$ and "multiplies" them to other elements of G. (In other symbols, $\cdot\colon G \times G \to G$). This operation must satisfy the following conditions:

(1) $a \cdot (b \cdot c) = (a \cdot b) \cdot c \forall a, b, c \in G$ (associativity);
(2) there is an *identity* or *neutral element* $e \in G$ such that $\forall g \in G$ one has $e \cdot g = g \cdot e = g$;
(3) $\forall g \in G$ there is an *inverse* element $\tilde{g} \in G$ with $g \cdot \tilde{g} = \tilde{g}g = e$.

REMARK II.7. (1) As a matter of convenience we often skip the dot and just write ab for $a \cdot b$ (as we have done above for symmetries). Also, one usually writes g^{-1} for the $\tilde{g}$ in item (3) of the definition.

(2) Be warned, that one of the conditions you might have placed here is missing: we do not require that $ab = ba$ in general. If you think of group elements as prodecures that you compose, this is clear: it makes usually some difference whether you put on socks first and then shoes, or the other way round.

(3) Note the following quirk of this asymmetry: if $ab = c$ then $c^{-1} = b^{-1}a^{-1}$. Thinking of socks and shoes makes this more obvious. You also have seen this for taking inverses of matrices, and of course a matrix is just a procedure acting on a

vector (by multiplication), so this all fits together. The invertible $n \times n$ matrices with real entries are one of the standard examples of a group. It is called the *general linear group* $\mathrm{Gl}(n, \mathbb{R})$.

(4) Associativity implies that the product $g \cdot g \cdots g$ of many copies of the same element is uniquely defined and does not depend on in what order we multiply the copis. (For example, you could take 4 copies and multiply them like $((gg)g)g$ or like $((gg)(gg))$. For 3 factors, this is explicit from the associativity rule, and for more than 3 we discuss it in Lemma II.13 below).

THEOREM II.8. *The symmetries on any chosen object form a group.*

PROOF. The set G is the set of symmetries, the operation of G is composition of symmetries. The identity is the symmetry that does not move. The inverse of a symmetry is the symmetry done backwards. The associativity rule comes from the fact that it holds for composition of functions (where it boils down to reading the definition of composition). □

To each group one can write down a table similar to the tables we have looked at for symmetries. For group tables one uses the phrase *Cayley table*. You surely have noticed at this point that each row and column of such table contains each element (once). That is no accident:if you had $ac = bc$ for example, the same element showing up as product twice in the same column, then also $(ac)c^{-1} = (bc)c^{-1}$ and so $a(cc^{-1}) = b(cc^{-1})$, or $ae = be$ which entails $a = b$ according to the various group axioms. We say that *groups have the cancellation property.*

EXAMPLE II.9. Here is a list of important groups with their operations. The $*$ just indicates usual multiplication.

(1) $(\mathbb{Z}, +)$, the integers with addition.

(2) $(\mathbb{Z}/n\mathbb{Z}, +)$, modular addition (verification in HW);

(3) $(\mathbb{R}^n, +)$, the n-dimensional vector space has as part of its axioms the group axioms for $+$;

(4) $(\{1, -1\}, *)$ with a Cayley table similar to that of the dihedral group D_2;

(5) $(\mathbb{R}_{>0}, *)$, which contains the previous group and uses the same operation, identity and inverse;

(6) $(\mathbb{R}\backslash\{0\}, *)$, which contains the previous group and uses the same operation, identity and inverse;

(7) $(\mathrm{Gl}(n, \mathbb{R}), *)$ and $(\mathrm{Gl}(n, \mathbb{C}), *)$ as previously mentioned;

EXAMPLE II.10. We consider here the list of all possible groups with 4 or fewer elements.

(1) If $|G| = 1$ then G is just e and the Cayley table is that of the dihedral group D_1.

(2) If $|G| = 2$ then $G = \{e, f\}$ for some other element f, and by the cancellation rule ff can't be ef and so must be e. So G has a Cayley table essentially that of the dihedral group D_2.

(3) If $|G| = 3$, $G = \{e, a, b\}$. Since ab can't be $ae = a$, but also not $eb = b$, by cancellation, it must be $ab = e$. Then we are forced to concede $aa = b$ and $bb = a$, and so $a^3 = e$. So the table is the one you get from the rotational symmetries of the equilateral triangle alone:

e	a	a^2
a	a^2	e
a^2	e	a

, with $b = a^2$. This is essentially C_3.

(4) If $|G| = 4$, with elements e, a, b, c then by the same reasoning as before, ab is e or c.

First case: $ab = c$. Then $ae = a$, $ab = c$ and since ac can't be c (since ec is c), we conclude $ac = b$. That then settles it, using associativity: we get

e	a	b	c
a	e	c	b
b	c	e	1
c	b	a	e

This is, up to relabeling a to $\leftrightarrow$, and b to $\updownarrow$, and c to $\curvearrowleft$, the same table as that of KV_4.

Second case: $ab = e$. Then a and b are mutual inverses. Since e is its own inverse $ee = e$, c must also be (for lack of other partners) its own inverse $cc = e$. Moreover, for cancellation reasons, ac can't be a or c and it is not e (since the inverse of c is not a but c) and so $ac = b$. We now know $ae = a$, $ab = e$, $ac = b$. Thus, $aa = c$. Next, in the same way we found $ac = b$ we also find $ca = b$. That forces $cb = a$ since $cc = e, ca = b, ce = c$. At this point, our knowledge is:

e	a	b	c
a	c	e	b
b	e		
c	b	a	e

But now the b-row is automatic. In particular, one sees $a = a^1, c = a^2, b = a^3, e = a^4$. This is the same table as for C_4, just the letter a replacing the letter ℓ.

DEFINITION II.11. If two groups G and G' of equal size permit a pairing of their elements such that renaming the elements of G by their partner elements of G' turns the Cayley table of G into the Cayley table of G', then we call G and G' *isomorphic* and write $G \cong G'$.

For example, Sym(S) $\cong D_2 \cong$ Sym(A) although the actual motions that carry S to S (the roation $\curvearrowleft$) and A to A (the flip $\leftrightarrow$) are very different. We only care about the abstract relationships of the symmetries, and they are in both cases the table

e	x
x	e

, with x being $\curvearrowleft$ in one case, and $\leftrightarrow$ in the other.

3. Cyclic and Abelian groups

DEFINITION II.12. A group $(G, \cdot)$ is called *cyclic* isf there is some element $g \in G$ such that every other element $g' \in G$ is a (possibly negative) power of g.

The element g is a *generator* for G.

The standard example is $(\mathbb{Z}, +)$ where there are two generators: every integer is a multiple of 1, but it also a multiple of -1.

Other examples include the group of rotational symmetries on a regular n-gon (these are the rotations that form 50% of the dihedral group D_n, $n \geq 3$), and the groups $(\mathbb{Z}/n\mathbb{Z}, +)$ for any $n \in \mathbb{N}$.

Writing down the Cayley tables for these cyclic groups one notices that these Cayley tables are all symmetric. In other words, $ab = ba$ for all a, b in such a group. This is no accident as we show now.

LEMMA II.13. *If G is cyclic, generated by $g \in G$, then for all elements $a, b \in G$ we have $ab = ba$.*

PROOF. In fact, we pay back a debt here on the meaning of g^i. We denote g^2 the product gg, and g^3 the product $g(gg) = (gg)g$, the results being the same by

associativity. For higher powers, argue as follows. Suppose we have proved that the product of k copies of g is independent of the placement of parentheses. Then, for $i+j=k+1$ and $i,j>0$ we have $(g^i)(g^j)=(g^i)(g(g^{j-1}))=((g^i)g)(g^{j-1})=(g^{i+1})(g^{j-1})$. So one may shuffle one copy of g after the other from one factor to the other without changing the product. So, a product of k copies of g only depends on g and k but not the placing of parentheses.

Let a,b be in a cyclic group generated by g. According to the definition of a cyclic group, there are numbers $i,j\in\mathbb{Z}$ such that $a=g^i, b=g^j$. But then $g^ig^j=g^jg^i$ since they are both the product of $i+j$ copies of g. □

DEFINITION II.14. If in a group $(G,\cdot)$ it is true that $gh=hg$ for all $g,h\in G$ then G is *Abelian*.

Cyclic groups are Abelian, but lots of groups are not, such as Sym(△). (the elements a,b,c only have two different powers, e and themselves, and ℓ,r only have the three powers e,ℓ,r). Also, Sym(H) is not cyclic as one sees easily from the squares.

The question when a power of an element is e seems to be important:

DEFINITION II.15. For an element g of the group $(G,\cdot)$, the smallest number $k\in\mathbb{N}_{>0}$ such that $g^k=e$ is its *order* $\mathrm{ord}(g)$. (There might not be such a k (like for $3\in(\mathbb{Z},+)$ for example. We then say $\mathrm{ord}(g)=0$ or $\mathrm{ord}(g)=\infty$.)

We call $|G|$ the *order of the group*.

Inside Sym(△), both the powers of ℓ and the powers of a form what we call a subgroup.

DEFINITION II.16. If $(G,\cdot)$ is a group, then a *subgroup* is a subset $H\subseteq G$ which, when equipped with the multiplication of G, is a group in its own right.

As mentioned, $H_1=\{e,\ell,r\}$ and $H_2=\{e,a\}$ are subgroups of Sym(△). The Cayley table of a subgroup is simply the appropriate subtable of the Cayley table for G.

Note that G counts as a subgroup of G, but he empty set is not a subgroup. This is because one group axiom postulates the existence of an identity in G, so $\{e\}$ is the smallest subgroup of any G (called the "trivial subgroup"). A subgroup different from G and $\{e\}$ is called a *proper subgroup*.

REMARK II.17. (1) If you recall the idea of a vector subspace, there was a criterion that said "if $W\subseteq V$ is a subset then it is a subspace provided that W is closed under addition, and under scaling by real numbers". There is a similar test for subgroups: $\emptyset\neq H\subseteq G$ is a subgroup if for all $h_1,h_2\in H$ the element $h_1^{-1}h_2$ is again in H.

Why? Associativity is inherited from G; e is in H because if $h\in H$ then by the test, $h^{-1}h=e$ is in H; if $h\in H$ then $h^{-1}e=h^{-1}$ is also in H.

(2) If $H\subseteq G$ is a subgroup and $h\in H$ then the order of h as element of H is the same as the order of h as element of G, since we use the same operation.

EXAMPLE II.18. This is rehashing a previous remark. Suppose G and G' are groups of the same size, and assume further that there is a bijection between the elements of G and the elements of G' that turns one Cayley table into the other. (We called such groups isomorphic).

If you take an element $g \in G$ then the order of g in G is the same as the order of its twin in G'. This follows from the translation of the Cayley tables. Basically, this says that if ϕ is the bijection then $\phi(a \cdot_G b) = \phi(a) \cdot_{G'} \phi(b)$.

The upshot is that one can use order to discriminate between groups. For example, KV_4 is not C_4 because KV_4 has 3 elements of order 2, and C_4 only one.

One can also count subgroups and compare: KV_4 has 5 subgroups, namely $\{e\}, \{e, \leftrightarrow\}, \{e, \updownarrow\}, \{e, \curvearrowleft\}, KV_4$. But C_4 has only three: $\{e\}, \{e, \ell^2\}, C_4$. So these two groups cannot be isomorphic.

Recall that for sets A, B the Cartesian product $A \times B$ is the set of all ordered pairs (a, b) with $a \in A, b \in B$.

DEFINITION II.19. If G, G are groups, then $G \times G'$ is also a group, with multiplication $(g_1, g_1') \cdot (g_2, g_2') = (g_1 \cdot_G g_2, g_1' \cdot_{G'} g_2')$.

For example, $(\mathbb{R}^2, +)$ is simply $(\mathbb{R}, +) \times (\mathbb{R}, +)$.

EXAMPLE II.20. The cyclic groups $C_2 = \{e, a\}$ with $a^2 = e$ and $C_3 = \{e, b, b^2\}$ with $b^3 = e$ have Cayley tables as discussed earlier. In these groups, e has order 1, a has order 2 and b has order 3. What about elements of $C_2 \times C_3$?

The list of elements has 2×3 members, and they are $(e, e), (e, b), (e, b^2), (a, e), (a, b), (a, b^2)$. One sees easily that (e, e) has order $1 = \mathrm{lcm}(1, 1)$; (e, b) and (e, b^2) have order $3 = \mathrm{lcm}(1, 3)$; (a, e) has order $2 = \mathrm{lcm}(2, 1)$; and (a, b) and (a, b^2) have order $6 = \mathrm{lcm}(2, 3)$.

(We explain the lcm statements: in general one has $\mathrm{ord}(x, y) = \mathrm{lcm}(\mathrm{ord}(x), \mathrm{ord}(y))$. Why? Surely, the lcm is a power that sends (x, y) to (e, e). The powers satisfy $x^k = x^{\mathrm{ord}(x)+k}$ and $y^k = y^{\mathrm{ord}(y)+k}$. So $y^k = e$ implies $y^k = y^{\mathrm{ord}\, y} = e$ and so $y^{\gcd(\mathrm{ord}(y),k)} = e$. But the gcd can't be bigger than $\mathrm{ord}(y)$ because it needs to divide it, and it can't be smaller than the order because of the definition of order. The only way out is that the order is the gcd. So the order of y divides any k with $y^k = e$. Similarly, the order of x divides any exponent i with $x^i = e$ and the order of (x, y) divides any exponent with $(x^i, y^i) = (e, e)$. So whatever the order of (x, y) is, it must be a multiple of $\mathrm{ord}(x)$ and $\mathrm{ord}(y)$, while being as small as possible. That is simply the lcm.

4. Automorphisms

DEFINITION II.21. An *automorphism* of a group G is a a relabeling of its elements that preserves the Cayley table.

For example, C_3 is the group $\{e, a, b\}$ with rules $ab = ba = e, ea = ae = a, be = eb = b$. This is completely symmetric in a, b. So the bijection

$$\begin{aligned} a &\mapsto b, \\ b &\mapsto a, \\ e &\mapsto e \end{aligned}$$

is an automorphism of C_3. (Geometrically, this switches left roation with right rotation in the rotational symmetries of an isosceles triangle).

In principle, an automorphism is just a special permutation of the elements. So one can search through all permutations and just keep those that preserve the Cayley table. This is not efficient if G has many elements, one should use the group structure in the search.

Note, that one possible automorphism is always just to leave everything as is. That is like the e in a group. In fact, automorphisms do form a group in their own right. $e \in \mathrm{Aut}(G)$ is the relabeling that sends every element of G to itself; multiplication of two automorphisms is just doing one relabeling after the other; the inverse of an automorphism is the relabeling done backwards.

Looking at C_3: there are only two automorphisms, the identity on C_3, and the switch discussed above. This is because e_G must be sent to e_G ($yx = x$ for all x is something only the element $y = e$ does, and relabelings must preserve products!). Composing the switch with itself gives the identity on C_3. So, it is fair to say that $\mathrm{Aut}(C_3)$ is basically the group with table as in Example II.10 part (2).

Another interesting example occurs in C_4, which is the group of symmetries of the oriented square  , with elements $a = \ell, b = \ell^2, c = \ell^3$ and the understanding $\ell^4 = e$. Here one can interchange a and c while keeping e, b fixed:

$$\begin{aligned} a &\mapsto c, \\ c &\mapsto a, \\ b &\mapsto b, \\ e &\mapsto e. \end{aligned}$$

It is easy to check that this relabeling preserves the table when written with a, b, c, e.

Again, this is the only automorphism since we must send e to e (the only element of order one) and b to b (the only element of order two). Aside, of course, of the identity on C_4 leaving everything fixed. So, $\mathrm{Aut}(C_4)$ is the "same" group as $\mathrm{Aut}(C_3)$, both isomorphic to C_2, sameness in the sense that ther Cayley tables are the same after renaming.

Example II.22. The automorphisms of KV_4 are more interesting.

Suppose we fix a. Then we could fix b but that also forces us to fix c as the only remaining element of order 2. That then comes down to the identity, sending each element of KV_4 to itself. Alternatively, if we do not fix b, the only open destination for b is c. So $a \mapsto a, b \mapsto c, c \mapsto b$.

Alternatively, we can try sending $a \mapsto b$. If we fix c then we are in a similar situation as before, because then b must go to a. On the other hand, we could send $a \mapsto b$ and $b \mapsto c$ which forces $c \mapsto a$.

The cases where $a \mapsto c$ are similar, with the letters b and c exchanged.

Altogether, the 6 options are summed up in the following table, where each row represents an automorphism, and where it sends the elements of G is recorded in the row.

	e	a	b	c
ψ_e	e	a	b	c
ψ_a	e	a	c	b
ψ_b	e	c	b	a
ψ_c	e	b	a	c
ψ_ℓ	e	b	c	a
ψ_r	e	c	a	b

For notation: ψ_e keeps everyone fixed; ψ_x for $x \in \{a, b, c\}$ keeps e, x fixed and switches the other two; ψ_ℓ encodes a rotation (b, c, a) of the letters a, b, c to the left in the sense that we read the sequence (b, c, a) as the instruction a goes where b

was, b goes where c was, and c goes where a was, which is now really a rotation to the left), and ψ_r moves them according to the instruction the right to make (c, a, b).

The notation is intentionally reminding you of Sym(△). Indeed, if you align ψ_x in $\mathrm{Aut}(KV_4)$ with $x \in \mathrm{Sym}(\triangle)$ then you find this to be an isomorphism (see Definition IV.7 below): it is a one-to-one correspondence between the elements of Sym(△) and the elements of KV_4. For example, ψ_ℓ after ψ_a first sends a to a, and then to b. And it sends b first to c and then that c is sent to a. So $\psi_\ell\psi_a$ is $e \mapsto e, a \mapsto b, b \mapsto a, c \mapsto c$. This is the same effect as that of psi_c, and so $\psi_{\ll}\phi_a = \psi_c$. If we compare to the Cayley table of Sym(△) then we also have correspondingly $\ell a = c$. Checking the entire list of products, we see $\mathrm{Aut}(KV_4) = \mathrm{Sym}(\triangle)$.

5. Free groups

Definition II.23. A group is *free* (on the elements $g_1, \ldots, g_k$) if, for some $k \in \mathbb{N}$, it is isomorphic to the group F_k of all words in the letter set $L_k = \{e, x_1, \ldots, x_k, y_1, \ldots, y_k\}$ with the rules (and no other rules) of

- $ez = z = ze$ for all $z \in L$;
- $x_i y_i = e = y_i x_i$ for all $1 \leq i \leq k$;
- associativity.

Here, the group operation is simply writing two words next to each other in the given sequence.

These groups are "free" because their elements have no other constraints aside from the group axioms. They are not Abelian for $k > 1$ (since we do not require $x_i x_j = x_j x_i$). In contrast, $F_1 = \{\ldots, y_1^2, y_1, e, x_1, x_1^2, \ldots\}$ is isomorphic to the Abelian group $(\mathbb{Z}, +)$ via the identificatiuon $x_1^k \leftrightarrow k \in \mathbb{Z}, y_1^k \leftrightarrow -k \in \mathbb{Z}$.

There are also free groups on infinite numbers of letter. We will not look at them much.

It is a fact that all subgroups of a free group are free (basically, because there are no relations, but the proof is not so easy), and somewhat shockingly, F_2 contains subgroups isomorphis to $F_3, F_4, \ldots$. We won't discuss this phenomenon.

It is also a fact that one can take any group G and interpret it as a free group "with extra rules".

Definition II.24. If G is a group we call a list L of elements a *generating set* if every element of G is a product of elements from $L \cup L'$ where L' is the list of inverses of L.

If such list has been chosen, we refer to elements of L as *generators*.

Evidently, $L = G$ is a generating set, although usually not an interesting one.

Example II.25. $\mathbb{Z} \times \mathbb{Z}$ is generated by $\{(1, 0), (0, 1)\}$. Because of this we can view $\mathbb{Z} \times \mathbb{Z}$ as "F_2 with the additional rules $x_1 x_2 = x_2 x_1$ and $x_1 y_2 = y_2 x_1$ and $x_2 y_1 = y_1 x_2$ and $y_1 y_2 = y_2 y_1$.

To se this note first that we get the relations $x_1 y_1 = y_1 x_1$ and $x_2 y_2 = y_2 x_1$ for free, because all four of these products give e.

Secondly, we read x_1 as $(1, 0)$, and x_2 as $(0, 1)$, which then suggests y_1 is $(-1, 0)$ and y_2 is $(0, -1)$. Then all additional rules imposed on F_2 above correspond to $\mathbb{Z} \times \mathbb{Z}$ being Abelian.

CHAPTER III

Week 3: $\mathbb{Z}/n\mathbb{Z}$ and cyclic groups

The main hero in this week is the group $\mathbb{Z}/n\mathbb{Z}$ with addition, where $n \in \mathbb{N}$. Recall that it is a cyclic group, generated by the coset of 1. The order of the element $1 + n\mathbb{Z}$ is n as one easily sees, and the order of $\mathbb{Z}/n\mathbb{Z}$ is also n.

All groups $\mathbb{Z}/n\mathbb{Z}$ are Abelian, because $\mathbb{Z}$ is Abelian and we just install new rules in order to make $\mathbb{Z}/n\mathbb{Z}$ from $\mathbb{Z}$.

1. Subgroups of cyclic groups

We want to study first how different the elements in $\mathbb{Z}/n\mathbb{Z}$ are for the purpose of generating subgroups.

EXAMPLE III.1. Let $G = \mathbb{Z}/12\mathbb{Z}$. We check for each element of G what group it generates inside G. We find:

- $1 + 12\mathbb{Z}, 5 + 12\mathbb{Z}, 7 + 12\mathbb{Z}, 11 + 12\mathbb{Z}$ all generate all of G. For example, the multiples of $\overline{7}$ are $\{\overline{7}, \overline{2}, \overline{9}, \overline{4}, \overline{11}, \overline{6}, \overline{1}, \overline{8}, \overline{3}, \overline{10}, \overline{5}, \overline{0}\}$ in that sequence.
- $2 + 12\mathbb{Z}, 10\mathbb{Z}$ both generate the subgroup of cosets of even numbers.
- $3 + 12\mathbb{Z}, 9 + 12\mathbb{Z}$ both generate the subgroup of cosets of numbers divisible by 3.
- $4 + 12\mathbb{Z}, 8 + 12\mathbb{Z}$ both generate the subgroup of cosets of numbers divisible by 4.
- $6 + 12\mathbb{Z}$ generates the subgroup of cosets of numbers divisible by 6.
- $0 + 12\mathbb{Z}$ both generate the subgroup of cosets of numbers divisible by 12.

Note that the elements listed in the same item above always have the same order (this is kind of obvious since the order of an element is precisely the order of the cyclic group it generates, and we hqave grouped in the same item the elements that generate the same group).

Note also that if we had asked "classify the elements of $\mathbb{Z}/12\mathbb{Z}$ by their order", we would have written the same exact list. This is because to each possible subgroup size (namely, $0, 1, 2, 3, 4, 6, 12$) there is exactly one subgroup of that size, even though there are usually several different ways to generate that subgroup.

It is natural to ask at this point how one can predict which elements will generate the same subgroup. But perhaps an easier question is "if I take $k + n\mathbb{Z}$, what is its order?". We now consider these questions. For this we collect some facts.

LEMMA III.2. *If* $\operatorname{ord}(g) = n > 0$ *then the exponents* i *with* $g^i = e$ *are precisely the multiples of* n.

In other words, $g^i = g^j$ *if and only if* $n|(i - j)$.

PROOF. If $i = kn$ then $g^i = (g^n)^k = e^k = e$. Conversely, if $g^i = e$ (and $i > 0$) and also $g^n = e$ then write the gcd of i, n as a linear combination $an + bi$ with

$a, b \in \mathbb{Z}$. Note that this gcd is positive since n, i are. Then compute $g^{an+bi} = (g^n)^a(g^i)^b = e^a e^b = ee = e$. So $\gcd(n, i)$ is an exponent that when used over g gives e. But $n = \text{ord}(g)$ is supposedly the smallest positive exponent of this sort. So, $\gcd(n, i) = n$ and so $n|i$.

For the last part, $g^i = g^j$ implies, when multiplying witht he inverse of g^j, that $g^{i-j} = e$, which then by the first part gives $n|(i - j)$. If on the other hand we have $n|(i - j)$ then $g^{i-j} = e$ and so $g^i = g^j$. □

DEFINITION III.3. If $g \in G$ we write $\langle g\rangle$ for the group of al powers—negative and positive—of g in G. This is the *cyclic subgroup generated by g*.

COROLLARY III.4. *Up to renaming,* $(\langle g\rangle, \cdot)$ *is* $(\mathbb{Z}/\text{ord}(g)\mathbb{Z}, +)$ *in the sense that the renaming identifies the Cayley tables.*

PROOF. Let $n = \text{ord}(g)$. Then we associate to $g^i \in \langle g\rangle$ the element $1 + n\mathbb{Z}$ in $\mathbb{Z}/n\mathbb{Z}$. Then $g^i \cdot g^j = g^{i+j}$ corresponds to $(i + n\mathbb{Z}) + (j + n\mathbb{Z}) = (i + j) + n\mathbb{Z}$, and $g^n = e$ to $\underbrace{(1 + n\mathbb{Z}) + \ldots + (1 + n\mathbb{Z})}_{n \text{ copies}}$. □

The next result then tells us how the groups generated by powers of $g \in G$ will look.

COROLLARY III.5. *Let* $g \in G$*have order* n*. Then the group* $\langle g^k\rangle$ *generated by* g^k *is the same group as the group* $\langle g^{\gcd(n,k)}\rangle$ *that is generated by* $g^{\gcd(n,k)}$*. Moreover, abstractly this group is the same as the cyclic group* $C_{n/\gcd(n,k)}$*.*

PROOF. By the same argument as in the previous proof, $\langle g^k\rangle$ contains $g^{\gcd(n,k)}$, and so also all its powers. Conversely, $\gcd(n, k)$ divides k and so of course $\langle g^{\gcd(n,k)}\rangle$ contains g^k and all its powers. So, the groups $\langle g^k\rangle$ and $\langle g^{\gcd(n,g)}\rangle$ are contained one in the other in both directions and hence equal.

Let $h = g^{\gcd(n,k)}$. What could the order of h be? Write $n = d \cdot \gcd(n, k)$; then $h^d = (g^{\gcd(n,k)})^d = g^n = e$ and so the order of h is no more than d. But if $h^i = e$ for some $i < d$ then we also have $e = (h^i) = (g^{\gcd(n,k)})^i$, and this would contradict $\text{ord}(g) = n$ since $\gcd(n, k) \cdot i < \gcd(n, k) \cdot d = n$. □

We can now complete a table from above on subgroups of $\mathbb{Z}/12\mathbb{Z}$:

$g := k \bmod 12\mathbb{Z}$	size of $\langle g\rangle$	$\gcd(n, k)$	$n/\gcd(k, n) = \text{ord}(k + 12\mathbb{Z})$
$\overline{1}, \overline{5}, \overline{7}, \overline{11}$	12	1	12
$\overline{2}, \overline{10}$	6	2	6
$\overline{3}, \overline{9}$	4	3	4
$\overline{4}, \overline{8}$	3	4	3
$\overline{6}$	2	6	2
$\overline{0}$	1	12	1

Looking at this table, the next natural question is: how do you predict the exponents i that give an equation $\langle g\rangle = \langle g^i\rangle$?

As a starter, let's ask for the generators of $\mathbb{Z}/n\mathbb{Z}$, the guys for which $\langle g\rangle$ is the entire group $\mathbb{Z}/n\mathbb{Z}$. For $n = 12$, the relevant cosets are $\overline{1}, \overline{5}, \overline{7}, \overline{11}$. These are the numbers that are *coprime* to 12. (Note: any representative in $\overline{k}$ is coprime to n when k is coprime to n. For example, $\gcd(5, 12) = 1$ and so also $\gcd(5 + 127 \cdot 12, 12) = 1$ and $5 + 127 \cdot 12$ lives in the same coset as 5.)

The magic therefore lies in coprimeness.

DEFINITION III.6. For $n \in \mathbb{Z}$ let $\phi(n)$ be the *Euler ϕ-functin* that counts the number of cosets in $\mathbb{Z}/n\mathbb{Z}$ that consist of representatives coprime to n.

For example, $\phi(12) = 4$ since modulo 12 the cosets $1+12\mathbb{Z}, 5+12\mathbb{Z}, 7+12\mathbb{Z}, 11+12\mathbb{Z}$ are those that are made of numbers coprime to 12.

LEMMA III.7. *If $G = \langle g \rangle$ is cyclic of order n then the generators of G are exactly the elements g^k with $\gcd(n,k) = 1$.*

PROOF. Any element h of G is some power $h = g^k$ of g since $G = \langle g \rangle$. A generator is an element g^k of G with $\langle g^k \rangle = G$, which is the case exactly when $\mathrm{ord}(g^k) = n$. But $\mathrm{ord}(g^k) = n/\gcd(n,k)$, and so we find that g^k is a generator if and only if $\gcd(n,k) = 1$. So counting the generators is the same as counting the cosets of $\mathbb{Z}/n\mathbb{Z}$ that are mnade of numbers coprime to n. □

We can now move and ask when $\langle g^i \rangle = \langle g^j \rangle$ for some exponents i, j. Since the size of $\langle g^i \rangle$ is $n/\gcd(n,i)$ we find the implication

$$[\langle g^i \rangle = \langle g^j \rangle] \Rightarrow [\gcd(n,i) = \gcd(n,j)].$$

In reverse, if the gcd equality holds, then $\gcd(n,i) = \gcd(n,j)$ is a divisor of j which forces g^j inside $\langle g^{\gcd(n,i)} \rangle$ and so $\langle g^i \rangle = \langle g^{\gcd(n,i)} \rangle$ contains $\langle g^j \rangle$. Exchenging i, j gives the reverse containment, hence an equality.

We have now seen all parts of

THEOREM III.8. *Let g be an element of order n. So $\langle g \rangle$ is C_n up to relabeling.*

1. *Subgroups of cyclic groups are always cyclic.*
2. *For all $i \in \mathbb{Z}$, $\mathrm{ord}(g^i)$ divides n and equals $n/\gcd(n,i)$.*
3. *If $k|n$ then there is a unique subgroup of size k inside $\langle g \rangle$, and it is exactly the set of n/k-th powers $\langle g^{n/k} \rangle$ of $g^{n/k}$.*
4. *If $k|n$ then he number of elements of order k inside $\langle g \rangle$ is equal to ϕk. If $k \not| n$, no elements have order k.*
5. *Obviously, if g^i generates a subgroup of order k then it does not generate a subgroup of order different from k. It follows from the previous item, that $n = \sum_{d|n} \phi(d)$.*

To see the last part in action, look at $\mathbb{Z}/n\mathbb{Z}$. Our table above on elements and groups they generate runs in the left column through all the cosets and puts them into one row if they generate the same subgroup. There are 12 such elements, they get grouped as

$$12 = \underbrace{4}_{=\phi(12)} + \underbrace{2}_{=\phi(6)} + \underbrace{2}_{=\phi(4)} + \underbrace{2}_{=\phi(3)} + \underbrace{1}_{=\phi(2)} + \underbrace{1}_{=\phi(1)}.$$

2. Products and simultaneous modular equations

EXAMPLE III.9.

- $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ is not cyclic, since no element can have order 4.
- $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ is generated by $(1,1)$.

LEMMA III.10. *$G := \mathbb{Z}/n_1\mathbb{Z} \times \mathbb{Z}/n_2\mathbb{Z} \times \ldots \times \mathbb{Z}/n_k\mathbb{Z}$ is cyclic if and only if $\gcd(n_i, n_j) = 1$ for every pair $i \neq j$.*

PROOF. If the gcd condition is in force, take the element $g = (\overline{1}, \ldots, \overline{1})$. Its order is a multiple of every n_i, but as they have no common factor, it is a multiple of the product $n_1 \cdots n_k$, which is $|G|$. But no element can have order greater than $|G|$, so $\mathrm{ord}(g) = n_1 \cdots n_k$ and so g generates G.

On the other hand, any element of G is always of order at most $\mathrm{lcm}(n_1, \ldots, n_k)$, since this power creates the neutral element in very component of the product. If $\gcd(n_i, n_j) > 1$ for any $i \neq j$ then this lcm cannot be the product $n_1 \cdot n_2 \cdots n_k = |G|$, so everyones order is less than $|G|$. So G will then have no element of order $|G|$. $\square$

In particular, this says that a product $\mathbb{Z}/(p_1)^{e_1}\mathbb{Z} \times \mathbb{Z}/(p_2)^{e_2}\mathbb{Z} \times \ldots \times \mathbb{Z}/(p_k)^{e_k}\mathbb{Z}$ for *distinct* primes $p_1 < p_2 < \ldots < p_k$ is cyclic.

Note that our first example showed that distinctness is crucial.

EXAMPLE III.11. Let's try to make this more explicit. We know that $\mathbb{Z}/7\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$ is cyclic, and must be of order $7 \times 5 = 35$. So abstractly we know $\mathbb{Z}/7\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$ is $\mathbb{Z}/35\mathbb{Z}$ in disguise. But can we see that inside $\mathbb{Z}/35\mathbb{Z}$?

We are looking for an identification of $\mathbb{Z}/35\mathbb{Z}$ with the product $\mathbb{Z}/7\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$ that preserves the Cayley table (which means it has to preserve the group operation $+$). Let's make a naïve guess: take $i + 35\mathbb{Z}$ and attach to it the element $(i + 7\mathbb{Z}, i + 5\mathbb{Z})$ in $\mathbb{Z}/7\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$. Surely, this attachment will respect addition since $(i + 35\mathbb{Z}) + (j + 35\mathbb{Z})$ would be attached to $(i + 7\mathbb{Z}, i + 5\mathbb{Z}) + (j + 7\mathbb{Z}, j + 5\mathbb{Z}) = ((i + j) + 7\mathbb{Z}, (i + j) + 5\mathbb{Z})$ as you would expect. We write π for this recipe, $\pi(i + 35\mathbb{Z}) = (i + 7\mathbb{Z}, i + 5\mathbb{Z})$.

(Important note here: in $\mathbb{Z}/35\mathbb{Z}$, we have grouped numbers together into a coset whenever they differ by a multiple of 35. Since multiples of 35 are also multiples of both 5 and 7, we can make "cosets of cosets" and read for example the cosets $3 + 35\mathbb{Z}, 8 + 35\mathbb{Z}, 13 + 35\mathbb{Z}, 18 + 35\mathbb{Z}, 23 + 35\mathbb{Z}, 28 + 35\mathbb{Z}, 33 + 35\mathbb{Z}$ as a partition of the coset $3 + 5\mathbb{Z}$ in $\mathbb{Z}/5\mathbb{Z}$. So, moving from $i + 35\mathbb{Z}$ to $i + 5\mathbb{Z}$ actually makes sense since it does not destroy cosets but preserves them and makes them even larger. So it is actually legal to go from $\mathbb{Z}/35\mathbb{Z}$ to $\mathbb{Z}/5\mathbb{Z}$ by the assignment "$i + 35\mathbb{Z}$ becomes $i + 5\mathbb{Z}$". Same argument for going from $\mathbb{Z}/35\mathbb{Z}$ to $\mathbb{Z}/7\mathbb{Z}$. But you could not, for example, go from $\mathbb{Z}/35\mathbb{Z}$ to $\mathbb{Z}/6\mathbb{Z}$: in $\mathbb{Z}/35\mathbb{Z}$, 3 and 38 belong to the same coset, but in $\mathbb{Z}/6\mathbb{Z}$ they do not. Destroying cosets isnot legal when moving groups about.)

So we have a way to go from $\mathbb{Z}/35\mathbb{Z}$ to $\mathbb{Z}/7\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$. Big question, how do we go back? In other words, given a pair $(a + 7\mathbb{Z}, b + 5\mathbb{Z})$ in $\mathbb{Z}/7\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$, how do we find $i + 35\mathbb{Z}$ such that $(a + 7\mathbb{Z}, b + 5\mathbb{Z}) = \pi(i + 35\mathbb{Z})$?

What we know is that this is supposed to work based on the fact that 5 and 7 are coprime. So $\gcd(7, 5) = 1$ must get used somewhere. The Euclidean algorithm says that there are numbers $x, y \in \mathbb{Z}$ with $1 = 7x + 5y$. (Specifically, $x = -2$ and $y = 3$ works). Then let's consider the number $i = a \cdot y \cdot 5 + b \cdot x \cdot 7$. (That one should look at this is not obvious and only becomes clear after a good number of examples). Then we compute:

$$\begin{aligned}(a \cdot y \cdot 5 + b \cdot x \cdot 7) + 7\mathbb{Z} &= a \cdot y \cdot 5 + 7\mathbb{Z} = a(1 - 7x) + 7\mathbb{Z} = a + 7\mathbb{Z},\\ (a \cdot y \cdot 5 + b \cdot x \cdot 7) + 5\mathbb{Z} &= b \cdot x \cdot 7 + 5\mathbb{Z} = b(1 - 5y) + 5\mathbb{Z} = b + 5\mathbb{Z}.\end{aligned}$$

We have basically proved:

LEMMA III.12. *If m, n are relatively prime and $a, b \in \mathbb{N}$ are given, then the simultaneous equations*

$$\begin{aligned}i \bmod m\mathbb{Z} &= a \bmod m\mathbb{Z},\\ i \bmod n\mathbb{Z} &= b \bmod n\mathbb{Z}\end{aligned}$$

have a solution given by $i = a \cdot y \cdot n + b \cdot x \cdot m$ *where* $1 = mx + ny$. □

REMARK III.13. If three pairwise number are m, n, p given, one can also solve simultaneous equations

$$\begin{aligned} i \bmod m\mathbb{Z} &= a \bmod m\mathbb{Z}, \\ i \bmod n\mathbb{Z} &= b \bmod n\mathbb{Z}, \\ i \bmod p\mathbb{Z} &= c \bmod p\mathbb{Z}. \end{aligned}$$

First deal with two equations, then throw in the last.

3. $U(n)$: Automorphisms of $\mathbb{Z}/n\mathbb{Z}$

We have seen that the generators of $(\mathbb{Z}/n\mathbb{Z}, +)$ are the cosets $a+n\mathbb{Z}$ for elements a that have the property $\gcd(n, a) = 1$. So for example, we can think of $\mathbb{Z}/5\mathbb{Z}$ as the group $\langle 1 + 5\mathbb{Z}\rangle$ generated by $1 + 5\mathbb{Z}$ as we usually do, but also as the group $\langle 3 + 5\mathbb{Z}\rangle$. Abstractly, there is no difference how we think. The two interpretations align any coset $a + 5\mathbb{Z}$ with the coset of $3a + 5\mathbb{Z}$, since we are required to respect group operation $+$ and so $(\underbrace{1 + \ldots + 1}_{a \text{ copies}} + 5\mathbb{Z})$ must correspond to $(\underbrace{3 + \ldots + 3}_{a \text{ copies}} + 5\mathbb{Z})$.

Then this correspondence ψ is as follows:

g	$0 + 5\mathbb{Z}$	$1 + 5\mathbb{Z}$	$2 + 5\mathbb{Z}$	$3 + 5\mathbb{Z}$	$4 + 5\mathbb{Z}$
$\psi(g) = 3g$	$0 + 5\mathbb{Z}$	$3 + 5\mathbb{Z}$	$1 + 5\mathbb{Z}$	$4 + 4\mathbb{Z}$	$2 + 5\mathbb{Z}$

You could think of this as having a clock with 5 hours that fell off the table. Now the clockwork is still ok, but the face is broken. You try to reassemble the face in such a way that the clock still works, but you make a mistake and read "3" as "1" in the dark. It's still a clock with 5 hours, but made for aliens that count $3, 1, 4, 2, 5 = 0$ instead of how we count.

Instead of sending $1 + 5\mathbb{Z}$ to $3 + 5\mathbb{Z}$ we could have taken any other generator. BUT, we could not have send it to $0 + 5\mathbb{Z}$ since that is not a generator.

Going back to gcd tests for being generator, note that $\gcd(ab, n) = 1$ for $a, b \in \mathbb{Z}$ happens if and only if both $\gcd(a, n) = 1$ and $\gcd(b, n) = 1$. We conclude that if we take two generators $a + n\mathbb{Z}$ and $b + n\mathbb{Z}$ of the group $\mathbb{Z}/n\mathbb{Z}$ then their product is another such generator. That leads to the idea of taking the set of generators for $\mathbb{Z}/n\mathbb{Z}$ and to turn it into a group with multiplication.

DEFINITION III.14. Let $n \in \mathbb{Z}$ and define $U(n)$ to the subset of $\mathbb{Z}/n\mathbb{Z}$ whose elements are the cosets $a+n\mathbb{Z}$ with $\gcd(a, n) = 1$. We call $U(n)$ the *n-th unit group.*

Each element $u + n\mathbb{Z}$ of $U(n)$ corresponds to an automorphism of $\mathbb{Z}/n\mathbb{Z}$ that is determined by sending $1 + n\mathbb{Z}$ to $u + n\mathbb{Z}$ and then using additivity.

So multiplication is an operation $\cdot : U(n) \times U(n) \to U(n)$ (by the above gcd considerations) that is associative (because multiplcation of integers is already associative) and there is an identity element for this multiplication process (namely the coset $1 + n\mathbb{Z}$). The interesting claim is that $U(n)$ also has inverses. Namely, if $\gcd(a, n) = 1$ then we know from Euclid's algorithm that there are $x, y \in \mathbb{Z}$ with $ax + ny = 1$. This implies directly that $\gcd(x, n)$ is also 1 (since 1 is a linear combination of x and n and therefore is divided by the actual gcd) and also that $(a + n\mathbb{Z}) \cdot (x + n\mathbb{Z}) = (ax + n\mathbb{Z}) = ((1 - yn) + \mathbb{Z}) = 1 + \mathbb{Z}$. So $x + n\mathbb{Z}$ is an inverse for $a + n\mathbb{Z}$.

So, $U(n)$ is a group, and encodes the automorphisms of $\mathbb{Z}/n\mathbb{Z}$,

$$U(n) = \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z}).$$

You can think of making $U(n)$ from $\mathbb{Z}/n\mathbb{Z}$ by asking "if I want to make a mutliplication group from $\mathbb{Z}/n\mathbb{Z}$, what do I need to do"?

Answer:The new identity will be $1+n\mathbb{Z}$. Wanting inverses forces you to dump $0+n\mathbb{Z}$. And if n divides ab then $(a+n\mathbb{Z})(b+n\mathbb{Z}) = 0+n\mathbb{Z} = (0+n\mathbb{Z})(b+n\mathbb{Z})$ wounld contradict the cancellation property. So all $a+n\mathbb{Z}$ with $\gcd(a,n) > 1$ must also be kicked out.

Here are some examples.

EXAMPLE III.15. (1) if $n = 2$ then $U(n)$ is just the coset $1+2\mathbb{Z}$, which is its own inverse. So, $U(n)$ is up to relabeling the trivial group $\{e\}$.

(2) if $n = 3$, $U(n) = \{\underbrace{1+3\mathbb{Z}}_{e}, \underbrace{2+3\mathbb{Z}}_{a}\}$ with the rule that $aa = e$. So, $U(3)$ is the same as the group $\mathbb{Z}/2\mathbb{Z}$ and also the same as D_2.

(3) if $n = 4$, $U(n)$ is $\{\underbrace{1+4\mathbb{Z}}_{e}, \underbrace{3+4\mathbb{Z}}_{a}\}$ with the same Cayley table as $U(3)$.

(4) If $n = 5$ then $U(4)$ has $4 = 5-1$ elements (as 5 is prime) and since $2^2 = 4, 2^3 = 8 = 3+5, 2^4 = 16 = 1+3\cdot 5$, every element of $U(5)$ is a power of $2+5\mathbb{Z}$. So, $U(4)$ is cyclic and of order 4, so it must be C_4.

(5) If p is prime then $U(p^k)$ has $p^{k-1}(p-1)$ elements. Indeed, if you want to be coprime to p^k all you need to do is not have p as a factor. So out of any p consecutive numbers, only $p-1$ will make it into $U(n)$. Since $\mathbb{Z}/p^k\mathbb{Z}$ has p^k elements, $U(p^k)$ will have $p^k \cdot \frac{p-1}{p}$ elements.

(6) If p is a prime number then of corse $U(p)$ has $p-1$ element. We will see later that $U(p)$ is always cyclic. In fact, unless $p = 2$ we will also see that $U(p^n)$ is cyclic. (Recall from above that in contrast $U(4)$ is not cyclic, and in fact $U(2^k)$ is never cyclic for $k > 1$).

There are lots of non-prime numbers n for which $U(n)$ is cyclic.

EXAMPLE III.16. For example, $U(6)$ is $\mathbb{Z}/2\mathbb{Z}$. Let's try to understand that. Recall from last time that we proved that there is an assignment $\psi\colon \mathbb{Z}/6\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$ that sends the coset $a+6\mathbb{Z}$ to the coset pair $(a+2\mathbb{Z}, b+3\mathbb{Z})$ and that this map respects addition and multiplication, and that it is bijective. Since $\mathbb{Z}/6\mathbb{Z}$ is cyclic, generated for example by $1+6\mathbb{Z}$, we conclude that $\mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$ is also cyclic.

We make this more explicit. If you start with a coprime to 6 then this is the same as saying that a is coprime both to 3 and 2. So, if $a+6\mathbb{Z}$ actually lives in $U(6)$ then $a+2\mathbb{Z}$ lives in $U(2)$ and $a+3\mathbb{Z}$ lives in $U(3)$. This is also true conversely since $\gcd(a,2) = 1$ and $\gcd(a,3) = 1$ implies $\gcd(a,6) = 1$. What this means is that ψ sets up not just a correspondence between $\mathbb{Z}/6\mathbb{Z}$ on one side and $\mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$ on the other, but also that under this identification $U(6)$ corresponds to $U(2)\times U(3)$.

Explicitly, this correspondence relates $1+6\mathbb{Z}$ with $(1+2\mathbb{Z}, 1+2\mathbb{Z})$ and $5+6\mathbb{Z}$ with $(5+2\mathbb{Z}, 5+3\mathbb{Z}) = (1+2\mathbb{Z}, 2+3\mathbb{Z})$.

By making the above paragraphs more abstract (replace 2 by m, and 3 by n), one obtains the following theorem.

THEOREM III.17. *If m, n have $\gcd(m,n) = 1$ then $U(mn) = U(m)\times U(n)$.*

Again, if you have 3 or more coprime numbers, one gets a corresponding results on products of unit groups.

EXAMPLE III.18. How many elements does $U(750)$ have?

The bad way is to write them all out. The enlightened 453 student says: $750 = 3 \cdot 5^3 \cdot 2$, and so $U(750) = U(3) \times U(5^3) \times U(2)$. I know $|U(3)| = 2$, $|U(5^3)| = 5^{3-1}(5-1)$, and $|U(2)| = 1$. Hence $|U(750)| = (2) \cdot (25 \cdot 4) \cdot 1 = 200$.

REMARK III.19. Recall the Euler ϕ-function that counts for $n \in \mathbb{N}$ how many numbers from $1, \ldots, n$ are relatively prime to n. Recall also that a is relatively prime to n if and only if $a + n\mathbb{Z}$ is a generator of the group $\mathbb{Z}/n\mathbb{Z}$. (In other words, the order of $a + n\mathbb{Z}$ is n, or yet in other words, na is the lowest positive multiple of a that is divisible by n).

Since $U(n)$ is made of the cosets of $\mathbb{Z}/n\mathbb{Z}$ that come from numbers relatively prime to n, there are exactly $\phi(n)$ elements in $U(n)$. That means also that if m, n are relatively prime, then $\phi(mn) = \phi(m)\phi(n)$ because of the theorem above.

CHAPTER IV

Week 4: Cosets and morphisms

1. Equivalence relations

DEFINITION IV.1. Let S be a set. An equivalence relation is a binary relation $\simeq$ on S such that

- $a \simeq a$ for all $a \in S$ (*reflexivity*);
- $[a \simeq b] \Leftrightarrow [b \simeq a]$ for all $a, b \in S$ (*symmetry*);
- $[a \simeq b$ and $b \simeq c] \Rightarrow [a \simeq c]$ for all $a, b, c \in S$ (*transitivity*).

Examples of such equivalence relations are :

- the usual equality of numbers
- congruence of geometric figures;
- equality in the module caculation (this is really the relation $i \simeq j$ on $\mathbb{Z}$ whenever $n|(i-j)$.

An example of a realtion that is not an equivalence relation is the usual $\leq$, because it is not symmetric: $3 \leq 4$ but not $4 \leq 3$.

LEMMA IV.2. *If S is a set with equivalence relation $\simeq$ then one can partition S into* cosets/equivalence classes *where any coset contains all the elements that are mutually equivalent to one another.*

If we denote the cosets $S_1, S_2, \ldots$, then we have: $S_i \cap S_j$ is empty unless $S_i = S_j$. Moreover, S is the union of all S_i.

LEMMA IV.3. *Let G be any group, and pick $n \in \mathbb{N}$. Then let A be the collection of all group elements $a \in G$ that have order exactly n. Then $|A|$ is a multiple of $\phi(n)$.*

PROOF. If $a \in A$, then $\langle a \rangle$ is a cyclic group of order n. By last week's results, $\langle a \rangle$ contains exactly $\phi(n)$ elements whose order is exactly n, and these are exactly the generators of $\langle a \rangle$. So, make an equivalence relation on A where $x \simeq y$ if and only if $\langle x \rangle = \langle y \rangle$. Each equivalence class has size $\phi(n)$, they do not meet, and their union is A. So the coset size divides $|A|$. □

Note: if G is cyclic and $n = |G|$, then $G = \langle g \rangle$ and $|A| = \phi(n)$ by last week.

2. Morphisms

Let G, G' be two groups.

DEFINITION IV.4. A *morphism* (or *homomorphism*) is a function $\psi \colon G \to G'$ that respects the group operations:

$$\psi(g_1 \cdot_G g_2) = \psi(g_1) \cdot_{G'} \psi(g_2)$$

for all $g_1, g_2 \in G$.

You have seen many examples already.

DEFINITION IV.5. Denote $\mathbb{R}^\times$ and $\mathbb{C}^\times$ the nonzero real numbers and the nonzero complex numbers respectively.

Here is a list of morphisms that you have seen at least in part.

- $G = \mathbb{Z} = G'$, ψ =multiplication by 42.
- $G = GL(2)$ = the invertible 2×2 matrices with matrix multiplication , $G' = (\mathbb{R}^\times, \cdot)$ and ψ =the determinant: $\det(AB) = \det(A)\det(B)$.
- the exponential map $(\mathbb{R}, +) \to (\mathbb{R}^\times, \cdot)$, since $\exp(x+y) = \exp(x) \cdot \exp(y)$.
- The logarithm function $\ln\colon (\mathbb{R}^\times, \cdot) \to (\mathbb{R}, +)$, since $\ln(a \cdot b) = \ln(a) + \ln(b)$.
- The square root function $\sqrt{\ }\colon (\mathbb{R}^\times, \cdot) \to (\mathbb{R}^\times, \cdot)$, since $\sqrt{a \cdot b} = \sqrt{a} \cdot \sqrt{b}$.
- The third power map $(-)^3\colon (\mathbb{R}^\times, \cdot) \to (\mathbb{R}^\times, \cdot)$, since $(a \cdot b^3 = a^3 \cdot b^3$.
- The third power map $(-)^3\colon U(7) \to U(7)$ since $((a + 7\mathbb{Z})(b + 7\mathbb{Z}))^3 = (a + 7\mathbb{Z})^3 \cdot (b + 7\mathbb{Z})^3$.

EXAMPLE IV.6. Suppose we want to make a morphism $k\colon \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ that sends $a + m\mathbb{Z}$ to $ka + n\mathbb{Z}$. That means that every element of the form $a + tm$ with $t \in \mathbb{Z}$ should be turned by multiplication by k into an element of the form $ka + sn$ where $s \in \mathbb{Z}$.

Let's examine this. For example, if $a = 0$ and $t = 1$ this means that km should look like sn for a suitable $s \in \mathbb{Z}$. This just asks that km is a multiple of n. As one can check, $n|km$ is also enough to make sure everything else goes well. We conclude:

> Multiplication by k sets up a morphism $k\colon \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ if and only if n divides mk.

DEFINITION IV.7. A morphism ψ is a *isomorphism* if it is bijective (= injective + surjective= into + onto). It is an *automorphism* if it is an *isomorphism* where $G' = G$.

Note that this forces $\psi(e_G) = e_{G'}$, since $e_g e_g = e_G$ forces $\psi(e_G)\psi(e_G) = \psi(e_G)$ and (because of the cancellation property in G') such an equation can only be satisfied by $e_{G'}$.

An automorphism is a relabeling of G that preserves the group structure. An isomorphism is a way of linking in tiwn pairs the elements of G and G' while making sure that products work in both groups the same way. That means, an isomorphism is the matching of 2 Cayley tables, and an automorphism is a switch in columns and rows of a Cayley table that reproduces the same Cayley table.

For example, $\exp\colon (\mathbb{R}, +) \to (\mathbb{R}^\times, \cdot)$ is an isomorphism, but not an automorphism. The map the multiplies by 5 is an automorphism of $\mathbb{Z}/12\mathbb{Z}$ (since it sends the generator $1 + 12\mathbb{Z}$ to the generator $5 + 12\mathbb{Z}$).

Here is a list of things an isomorphism ψ needs to do/have. This is a good list for checking whether isomorphisms between G and G' can exist at all.

- $|G| = |G'|$;
- both G and G' are cyclic, or neither one is cyclic;
- both are Abelian, or neither is;
- the number of elements of G that have order k is the same as the number of elements of G' that have order k (for any k).

EXAMPLE IV.8. If G, G' are both cyclic of the same order n then they are isomorphic. Namely, if $G = \langle g \rangle$ and $G = \langle g' \rangle$, let the morphism send g^i to $(g')^i$.

We saw last week that the automorphisms of $\mathbb{Z}/n\mathbb{Z}$ are labelled by the cosets $k + n\mathbb{Z}$ with $\gcd(k, n) = 1$. In other words,

$$\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z}, +) = U(n).$$

EXAMPLE IV.9. While $U(12)$ and $U(10)$ both have 4 elements, they are not isomorphic. Indeed, $U(10)$ is cyclic generated by 3, and $U(12)$ has no element of order 4.

One natural way of making automorphisms is the following:

DEFINITION IV.10. Let $a \in G$ be a group element. Define a map $\psi_a \colon G \to G$ by setting

$$\psi_a(g) = aga^{-1}.$$

This is the *inner automorphism on G induced by a.*

Note first that this indeed respects multiplication: $\psi_a(g_1)\psi_a(g_2) = ag_1a^{-1}ag_2a^{-1} = ag_1g_2a^{-1} = \psi_a(g_1g_2)$.

Note next that if G is Abelian, then $\psi_a(g) = aga^{-1} = aa^{-1}g = ea = a$ for any choice of g and a. So in an Abelian group, every inner automorphism is the identity map.

EXAMPLE IV.11. If $G = \mathrm{Sym}(\triangle)$, then the 6 inner automorphisms are as follows:

- ψ_e fixes every element;
- ψ_ℓ sends $e \to e, r \to r, \ell \to \ell, a \to c, b \to a, c \to b$;
- ψ_r sends $e \to e, r \to r, \ell \to \ell, a \to b, b \to c, c \to a$;
- ψ_a, ψ_b, ψ_c are quite similar: ψ_x fixes e, r, ℓ, x and interchanges the two remaining elements of G.

If ψ_x, ψ_y are inner automorphisms of G, they can be composed: $\psi_x(\psi_y(z)) = xyzy^{-1}x^{-1} = \psi_{xy}(z)$. Thus, we find:

LEMMA IV.12. *The assignment $a \mapsto \psi_a$ is a morphism inn_G from the group G to the group of its inner automorphisms $\mathrm{Inn}(G)$.*

We will see later that since $\mathrm{Inn}(\mathrm{Sym}(\triangle))$ has 6 different elements just like $\mathrm{Sym}(\triangle)$ itself, then the conclusion is that inn_G is actually an isomorphism, so that as abstract groups there is no difference between $\mathrm{Sym}(\triangle)$ and $\mathrm{Inn}(\mathrm{Sym}(\triangle))$.

3. Cosets for subgroups

DEFINITION IV.13. Let H be a subgroup of G and choose $g \in G$. We write gH for the set of all products gh with $h \in H$. We call gH the *left coset* of H to g. The set of products Hg is the *right coset* of H to g.

Note that if G is Abelian, then left and right cosets agree, $gH = Hg$. Note also that (because of the cancellation property) gH and Hg contain equally many elements, namely $|H|$ many.

EXAMPLE IV.14. Let $G = \mathrm{Sym}(\triangle)$ and $H = \{e, a\}$. Then $e \cdot H = a \cdot H = H = \{e, a\}$, $r \cdot H = b\dot{H} = \{b, r\}$, and $\ell \cdot H = c \cdot H = \{c, \ell\}$.

Note that these are disjoint, one of them is H, and their union is G.

These observations generalize as follows.

LEMMA IV.15. *Let H be a subgroup of G. For all $a, b \in G$ we have:*

1. *$a \in aH$ (since $1 \in H$).*
2. *aH meets H if and only of $a \in H$ (since $ah = h'$ gives $a = h'h^{-1}$).*
3. *$aH = bH$ or $aH \cap bH = \emptyset$ (since $c = ah = bh'$ implies $b^{-1}a = h'h^{-1} \in H$ and so $b^{-1}aH = H$ by the previous item, and so $aH = bH$.*
4. *$|aH| = |H|$ (since cancellation dictates that the map $h \to ah$ is injective, and so is the map $ah \to ahh^{-1} = a$).*
5. *From the definitions, $aH = Ha$ if and only if $aHa^{-1} = H$ if and only if H is stable (as set, not necessarily element by element) under the inner automorphism ψ_a.*
6. *aH is a subgroup of G iff $a \in H$ (since a subgroup needs e, and $e \in aH$ means $a^{-1} \in H$, hence $a \in H$).*

The main upshot of this lemma is

THEOREM IV.16. *Let G be a finite group, H a subgroup. Then $|H|$ divides $|G|$, and the number of left cosets of H is equal to $|G|/|H|$.*

PROOF. The various cosets gH with $g \in G$ are either equal to one another, or disjoint. So G is the disjoint union of a bunch of cosets, and they all have size $|H|$ but the lemma. □

DEFINITION IV.17. The quotient $|G|/|H|$ from the theorem is denoted $[G : H]$ and called the *index of H in G*.

COROLLARY IV.18 (Lagrange's Theorem). *If $g \in G$ then the order of g divides the order of G.*

PROOF. The cyclic group $\langle g \rangle$ is a subgroup of G. It has $\mathrm{ord}(g)$ elements and by the theorem this number divides $|G|$. □

COROLLARY IV.19. *If a group has a prime number of elements, it must be cyclic.*

PROOF. Take an element $g \in G$. Its order divides $|G|$, which is supposed to be prime. So the choices are $\mathrm{ord}(g) = |G|$ or $\mathrm{ord}(g) = 1$. In the second case, $g = e$ must be the identity. So, take another element that is not the identity. Now $\mathrm{ord}(g)$ cannot be 1, so it must be $|G|$ as in the first case. But if an element has order $|G|$ then the cyclic subgroup it generates has $|G|$ elements, and that means it fills out G completely. So $G = \langle g \rangle$ for any g different from e. □

THEOREM IV.20 (Fermat's little theorem). *If p is a prime number then $p|(a^{p-1} - 1)$ for all $a \in \mathbb{Z}$. In particular, $a^p_p\mathbb{Z} = a \bmod p\mathbb{Z}$.*

PROOF. The group of units $U(p)$ has $p - 1$ elements. That means, that the order of every element in $U(p)$ divides $p - 1$. In other words, g^{p-1} is the identity for all $g \in U(p)$. Unraveling this gives $a^{p-1} + p\mathbb{Z} = 1 + p\mathbb{Z}$ for all $a \in \mathbb{Z}$. The second statement follows from multiplication by a. □

Note that p prime is essential: Fermat's little theorem fails for $p = 4$. Question: are there non-primes for whch this does work?

4. Kernels and normal subgroups

Recall from homework the concept of conjugation:

DEFINITION IV.21. If H is a subgroup of G, and if $a, g \in G$ then the *conjugate* of g with respect to a is the element aga^{-1}. The set of conjugates of elements of H, aHa^{-1}, is the *conjugate* of H with respect to a.

We note that $aga^{-1}ag'a^{-1} = agg'a^{-1}$ and so that products of conjugates are conjugates. In particular, a conjugates e_G to e_G and the inverse of aga^{-1} is $ag^{-1}a^{-1}$. In fact, conjugation by a provides an inner automorphism ψ_a of G and aHa^{-1} is a subgroup of G.

DEFINITION IV.22. Suppose $\phi\colon G \to G'$ is a morphism. Set $\ker(\phi) = \{g \in G | \phi(g) = e_{G'}\}$ be the *kernel* of ϕ.

THEOREM IV.23. *For any morphism $\phi\colon G \to G'$, the kernel $\ker(\phi)$ is a subgroup of G. Moreover, $\ker(\phi)$ is stable under all inner automorphisms ψ_x for $x \in G$.*

PROOF. For being a subgroup we need to show that $\ker(\phi)$ is closed under G-multiplication, and under taking inverses. We will use that we already proved that a morphism must take e_G to $e_{G'}$. I will be very explicit about where multiplications happen, in G or in G'.

So let $g_1, g_2 \in \ker(\phi)$. By definition that means $\phi(g_1) = \phi(g_2) = e_{G'}$, the identity in G'. The morphism property then gives $\phi(g_1 \cdot_G g_2) = \phi(g_1) \cdot_{G'} \phi(g_2) = e_{G'} \cdot_{G'} e_{G'} = e_{G'}$. Moreover, if $g \in G$ then $e_{G'} = \phi(e_G) = \phi(g \cdot_G g^{-1}) = \phi(g) \cdot_{G'} \phi(g^{-1})$ which shows that $\phi(g)$ and $\phi(g^{-1})$ are always inverse to one another in G'. In particular, if $\phi(g) = e_{G'}$ then the same is true for g^{-1}. That shows that if $g \in \ker(\phi)$ then $g^{-1} \in \ker(\phi)$. We have therefore shown that $\ker(\phi)$ is a subgroup that we denote H for brevity in the rest of the proof.

Now consider conjugation by $x \in G$. All elements of xHx^{-1} have the form xgx^{-1} with $g \in \ker(\phi)$, so $\phi(g) = e_{G'}$. Then we need to show that xgx^{-1} is also in the kernel of ϕ. So we test it: $\phi(x \cdot_G g \cdot_G x^{-1}) = \phi(x) \cdot_{G'} \phi(g) \cdot_{G'} \phi(x^{-1}) = \phi(x) \cdot_{G'} e_{G'} \cdot_{G'} \phi(x^{-1}) = \phi(x) \cdot_{G'} \phi(x^{-1}) = e_{G'}$. So, indeed we have $xgx^{-1} \in H$. So H is stable under conjugation. □

DEFINITION IV.24. If $H \subseteq G$ is a subgroup that is stable under all inner automorhisms, we call H a *normal subgroup*.

As a side remark, this is not "normal" behavior in the usual sense of language. Looking at all subgroups H of a given group G, it is usually quite unnormal for H to be normal. Normal subgroups are quite special.

Note that $aHa^{-1} = H$ is equivalent to $aH = Ha$ so that left and right cosets agree for each $a \in G$ precisely when H is normal.

EXAMPLE IV.25. The kernel of any morphism is normal as we proved in the theorem above.

EXAMPLE IV.26. The subgroup $\{e, a\} \subseteq \mathrm{Sym}(\triangle)$ is not normal. Indeed, $a \in H$ but $\ell a \ell^{-1} = c\ell^{-1} = cr = b$ is not in H.

One can check that there are not many normal subgroups of $\mathrm{Sym}(\triangle)$: the only ones stable unde all conjugations are the trivial group $\{e\}$, the rotation subgroup, and the whole group. (These latter two are always normal and never interesting as subgroups).

REMARK IV.27. A subgroup is normal if and only if the left cosets aH agree with the right cosets Ha. This follows because $[aH = Ha] \Leftrightarrow [aHa^{-1} = H]$ as one sees by multiplying with a^{-1} on the right.

EXAMPLE IV.28. Let G be the 2×2 invertible matrices with real entries, with matrix multiplication as group operation. Let $\phi\colon G \to \mathbb{R}^\times$ be the morphism that takes determinants. Linear algebra says that $\det(ABA^{-1}) = \det(A)\det(B)/\det(A) = \det(B)$. So if $\det(B) = 1$ then this is also true for all its conjugates.

CHAPTER V

Week 5: Permutations and the symmetric group

DEFINITION V.1. The symmetric group S_n is the group of all permutations on n elements. It makes no difference what the permuted n things are. We usually assume they are the numbers $1 \ldots, n$.

Note that S_n has $n!$ elements.

EXAMPLE V.2. We have met S_3 as Sym($\triangle$). We denote the elements of S_n by arrays. For example, if our triangle has the letters A, B, C written on the vertices in counterclockwise order, then we have the correspondence

$$e \leftrightarrow \begin{pmatrix} A & B & C \\ A & B & C \end{pmatrix}, r \leftrightarrow \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix}, l \leftrightarrow \begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix}$$
$$a \leftrightarrow \begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix} b \leftrightarrow \begin{pmatrix} A & B & C \\ C & B & A \end{pmatrix} c \leftrightarrow \begin{pmatrix} A & B & C \\ B & A & C \end{pmatrix}$$

between symmetries of the triangle (on the left) and the permutations of A, B, C.

The meaning of for example $\begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix}$ is that the lower row indicates the letter that is replaced by the one on top of it. So, B (below) is replaced by A (above it) and so on. Another way of saying this is: the letter A (above) moves to where the letter B (below A) was. Better yet is to say "what used to be in the A-bucket is now moving in the B-bucket.

If one composes, one gets for example

$$ra = b \leftrightarrow \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix} \begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix} = \begin{pmatrix} A & B & C \\ C & B & A \end{pmatrix}.$$

It takes some practice to read this product correctly. The important bit is that one again carries it out *right to left*. So, if you want to know what this product on the right does to the letter B you first check what the right factor $\begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix}$ does to B. And you find, it sends it to where C used to be. So B is now in bucket 3. Next, you ask what the left factor $\begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix}$ does to stuff in bucket 3, so you look at the column labeled C. And it says that stuff in bucket 3 is being moved to bucket 2, since under the C is a B. So, combining both steps, B first moves to bucket 3 and then back to bucket 2. So in the product one should have B above B, which is exactly right.

Similarly, A in the right factor moves to bucket 1, and then with the left factor to bucket 3. So A should be above C in the product. Finally, C moves with the right factor to bucket 2, and then with the left factor to bucket 1. So, C in the product should stand above A.

DEFINITION V.3. There is another way to wrote permutatins called *cycle notation*. You start with some letter (A, and then record where A goes. For example, for the right rotation $r = \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix}$ we write down $(A, C$. Next you ask where C goes, and under r that is B. So we continue to $(A, C, B$. But B now is moved to A and that "closes the cycle, so we write (A, B, C).

If a cycle closes before you have witten down what happens to all elements, just open another cycle. So, the permutation $\begin{pmatrix} A & B & C & D & E & F \\ B & C & A & D & F & E \end{pmatrix}$ has cycle notation $(A, B, C)(D)(E, F)$. It rotates A, B, C in a 3-cycle and also rotates E, F in a 2-cycle, and leaves D put.

One may or may not indicate 1-cycles (since they are talking about elements that do not move, the assumption is that if an element does not show in a cycle that you wrote, then it is not moving. For example $(1, 3, 5)$ is a permutation that leaves 2 and 4 fixed.)

A cycle of length 2 is a *transposition*.

How does one compose cycles? Just the same as always: start on the right. So, $(1, 4, 5)(2, 3, 4, 1)(3, 5)$ is decoded as follows. Start with 1. under (3,5) it goes to position 1, then 1 goes under (2,3,4,1) to position 2. So the 1 we started with is now in position 2. Stuff in position 2 moves under (1,4,5) not al all, so position 2 is the final destination of 1. So we start writing the prouct as $(1, 2$.

Next we redo this all with input 2. Under (3,5), 2 stays put. Under (2,3,4,1) stuff in bucket 2 moves to bucket 3. And then under (1,4,5) stuff in bucket 3 stays put. So overall, 2 moves to bucket 3. So we are now at $(1, 2, 3$.

Restart with input 3. Under (3,5), 3 moves to bucket 5, and under (2,3,4,1) bucket 5 stays put. The at the end (1,4,5) mive bucket 5 to bucket 1, and that means our 3 lands in bucket 1. So, we have found the first part of the product cycle as (1,2,3).

This does not yet explain what happens to 4 and 5 under the product. Let's check 4. Under (3,5) the bucket 4 stays put. Then it is moved to bucket 1 under (2,3,4,1). And bucket 1 is moved to bucket 4 under (1,4,5). Hence the number 4 stays put overall. That means, 5 also must stay put since there is no more open space. So, the product is $(1, 2, 3)(4)(5)$.

REMARK V.4.

- What our product procedure produces is *disjoint cycles*. That is, the cycles we write down as anser are such that no number occurs in more than one cycle. Disjoint cycles are prefereable since we "understand" better.
- For example, the order of any cycle (in the group theory sense) is its own length: if you rotate left a bunch of k people on a round table, you need to repeat this k times until everyone is back to his own seat. Moreover, if you have the product of a bunch of disjoint cycles, then the order of this product is th lcm of the cycle lengths. For example, the order of (1,2,3)(4,5) is 6, because only iteration multiples of 3 make 1,2,3 go back home, and only even numbers of iterations make 4,5 go back home.

  In contrast, (1,2,3)(2,3,4)=(1,3)(2,4) has order 2, not 3 (the lack of disjointness messes with thing!)

THEOREM V.5. *Any permutation is a product of 2-cycles (usually not disjoint!).*

PROOF. It is enough to show that any single cycle can be made from 2-cycles. If $n = 2$ this is clear (except that you need to say that the identity is writeable as $(1,2)(1,2)$.)

If $n \geq 3$ check that $(a_1, \ldots, a_k) = (a_1, a_3, \ldots, a_k)(a_1, a_2)$. So the theorem follows from induction. □

LEMMA V.6. *If you take a permutation σ and write it as product of permutations, then the number of transposition is not determined by σ, but the number of permutations for the same σ is either always odd or always even.*

Before we embark on a proof, one more concept:

DEFINITION V.7. Let σ be a permutation of $1, \ldots, n$. We say that $[i,j]$ is a *switch* of σ if $i < j$ but σ places i in a position behind where it places j. In other woords, if you write $\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma_1 & \sigma_2 & \ldots & \sigma_n \end{pmatrix}$ then $i < j$ but $\sigma_i > \sigma_j$.

The *disorder* of σ is the number of switches of σ. The *parity* of σ is the answer to the question "Is the disorder of σ even or odd'?"

For example, the cycle (1,2,3,4) is the permutation $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$ and so has switches $[1,4], [2,4], [3,4]$ and so has disorder 3 and is an odd permutation (has odd parity)

Proof of the lemma: It is enough to prove that the identity cannot be written as the product of an odd number of transpositions. (Since $e = (1,2)(1,2)$ is an even way of writing the identity).

The main idea is that if some σ is composed with a transposition then its parity changes. Let's check that. So we imagine, for $1 \leq i < j \leq n$, composing the transposition (i,j) with the permutation $\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma_1 & \sigma_2 & \ldots & \sigma_n \end{pmatrix}$ and we count the change in the disorder.

Let's say the output of σ is the sequence

$$s_1, \ldots, s_{i-1}, s_i, s_{i+1}, \ldots, s_{j-1} s_j, s_{j+1}, \ldots, s_n.$$

Then the output of $(i,j)\sigma$ is

$$s_1, \ldots, s_{i-1}, s_j, s_{i+1}, \ldots, s_{j-1} s_i, s_{j+1}, \ldots, s_n.$$

We consider the change in the number of switches.

If a switch of σ does not involve i nor j then it is a switch also of the composition $(i,j)\sigma$. So we need to focus on switches that involve either i or j or both. We next study when a switch involves one of i, j.

If $k < i$, the number of s_t that appear to the right of s_k but are smaller than s_k does not change if we interchange s_i ith s_j.

If $k > j$ then the number of s_t that are to the right of s_k and are smaller than s_k chenges even less.

If $i < k < j$, there are 4 cases:

(1) If $s_k < s_i$ and $s_k < s_j$ then $[k,i]$ is a switch in σ and $[k,j]$ is a switch in $(i,j)\sigma$.
(2) If $s_k < s_i$ and $s_k > s_j$ then $[k,i], [k,j]$ are a switch in σ and neither of $[k,i], [k,j]$ is a switch in $(i,j)\sigma$.

(3) If $s_k > s_i$ and $s_k < s_j$ then neither of $[k,i],[k,j]$ is a switch in σ and both $[k,i],[k,j]$ are a switch in $(i,j)\sigma$.
(4) If $s_k > s_i$ and $s_k > s_j$ then $[k,j]$ is a switch in σ and $[k,i]$ is a switch in $(i,j)\sigma$.

In all cases then, so far, the change of the number of switches in σ versus $(i,j)\sigma$ is even.

Finally, consider the pair i,j. If it is not a swutch for σ then it must be one for $(i,j)\sigma$, and conversely. So overall, the number of switches is an even number *plus one*, and hence odd.

What this means is that if you write any σ as product of transpositions, the number of transpositions must agree with the parity of σ (which does not depend on how you write σ as such product!). So, the number of transpositions used is even if and only of the parity of σ is even. □

DEFINITION V.8. Let A_n be the *altermating group* of all even permutations.

Note that we just proved that this definition makes sense since products of even permutations are even (just as odd times even or even times odd is odd, and odd times odd is even).

Now recall the Cayley table to the group D_2, and line it up as $e \leftrightarrow$even $f \leftrightarrow$odd. Note that D_2 can be viewed as $(\mathbb{Z}/2\mathbb{Z},+)$. That means there is a morphism

$$\text{sign}\colon S_n \to (\mathbb{Z}/2\mathbb{Z},+)$$

that sends even permutations to 0 mod $2\mathbb{Z}$, odd permutations to $1+2\mathbb{Z}$, and turns composition of permutations into addition of signs. The kernel of this morphism is A_n.

EXAMPLE V.9. For $n=3$, A_n =the rotations. Note that indeed composition of A_n-elements (rotations) gives you other A_n-elements (rotations).

Note that for $n>3$ the rotations do not fill out A_n, although they do belong into A_n. For example, $(1,2)(3,4)$ is not a rotation but still even.

The following explains the special position of permutation groups.

THEOREM V.10. *Any group G can be viewed as a permutation group.*

PROOF. Take the base set for the permutations to be the elements of G. Then take $g \in G$ and read it as a permutation σ^g by recording in the permutation σ^g the products $g\cdot g'$, letting g' run through all of G. By the cancellation property you do indeed get a permutation. Multiplication in G then corresponds to composition of permutations. □

EXAMPLE V.11. Recall that KV_4 is the symmetry group of the letter H. We can make it a subgroup of S_4 as follows. Take as symbols of the group the "letters" $e, \leftrightarrow, \updownarrow, \curvearrowleft$. Now ask what the effect of multiplying by the group elements is on the sequence $\{e, \leftrightarrow, \updownarrow, \curvearrowleft\}$. We find

$$\begin{aligned} e\cdot\{e,\leftrightarrow,\updownarrow,\curvearrowleft\} = \{ee, e\leftrightarrow, e\updownarrow, e\curvearrowleft\} &= \{e,\leftrightarrow,\updownarrow,\curvearrowleft\} \\ \updownarrow\cdot\{e,\leftrightarrow,\updownarrow,\curvearrowleft\} &= \{\updownarrow,\curvearrowleft,e,\leftrightarrow\} \\ \leftrightarrow\cdot\{e,\leftrightarrow,\updownarrow,\curvearrowleft\} &= \{\leftrightarrow,e,\curvearrowleft,\updownarrow\} \\ \curvearrowleft\cdot\{e,\leftrightarrow,\updownarrow,\curvearrowleft\} &= \{\curvearrowleft,\updownarrow,\leftrightarrow,e\}. \end{aligned}$$

If one now reads these as permutations, one can write them as cycles as:

$$\begin{array}{rcl} e & \text{becomes} & () \\ \updownarrow & \text{becomes} & (e, \updownarrow)(\leftrightarrow, \curvearrowleft) \\ \leftrightarrow & \text{becomes} & (e, \leftrightarrow)(\updownarrow, \curvearrowleft) \\ \curvearrowleft & \text{becomes} & (e, \curvearrowleft)(\updownarrow, \leftrightarrow). \end{array}$$

If one translates into symbols $1, 2, 3, 4$ we get

$$KV_4 = \{(), (1,2)(3,4), (1,3)(2,4), (1,4)(2,3)\}.$$

REMARK V.12. In many cases, there is a more obvious way of embedding a given group into a symmetric group. For example, the symmetry group of a cube is a naturally a subgroup of S_8 since the symmetries of the cube move around the 8 vertices. But that is sort of an accident: not every group *comes to us* as the symmetry group of a small set of things (with certain constraints). If someone hands us the symmetry group of a cube without saying what it really is, and if we don't notice it, we would have to take recourse to the recipe of the proof of the proposition. And that would view the symmetry group of the cube (with 48 elements) as a subgroup of S_{48}, a rather unpleasant idea. So the proposition conveys a principle, but it pays to be opportunistic.

CHAPTER VI

Week 6: Quotients and the Isomorphism Theorem

Let me start with recalling some ideas from the past. If $H \subseteq G$ is a subgroup (same identity element, same multiplication) then is is a normal subgroup if it has no conjugate subgroups aside form itself. This is saying, that $aHa^{-1} = H$ (or $aH = Ha$)for any $a \in G$. Note that this says that aha^{-1} is again in H, but it does not require that $aha^{-1} = h$ (but it also does not say this should not be true)

Let us also recall that H can be used to make H-formed clusters in G by looking at the left cosets aH; any element of G belongs to one such coset, so their union is all of G, and two cosets either do not meet at all, or agree completely. No partial agreement is possible (because of the cancellation property).

1. Making quotients

DEFINITION VI.1. Let us denote the collection of all left H-cosets in G by G/H.

Note the similarities: when $G = \mathbb{Z}$ is all integers, and $H = n\mathbb{Z}$ the subgroup of integers divisible by n then $G/H = \mathbb{Z}/n\mathbb{Z}$ is exactly the collection of cosets $a + n\mathbb{Z}$ with $a \in \mathbb{Z}$.

Note also that $\mathbb{Z}/n\mathbb{Z}$ is a group itself; we would like to arrange for G/H to be a group as well. The natural plan would be to define $(aH) * (bH) = abH$. Lets look in an example what that is like.

EXAMPLE VI.2. Let $G = S_4$ be the symmetry group of the equilateral tetrahedron (also known as the permutation on 4 elements) and take as H the group of permutations $\{(), (12)(34), (13)(24), (14)(23)\}$. We saw at the end of last class that this group can be identified with KV_4, the symmetry group of the letter H.

As S_4 has 24 elements and H has 4, the clusters we make for cosets have size 4 and there will be 6 such cosets. They are: (no shortcut here, I just sat down and computed each set aH by hand):

$$\begin{aligned} E := H &= \{(), (1,2)(3,4), (1,3)(2,4), (1,4)(2,3)\}, \\ \gamma := (12)H &= \{(1,2), (34), (1,3,2,4), (1,4,2,3)\}, \\ \beta := (13)H &= \{(1,3), (1,2,3,4), (2,4), (1,4,2,3)\}, \\ \alpha := (14)H &= \{(1,4), (1,2,4,3), (1,3,2,4), (2,3)\}, \\ \lambda := (123)H &= \{(1,2,3), 1,3,4), (2,4,3), (1,4,2)\}, \\ \rho := (124)H &= \{(1,2,4), (1,4,3), (1,3,2), (2,3,4)\}. \end{aligned}$$

Now we would like to make these 6 clusters into a group. As mentioed above, we aim for $(aH)(bH) = abH$. In order to avoid problems such as we met in Assignment 4a when we were looking at morphisms from $\mathbb{Z}/m\mathbb{Z}$ to $\mathbb{Z}/n\mathbb{Z}$ that were not even functions (because they destroyed cosets), we need to keep cosets together. More

explicitly, we need that for all choices of $a, a' \in G$ and $h, h' \in H$ we have that $ghg'h' \in gg'H$. (Multiplication should not depend on the specific representative we picked; if it does, multiplication would destroy cosets). But

$$\begin{aligned} & [ghg'h' && \in && H] \\ \Leftrightarrow\ & [hg'h' && \in && g'H] (\text{ cancelling a } g) \\ \Leftrightarrow\ & [g'^{-1}hg'h' && \in && H] (\text{ left-multiply by } g'^{-1}) \\ \Leftrightarrow\ & [g'^{-1}hg' && \in && H] (\text{ right-multiply by } h'^{-1}) \\ \Leftrightarrow\ & [aHa && \in && H] (\text{ renaming } g' \text{ to } a) \\ \Leftrightarrow\ & [&& && H \text{ is normal in } G. \end{aligned}$$

So, we should check whether H is normal. Since every element of S_4 is a product of transpositions (i, j), we do not need to test all 24 elements $a \in G$ whether $aH = Ha$, but only tr transpositions. And since H stays H when you arbitrarily rename the permuted objects $1, 2, 3, 4$, it suffices to check that $aH = Ha$ for $a = (1, 2)$. (Note: H consists of the identity, and all 3 possible products of disjoint 2-cycles. This description takes no recourse to the name of actual elements, so renaming keeps H stable).

We compute:

$$\begin{aligned} (1,2)()(1,2)^{-1} &= (), \\ (1,2)((1,2)(3,4))(1,2)^{-1} &= (1,2)(3,4), \\ (1,2)((1,3)(2,4))(1,2)^{-1} &= (1,4)(3,2), \\ (1,2)((1,4)(2,3))(1,2)^{-1} &= (1,3)(2,4). \end{aligned}$$

So, the conjugate by $(1, 2)$ of every element of H is again an elemnt of H. It follows that H is normal and our idea of setting $(aH)(bH) = abH$ will indeed work.

As a side remark, note that every element of H is an even permutation. (Since they are made of zero or of two 2-cycles). It follows that each coset aH is either completely even or completely odd.

Now that we know that our S_4/KV_4 is a group, it is a reasonable question to ask: what group is it? A first step towards this is always to compute the Cayley table. An easy but painstaking computation reveals that is is as follows:

$$\begin{pmatrix} E & \rho & \lambda & \alpha & \beta & \gamma \\ \rho & \lambda & E & \beta & \gamma & \alpha \\ \lambda & E & \rho & \gamma & \alpha & \beta \\ \alpha & \gamma & \beta & E & \lambda & \rho \\ \beta & \alpha & \gamma & \rho & E & \lambda \\ \gamma & \beta & \alpha & \lambda & \rho & E \end{pmatrix}$$

Now checking back all the way at the start of Week 2, if we use the translations

$$e \leftrightarrow E, a \leftrightarrow \alpha, b \leftrightarrow \beta, c \leftrightarrow \gamma, r \leftrightarrow \rho, \ell \leftrightarrow \lambda,$$

we see that up to the renaming we are looking at $\mathrm{Sym}(\triangle) = S_3$.

Could we have seen this somehow? Yes, I think so, and here is how. The fact that we have any group structure at all on G/H is because the normality of H assures us that whenever a' belongs to the coset aH and b' belongs to the coset bH, then $a'b'$ is in the coset abH. Now look at the 6 cosets, and pick out the elements in each coset that *do not use 4*. We find $E \in E$, $(1, 2) \in \gamma$, $(1, 3) \in \beta$, $(2, 3) \in \alpha$,

$(1,2,3) \in \lambda$ and $(1,3,2) \in \rho$. The remarkable fact is that there is exactly one in each coset. Composing or inverting these elements can only produce other elements that also do not use 4, so these 6 elements actually form a group by themselves, a subgroup of S_4. And it is easy to see that this subgroup is exactly S_3. The renaming was made in such a way that a Greek letter corresponds to the Roman letter that we have the element of S_3 sitting inside the coset indicated by the Greek letter.

Let us look at a somewhat easier example, easier because of commutativity.

EXAMPLE VI.3. Let $G = (\mathbb{Z}/24\mathbb{Z}, +)$ and let H be the subgroup formed by the multiples of 6. As in the previous example, $|G| = 24$ and $|H| = 6$. But in this case there is no question that H is normal, since G is Abelian and so even $ah = ha$ *element by element*, and not just $aH = Ha$ as a set.

The cosets are then

$$\{0 + 24\mathbb{Z}, 6 + 24\mathbb{Z}, 12 + 24\mathbb{Z}, 18 + 24\mathbb{Z}\}, \tag{3}$$

$$\{1 + 24\mathbb{Z}, 7 + 24\mathbb{Z}, 13 + 24\mathbb{Z}, 19 + 24\mathbb{Z}\}, \tag{4}$$

$$\{2 + 24\mathbb{Z}, 8 + 24\mathbb{Z}, 14 + 24\mathbb{Z}, 20 + 24\mathbb{Z}\}, \tag{5}$$

$$\{3 + 24\mathbb{Z}, 9 + 24\mathbb{Z}, 15 + 24\mathbb{Z}, 21 + 24\mathbb{Z}\}, \tag{6}$$

$$\{4 + 24\mathbb{Z}, 10 + 24\mathbb{Z}, 16 + 24\mathbb{Z}, 22 + 24\mathbb{Z}\}, \tag{7}$$

$$\{5 + 24\mathbb{Z}, 11 + 24\mathbb{Z}, 17 + 24\mathbb{Z}, 23 + 24\mathbb{Z}\}. \tag{8}$$

So, for example, the last of these cosets contains all numbers that leave rest 5 when divided by 6. If we call these collections $\overline{0}, \ldots, \overline{5}$ (in the given order), and recall that we are supposed to use addition, it is clear how we want to think of the group G/H: it is $\mathbb{Z}/6\mathbb{Z}$.

We formulate officially what we have seen in examples.

THEOREM VI.4. *If H is a normal subgroup of G then one can equip the collection of left cosets $\{aH | a \in G\}$ with a group structure. The multiplication in this group takes aH and bH and multiplies them to abH. The resulting group is denoted G/H and called the* quotient group of G by H.

If H is normal, the same construction can be carried out for the right cosets Ha, and that also leads to a group. One can check that these are the same groups, so that the symbol "G/H" is unambiguous. ◇

It is often good for understanding a definition when one sees a case where the defined concept is absent.

EXAMPLE VI.5. Let G be $S_3 = \mathrm{Sym}(\triangle)$ and take H to be the subgroup $H = \{1, a\}$. We have checked ultiple times that H is not normal. Let us see how this impacts G/H being a group.

The collection of left cosets is $\{1, a\} = aH$, $\{b, ba = r\} = bH$ and $\{c, ca\} = cH$. Let us name these cosets E, R, L in that order.

If we hope that we can mae $\{E, R, L\}$ a group, then the product structure must come from multiplication in G. That means for example, that we should be able to multiply any element of the coset E with any element of the coset R and get elements that all live in the same coset. (Presumably, that product coset should be R again, since E contains the identity e and therefore should be the coset that gives the identity element in G/H).

However, we find: $eb = b, er = r, ab = \ell, ar = c$. These four products do not lie in any one of the three cosets E, R, L but in fact cover two of them, E and L. Thus, there is no meaningful product $E \cdot R$ and we cannot hope to make G/H into a group.

Note how this failure is quite similar to looking for morphisms $\mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ using multiplication by k that does not satisfy $n|km$. The underlying theme is that one is only allowed to do operations on cosets that do not destroy the cosets. Friends should stay friends!

2. The isomorphism theorem

Now suppose $\psi\colon G \to G'$ be a morphism. We checked previously, that then $\ker(\psi)$ is a subgroup of G. We also checked that this subgroup is normal in G (since if $\phi(h) = e_{G'}$ then $\psi(aha^{-1}) = \psi(a)\psi(h)\psi(a^{-1}) = \phi(a)e_{G'}\phi(a)^{-1} = e_{G'}$ showing that aha^{-1} belongs to $\ker(\psi)$ as well).

It follows that $G/\ker(\psi)$ can be turned into a group. We want to find ut what this quotient group has to do with ϕ in concrete terms.

Example VI.6. Let $G = \mathbb{Z}/28\mathbb{Z}$, $G' = \mathbb{Z}/42\mathbb{Z}$ and $\psi\colon G \to G'$ be "multiplication by 9" in the sense that $\psi(a + 28\mathbb{Z}) = 9a + 42\mathbb{Z}$.

Start with noting that 42 indeed divides $9 \cdot 28$, so by our often-mentioned criterion "multiplication by 9" does indeed give a function that does not destroy cosets.

The kernel of ψ consists of those cosets $a + 28\mathbb{Z}$ for which $9a$ is a multiple of 42. But $42|9a$ if and only if $14|a$. So, $\ker(\psi)$ has two elements, $\{0 + 28\mathbb{Z}, 14 + 28\mathbb{Z}\}$. We call this subgroup H.

You might want to think of H as $\mathbb{Z}/2\mathbb{Z}$ "stretched by the factor 14: as a group of 2 elements they are isomorphic. The identity is $0 + 28\mathbb{Z}$ and the Cayley table is $\begin{pmatrix} \overline{0} & \overline{14} \\ \overline{14} & \overline{0} \end{pmatrix}$. Formally, that is the table of $\mathbb{Z}/2\mathbb{Z}$.

Now in G/H we make "cosets of cosets". For example, in the coset eH we throw together $0 + 28\mathbb{Z}$ and $14 + 28\mathbb{Z}$. Note that this boils down to lumping together all the multiples of 14 into one big family. And that this is going to be the identity element of the quotient group G/H. So, G/H is "G with 14 declared to be zero". But that just means $\mathbb{Z}/14\mathbb{Z}$.

Now that we have understood the quotient, let us see what else ψ can tell us. The morphism ψ involves the groups G and G', and we also have concocted the group $\ker(\psi)$. There is a fourth group lurking here, namely the group of all elements that are output of ψ. In the case at hand, that is all cosets of the form $9a + 42\mathbb{Z}$. So, these are the cosets modulo 42 of $0, 9, 18, 27, 36, 45, \ldots$. But in $\mathbb{Z}/42\mathbb{Z}$, 45 counts the same as 3=45-42. So this set of output representatives really reads $0, 9, 18, 27, 36, 3, 12, 21, 30, 39, 6, 15, 24, 33, 0$. Then it cycles, and so we get any $b + 42\mathbb{Z}$ for which $3|b$.

The collection of cosets $3 + 42\mathbb{Z}, 6 + 42\mathbb{Z}, \ldots, 0 + 42\mathbb{Z}$ is a group with addition and sits inside $\mathbb{Z}/42\mathbb{Z}$. It is called the *image* of ψ, written $\operatorname{im}\psi$), and it is made of all the possible outputs of ψ. You can view it as $\mathbb{Z}/14\mathbb{Z}$ "scaled up" by a factor of 3. But remember: $\mathbb{Z}/14\mathbb{Z}$ was also what the group G/H looked like! The two groups $G/\ker(\psi)$ and $\operatorname{im}(\psi)$ have the same structure!

The fact that this is typical behavior is our next theorem.

Before we state it, let me remind you that you have seen something like this before: if A is a real $m \times n$ matrix, you can view it as a way to turn vectors $v \in \mathbb{R}^n$ into vectors $A \cdot v$ of $\mathbb{R}^m$. Both $\mathbb{R}^n, \mathbb{R}^n$ are groups (the first three axioms that you learned for a vector space mean just that it is a group for addition of vectors!) The kernel of the matrix A used to be the vectors v that have $Av = 0$, and since the zero vector is the identity element for vector addition, this old "kernel" idea for vector spaces agrees exactly with our new one for groups. And you were also told that the image of A (you used to call it the column span) is a vector space (hence group!) of dimension $\mathrm{rk}(A)$. And to top it all off, you learned that rank lus nullity gives n. In new and fancy terms this can be phrased as "the kernel of A is a vector space of dimension $n - \mathrm{rk}(A)$, and $\mathbb{R}^n / \ker(A)$ is a vector space of dimension $n - (n - \mathrm{rk}(A)) = \mathrm{rk}(A)$. This quotient is precisely the column space of A, a vector space of dimension $\mathrm{rk}(A)$ just like $\mathbb{R}^n / \ker(A)$.

THEOREM VI.7 (The isomorphism theorem). *If*

$$\psi \colon G \to G'$$

is a morphism of groups with kernel $H := \ker(\psi)$ *sitting inside* G*, and with image* $\mathrm{im}(\psi)$ *sitting inside* G'*, then there is an isomorphism*

$$\overline{\psi} \colon G/\ker(\psi) \simeq \mathrm{im}(\psi)$$

where $\overline{\psi}(aH) = \psi(a)$.

Here, the group operation in G/H *is* $(aH)(bH) = abH$ *and the operation in* $\mathrm{im}(\psi)$ *is the one from* G'.

I will not prove this theorem in detail, but here is why you should think it is true:

(1) As you move from G to G' using ψ, products are preserved but all of H is crunched dwon to $e_{G'}$, basically by definition. Therefore if you want to relate stuff n G with stuff in G', you need to form the cosets G/H to account for the "lumping together" of anything in H.
(2) You are not going to be able to relate elements of G' that *are not* outputs of ψ to anything in G since ψ is your only comparison vehicle, and stuff in G' that ψ "does not see" is stuff that ψ has no opinion about.
(3) So, really the question is what $G/\ker(\psi)$ hasto do with $\mathrm{im}(\psi)$. And the function $\overline{\phi}$ I mentioned, which sends a coset aH to $\psi(a)$, can be shown to be a morphism (easy, since ψ is), and injective (confusing, but easy) , and surjective (easy). But that makes it an isomorphism.

In particular, if ψ is surjective,

$$G/\ker(\psi) \simeq \mathrm{im}(\psi).$$

EXAMPLE VI.8. Here I will talk through some examples.

(1) Let ψ be the morphism from $\mathbb{Z}/15\mathbb{Z}$ to $\mathbb{Z}/25\mathbb{Z}$ that multiplies by 5, sending $a + 15\mathbb{Z}$ to $5a + 25\mathbb{Z}$. (Recall that if k is to be used as morphism from $\mathbb{Z}/m\mathbb{Z}$ to $\mathbb{Z}/n\mathbb{Z}$ then we need that $n|km$).

    Then $\ker(\psi) = \{a+15\mathbb{Z} | 5a+25\mathbb{Z} = 0+25\mathbb{Z}\}$. This requires that $25|5a$ so that a must be a multiple of 5. So, $\ker(\psi) = \{0+15\mathbb{Z}, 5+15\mathbb{Z}, 10+15\mathbb{Z}\}$. You can view this as $\mathbb{Z}/3\mathbb{Z}$ "inflated by a factor of 5".

    The image $\mathrm{im}(\psi)$ of ψ are the cosets $\{5a + 25\mathbb{Z}\}$. That is a group of 5 elements. We know (5 is prime) that this is a cyclic group, and indeed

$5+25\mathbb{Z}$ is a generator as all other image elements are multiples of $5+25\mathbb{Z}$. Abstractly, $\mathrm{im}(\psi)$ looks like $\mathbb{Z}/5\mathbb{Z}$ therefore. And one could say that it is $\mathbb{Z}/5\mathbb{Z}$ "inflated by a factor of 5".

So, the isomorphism theorem says that $(\mathbb{Z}/15\mathbb{Z})/5 \circ (\mathbb{Z}/3\mathbb{Z}) \simeq 5 \circ (\mathbb{Z}/5\mathbb{Z})$, where the $5\circ$ means "inflate by 5, and $\circ 4$ means "inflate by 3".

(2) More generally, let $n|km$ and consider the morphism $k\colon \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ that multiplies by k. Then the kernel is the elements $a + m\mathbb{Z}$ with $n|ak$ and these are just the cosets in $\mathbb{Z}/m\mathbb{Z}$ corresponding to the multiples of $n/\gcd(n,k)$. (The lowest a with $n|ak$ is the mimimal a that satisfies: ak is a multiple of k; ak is a multiple of n. So we want the smallest a for which ak is a multiple of $\mathrm{lcm}(n,k)$, and of course that smallest ak that is a multiple of $\mathrm{lcm}(n,k)$ is just $\mathrm{lcm}(n,k)$. It follows that the corresponding a is $\mathrm{lcm}(n,k)/k$ and so equals $n/\gcd(n,k)$ since in general $xy = \mathrm{lcm}(x,y)\cdot\gcd(x,y)$.)

The number of elements in the kernel is $\kappa := m/(n/\gcd(n,k)) = m\cdot\gcd(n,k)/n = mk/\mathrm{lcm}(n,k)$. This group looks like $\mathbb{Z}/\kappa\mathbb{Z}$ inflated by $\mathrm{lcm}(n,k)/k$.

The image is the subgroup of $\mathbb{Z}/n\mathbb{Z}$ consisting of the cosets to elements of the form ak. Since $\gcd(k,n)$ is a linear combination of k and n, this image is the same as the subgroup generated by the coset $\gcd(k,n)+n\mathbb{Z}$. It can be viewed as the group $\mathbb{Z}/\nu\mathbb{Z}$ inflated by $\gcd(k,n)$, where $\nu := n/\gcd(k,n) = \mathrm{lcm}(k,n)/k$.

Altogether, the isomorphism theorem says:

$$\frac{(\mathbb{Z}/m\mathbb{Z})}{\nu\circ(\mathbb{Z}/\kappa\mathbb{Z})} \simeq \gcd(k,n)\circ(\mathbb{Z}/\nu\mathbb{Z}).$$

Note that $\kappa\cdot\nu = (mk/\mathrm{lcm}(n,k))\cdot(\mathrm{lcm}(k,n)/k) = m$ as it should.

(3) In the previous items, normality came for free since the groups were normal. Let now G be the octonians. Its center Z consists of the elements $\{\pm 1\}$ as is easy to see. The center of a group is made of th elements that commute with everyone, so in particular the center of G is a normal subgroup.

The quotient of the octonians by their center is a group of $8/2 = 4$ elements. Which group is it? We know it can only be $\mathbb{Z}/4\mathbb{Z}$ or KV_4 since these are the only groups of size 4.

If you look at the cosets, they are $E = \{\pm 1\}, I = \{\pm i\}, J = \{\pm j\}, K = \{\pm k\}$. Note that $I\cdot I$ is exactly E, and similarly $J\cdot J = E = K\cdot K$.It follows that no element of G/Z has order 4, and so G/Z must be KV_4. We can check explicitly (element by element) that $I\cdot J = J\cdot I = K, I\cdot K = K\cdot I = J, J\cdot K = K\cdot J = I$. So we can align G/Z with KV_4 by $E \leftrightarrow (), I \leftrightarrow (12)(34)$, $J \leftrightarrow (13)(24)$, $K \leftrightarrow (14)(23)$, preserving Cayley tables.

In fact, (one can check that) we can make a morphism $\pi\colon G \to KV_4$ by sending 1 and -1 to E, i and $-i$ to I, j and $-j$ to J, and k and $-k$ to K, respecting multiplication. The kernel of this morphism is $\{\pm 1\}$, and so the isomorphism theorem predicts $G/Z \simeq KV_4$.

CHAPTER VII

Week 7: Finitely generated Abelian groups

1. Row reduced echelon form over the integers

A linear transformation (in linear algebra) is a function $T\colon V \to W$ such that $T(\vec{v}+\vec{v}') = T(\vec{v}) + T(\vec{'})$ and $T(\lambda\vec{v}) = \lambda T(\vec{v})$ for all $v, v' \in V$ and all $\lambda \in \mathbb{R}$. A moment's thought shows that this is a (somewhat special) morphism from the group $(V,+)$ to the group $(W,+)$.

Suppose $\dim(V) = n, \dim(W) = m$, and suppose one has chosen bases $B_V = \{b_1^V, \ldots, b_n^V\}$ and $B_W = \{b_1^W, \ldots, b_m^W\}$ in V and W respectively. (You might want to think of B_V, B_W as matrices whose columns are the elements of the basis). Then to each $v \in V$ there is a coefficient vector $c^{B_V}(\vec{v}) \in \mathbb{R}^n$ such that $\vec{v}$ is the linear combination $B_V \cdot c^{B_V}(\vec{v}) = \sum c(\vec{v})_i^{B_V} b_i^V$.

Recall that if $T\colon V \to W$ is a linear transformation (like in linear algebra) then there is a real $m \times n$ matrix A such that if $c^{B_V}(\vec{v})$ is the coefficient vector for $\vec{v}$ relative to the basis B_V, then $A \cdot c^{B_V}(\vec{v})$ is the coefficient vector of $T(\vec{v})$ relative to the basis B_W, and this happens for all $\vec{v} \in \mathbb{R}^n$. In other words, $T(B_V \cdot c^{B_V}(\vec{v})) = B_W \cdot (A \cdot c^{B_V}(\vec{(v)}))$.

If we change the basis on the source and target space to B'_V and B'_W, then there are matrices $Q_V \in \mathbb{R}^{n\times n}$ and $Q_W \in \mathbb{R}^{m\times m}$ such that $B'_V Q_V = B_V$ and $B'_W Q_W = B_W$. The coefficient vector for $\vec{v}$ relative to B'_V is then the vector $c^{B'_V}(\vec{v})$ such that $B'_V \cdot c^{B'_V}(\vec{v}) = \vec{v} = B_V \cdot c^{B_V}(\vec{v})$, but as $B'_V Q_V = B_V$ this means $B'_V Q_V \cdot c^{B_V}(\vec{v}) = B_V \cdot c^{B_V}(\vec{v}) = \vec{v} = B'_V \cdot c^{B'_V}(\vec{v})$, hence $c_V^{B'_V}(\vec{v}) = Q_V \cdot c^{B_V}(\vec{v})$.

Then we have

$$\begin{aligned} T(\vec{v}) = \vec{w} = B^W c^{B_W}(\vec{w}) = B^W A c^{B_V}(\vec{v}) &= B'^W Q^W A c^{B_V}(\vec{v}) \\ &= B'^W Q^W A (Q^V)^{-1} Q^V c^{B_V}(\vec{v}) \\ &= B'^W [Q_W A (Q_V)^{-1}] c'^V(\vec{v}). \end{aligned}$$

This says that the transformation T (which exists independently of the choice of basis) is represented relative to the bases B'^W, B'^V by the matrix $A' = Q_W A Q_V^{-1}$.

(As a special case, if $V = W$ and one chooses $B_V = B_W$ then the change of coordinates has the effect of conjugation on A. In some sense this is clear: if you have a recipe tor a tranformation (called A) that works in one language (the bases B_V, B_W) and you want to use in a different language (the bases B'_V, B'_W) then you first translate the ingredients from the new into the old language (by Q_V^{-1}), then use the recipe (namely A), and then translate the result into the new language (by Q_W). Once again, this goes right to left because that is the way functions work).

The moral of this linear algebra story is that a transformation is not affected by the way we think of the input and the output space, but the tools we use

to compute what the transformation does (namely, A) do change, and do so in predictable manner.

The main motivation is that we don't care too much what the bases are that we use, but want to understand only the nature ot T, we can perhaps arrange them so that the matrix A looks very simple.

Recall now, that a change of basis requires that Q_V, Q_W are invertible (so that you can undo the change, with the inverse matrix). Recall also that in linear algebra you learned that row reduction leads to row reduced echeleon form and can be accomplished by three elementary row operation steps: (I) interchanging two rows, (II) adding some number of copies of one row to another, and (III) scaling a row by a an invertible number. Recall finally that the process of row reduction of a matrix A is mirrored by multiplication of A on the left by elementary matrices, corresponding to the three steps, and so the row reduced echelon form of A can be achieved as a product $E \cdot A$ where E is the product of all the elementary row operations used to row reduce A. Naturally, this E is invertible since each row reduction step can be reversed.

Similar to row operations, one can discuss column operations and column reduced echeleon form, which is practically the transpose of the row reduced echelon form of the transpose of A. Now imagine what the row reduced echeleon form turns into when you column reduce it. The row reduced echelon form has rank many nonzero rows, and they start with leading ones placed on a Northwest-to-Southeast "diagonal". If you now column reduce, all that remains are the rank many leading ones.

Now we need to make a leap of sorts: we need to consider what happens to all this when we do not use real numbers, but just integers.

The main issue that comes up is that we can't divide most of the time. In particular, our usual formula for an inverse matrix,

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A)$$

involving the adjoint matrix, indicates that most matrices cannot be inverted over the integers. It only will work if $\det(A) = \pm 1$. This rules out one of the basic row operation steps, the one that says "rescale row i by λ". So we will have to live without that. On the other hand, switching 2 rows or 2 columns, or adding multiples of one row to another row, or adding multiples of one column to another column, are all processes that can be inverted with integers. So we still get to use these 2 kinds of operations, but now on rows and on columns.

EXAMPLE VII.1. Suppose $G = \mathbb{Z}^3$, the set of all 3-vectors with integer coordinates, and we want to understand the quotient G/H by the supgroup H that is generated by the columns $(1, 0, -1)^T, (4, 3, -1)^T, (0, 9, 3)^T$ and $(3, 12, 3)^T$. So, H consists of all linear combinations of these 4 columns. The difficulty in understanding the ramifications of "setting elements of H to zero" in the process of going from G to G/H is that the individual coordinates of a vector in H are not independent from one another.

So make a matrix $\begin{pmatrix} 1 & 4 & 0 & 3 \\ 0 & 3 & 9 & 12 \\ -1 & -1 & 3 & 3 \end{pmatrix}$. Read it as a map from $\mathbb{Z}^4$ to $\mathbb{Z}^3$, sending $\vec{v} \in \mathbb{Z}^4$ to $A \cdot \vec{v}$ in $\mathbb{Z}^3$.

Row reduction says that the relations of these 4 columns *don't* change (and neither does the row span) if you add row 1 to row 3 the 1 to wipe out the -1, which leads to $\begin{pmatrix} 1 & 4 & 0 & 3 \\ 0 & 3 & 9 & 12 \\ 0 & 3 & 3 & 6 \end{pmatrix}$. Of course, it *does* have an effect on the column span of the matrix, so this amounts to a coordinate change in $\mathbb{Z}^3$ (the target of the map).

Now our row reduction can go on, with 3 as pivot, erasing 3 below it. We get $\begin{pmatrix} 1 & 4 & 0 & 3 \\ 0 & 3 & 9 & 12 \\ 0 & 0 & -6 & -6 \end{pmatrix}$. That is another change of basis in the target space $\mathbb{Z}^3$. We can now change the -6 to a 6, and then normal row reduction would stop.

The row steps are a reflection of a change of basis in the target of the transformation, but we can also change basis in the source. That is encoded by (invertible) column operations. For example, we can use the top left 1 to wipe out the other numbers in row I to get $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 9 & 12 \\ 0 & 0 & 6 & 6 \end{pmatrix}$. And now we can use the 3 to wipe out all that is to its right: $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 6 & 6 \end{pmatrix}$. And then the left 6 to kill the right 6: $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 6 & 0 \end{pmatrix}$.

DEFINITION VII.2. The shape of this matrix is called *Smith normal form*. It features: only nonzero entries on the diagonal, and from upper left to lower right the diagonal entries are multiples of the previous entries.

The business of base change in source and target does not change the structure of the quotient group (target/rowspan), although it changes how we think of it (as any coordinate change does). So, our quotient group G/H now turns out to be $\mathbb{Z}^3$ modulo the linear combinations of the columns of the last matrix above. In other words,

$$G/H \simeq (\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z})/\{(a, 3b, 6c) | a, b, c \in \mathbb{Z}\}.$$

The point of the row reduction work is that the stuff in H now has been "decoupled": the first coordinate of an element of H is any number, the second is any number divisible by 3, the last is any multple of 6. The coordinates do no longer "talk to each other", they have become independent.

This also makes clear what G/H is equal to: $(\mathbb{Z}/1\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/6\mathbb{Z})$. Note that $\mathbb{Z}/1\mathbb{Z}$ is the trivial group as $1\mathbb{Z} = \mathbb{Z}$.

Recall, that $\mathbb{Z}/6\mathbb{Z} = (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$ since $2, 3$ are coprime. So $G/H =$ (trivial group) $\times (\mathbb{Z}/3\mathbb{Z}) \times ((\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z}))$.

There is one big hurdle we did not meet in the previous example: our pivots came as a free gift. The following example shows what to do when lunch is not free.

EXAMPLE VII.3. Lets try this for H the subgroup of $G = \mathbb{Z}^3$ generated by $(10, -4, 8)^T, (-6, -6, -16)^T, (4, -10, -8)^T$, which yield the matrix $\begin{pmatrix} 10 & -6 & 4 \\ -4 & -6 & -10 \\ 8 & -16 & -8 \end{pmatrix}$. There is no element here that can be used as a pivot, because a pivot should divide all the other numbers it is used to wipe out (we don't have access to fractions...). This means, we have to make a pivot first, by clever row or column operations, or both.

The main question is what we can hope and aim for. Surely, we can't make a 1 here since all numbers are even. But we could hope for a 2, and that would divide every other number. And we can make a 2 by subtracting row III from row I, to get $\begin{pmatrix} 2 & 10 & 12 \\ -4 & -6 & -10 \\ 8 & -16 & -8 \end{pmatrix}$. Now clean out the front column: $\begin{pmatrix} 2 & 10 & 12 \\ 0 & 14 & 14 \\ 0 & -56 & -56 \end{pmatrix}$. Then one more row step leads to $\begin{pmatrix} 2 & 10 & 12 \\ 0 & 14 & 14 \\ 0 & 0 & 0 \end{pmatrix}$ and then 3 column operations produce $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

We infer that $G/H \simeq (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/14\mathbb{Z}) \times (\mathbb{Z}/0\mathbb{Z}) = (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/7\mathbb{Z}) \times (\mathbb{Z})$ since $2, 7$ are coprime.

Note that the zero on the diagonal is actually very important here, it tells us about $\mathbb{Z}$ being a factor of G/H (and so makes G/H have infinitely many elements).

DEFINITION VII.4. If A is an integer $m \times n$ matrix with $m \leq n$ then let A' be the Smith normal form of A. The diagonal elements of A' are the *elementary divisors* of A.

If $m > n$, first augment A with $m - n$ columns of zeros on the right, and then proceed to compute Smith normal form. (This has the effect of adding $m - n$ zeros to the set of elementary divisors). ◇

THEOREM VII.5 (FTFGAG, Part 1). *We assume that A is $m \times n$, with $m \leq n$. If $m > n$, augment A with $m - n$ columns of zeros. We start with properties of Smith normal form and elementary divisors.*

1. *The Smith normal form of A can be computed by row and column operations of types I and II.*
2. *The Smith normal form of A is determined by A alone, and not on how we compute the normal form by pivot choices.*
3. *The elementary divisors $d_1, \ldots, d_n$ of A are the m numbers on the diagonal of the Smith normal form A' of A.*
4. *The elementary divisors satisfy $d_i | d_{i+1} \forall i$.*

2. Generating groups

Recall that if a group Q is cyclic with generator g then the elements of H are powers of either g or g^{-1}. This gives a *presentation*

$$Q = \mathbb{Z}/\operatorname{ord}(g)\mathbb{Z}$$

as a quotient of $\mathbb{Z}$ by a suitable subgroup.

More generally, we say that a set $\{g_1, \dots, g_k\}$ of elements of H is a *generating set* if every element of Q is a product of powers of the g_i and/or their inverses.

For a general group Q with generating set $\{g_1, \dots, g_k\}$ one can make a surjective morphism from the free group F_k on k symbols to H, by sending the i-th symbol of F_k to g_i. If Q is Abelian, one can also make a surjective morphism $\mathbb{Z}^k \to Q$ by sending the i-th unit vector in $\mathbb{Z}^k$ to g_i. These surjections are called *presentations*.

THEOREM VII.6 (FTFGAG, Part 2). *We consider a subgroup H of $G = \mathbb{Z}^m$ and investigate the quotient G/H.*

(5) Any subgroup H of $\mathbb{Z}^m$ can be generated with a finite number of columns from a suitable matrix A.

(6) The group G/H is isomorphic to

$$(\mathbb{Z}/d_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/d_m\mathbb{Z})$$

where $d_1, \dots, d_m$ are the elementary divisors of A. They do not depend on how one chooses A.

(7) For comparisons of different groups $\mathbb{Z}^m/H$ and $\mathbb{Z}^{m'}/H'$, one can split $\mathbb{Z}/d_i\mathbb{Z}$ further using coprime factors.

(8) Two quotients $\mathbb{Z}^m/H$ and $\mathbb{Z}^{m'}/H'$ are isomorphic if and only if their lists of elementary divisors are equal after striking all appearances of $d_i = 1$ from both lists.

A comment on the last item: $\mathbb{Z}/1\mathbb{Z}$ is the trivial group, and so $\mathbb{Z}/1\mathbb{Z} \times G = G$ for any group G. So erasing instances of $\mathbb{Z}/1\mathbb{Z}$ from the third item of the theorem does not change anything.

All parts except the last one are cler from what we have done and said in examples. At the end of the section I explain why the last part is true (why different elementary divisors must come from non-isomorphic groups).

Let us ask what we can do for arbitrary Abelian groups. The answer is: with a bit of preprocessing, the exact same things.

EXAMPLE VII.7. Let $G = KV_4 = \{e, \updownarrow, \leftrightarrow, \curvearrowleft\}$. This group is Abelian, and has 3 elements aside from the identity. The main observation of this example is that we can make a morphism $\pi\colon \mathbb{Z}^3 \to KV_4$ that sends $(1,0,0)$ to $\updownarrow$, $(0,1,0)$ to $\leftrightarrow$ and $(0,0,1)$ to $\curvearrowleft$.

This map is surjective, but surely not an isomorphism (for example, because $\mathbb{Z}^3$ is infinite and KV_4 is not). What is in the kernel of π? These are the expressions in $\updownarrow, \leftrightarrow, \curvearrowleft$ that give the identity in KV_4. For example, $\updownarrow \cdot \updownarrow = e$ in KV_4 and so $(1,0,0) + (1,0,0) \in \ker(\pi)$. To understand how this came about, recall that we have the morphism rule $\pi(\vec{v} + \vec{w}) = \pi(\vec{v}) \cdot \pi(\vec{w})$. So if $\vec{v} = \vec{w} = (1,0,0)$ then $\pi((1,0,0) + (1,0,0)) = \pi((1,0,0)) \cdot \pi((1,0,0)) = \updownarrow \cdot \updownarrow = e$, placing $(1,0,0) + (1,0,0)$ in the kernel of π. (Recall: kernel is whoever is mapped to the identity).

Other elements in the kernel are: $(0,2,0), (0,0,2)$, basically for similar reasons. But there is another more interesting relation: since $\updownarrow \cdot \leftrightarrow = \curvearrowleft$, the corresponding relation is $(1,1,-1)$. It turns out that these 4 elements generate the kernel of π.

So let us run the elementary divisor business on H, the subgroup of $\mathbb{Z}^3$ spanned by the kernel of π, which is the column span of $\begin{pmatrix} 2 & 0 & 0 & 1 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & -1 \end{pmatrix}$. If we move the $(1,1,-1)$ column to the left and then clear out lower parts of the left column, we get

$\begin{pmatrix}1&2&0&0\\0&-2&2&0\\0&2&0&2\end{pmatrix}$. We then use the -2 as pivot to get $\begin{pmatrix}1&2&0&0\\0&-2&2&0\\0&0&2&2\end{pmatrix}$. At last, we do column operations to get to $\begin{pmatrix}1&0&0&0\\0&2&0&0\\0&0&2&0\end{pmatrix}$, which certifies that $KV_4 = \mathbb{Z}^3/H$ is isomorphic to $(\mathbb{Z}/1\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z}) = (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$. We can associate KV_4 with $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$ via

e	$\updownarrow$	$\leftrightarrow$	$\curvearrowleft$
$(0+2\mathbb{Z}, 0+2\mathbb{Z})$	$(1+2\mathbb{Z}, 0+2\mathbb{Z})$	$(0+2\mathbb{Z}, 1+2\mathbb{Z})$	$(1+2\mathbb{Z}, 1+2\mathbb{Z})$

and this assignment is an isomorphism.

One can use this result to count the number of Abelian groups of a certain order, and also compare different Abelian groups for being isomorphic.

Example VII.8. How many Abelian groups G with 168 elements are there? For each, find the elementary divisors.

$168 = 2^3 \cdot 3^1 \cdot 7^1$. By FTFGAG, G should be a product of some $\mathbb{Z}/2^{d_i}\mathbb{Z}$ and $\mathbb{Z}/3^{e_i}\mathbb{Z}$ and $\mathbb{Z}/7^{f_i}\mathbb{Z}$. Of course, in order to make the group indeed have 168 elements, we need the sum of the d_i to be 3, and the sum of the e_i to be 1 and the sum of the f_i to be 1 as well. That actually leaves very little choice, since an exponent of 0 can be ignored. We must have one e and one f of value 1. The only interesting bit is how we partition 3. As we know, this could be as $1+1+1$ or as $1+2$ or as 3.

So the possibilities are:

$$(\mathbb{Z}/8\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/7\mathbb{Z}),$$
$$(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/4\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/7\mathbb{Z}),$$
$$(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/7\mathbb{Z}).$$

The elementary divisors satisfy: the product is 168, they divide each other (and 1's can be ignored since they lead to $\mathbb{Z}/1\mathbb{Z}$ factors which are trivial). Since 3 and 7 appear with power 1 in 168, both appear only in the last (biggest) elementary divisor. The possibilities are: 168, or $2 \cdot 84$, or $2 \cdot 2 \cdot 42$. One can see that the partitions of the exponent 3 of 2 correspond to these factorizations: $1+1+1$ corresponds to $(2^1) \cdot (2^1) \cdot (2^1 \cdot 3^1 \cdot 7^1)$, $1+2$ to $(2^1) \cdot (2^2 \cdot 3^1 \cdot 7^1)$, and 3 to $(2^3 \cdot 3^1 \cdot 7^1)$. Of course, the same applies to the partitions of the exponent 1 over 3 and 7, but since 1 can't be partitioned nontrivially, that is not so thrilling and 3, 7 only appear in the last elementary divisor. ◇

We want to explain lastly, why for example $G = (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/4\mathbb{Z})$ is not isomorphic to $\mathbb{Z}/8\mathbb{Z}$, and in the process understand all similar questions with more or higher exponents.

The underlying reason is by finding elements that are "killed" by 2 (or its powers) in this case. By this we mean elements $g \in G$ that when you double them are zero. In $\mathbb{Z}/2^e\mathbb{Z}$, there is always exactly one element that is not zero but yet killed by 2, namely the coset of 2^{e-1}. More generally, we learned when we studied cyclic groups, that the number of elements in a cyclic group $\mathbb{Z}/n\mathbb{Z}$ of exact order d is either zero (when d does not divide n) or (if $d|n$) equals $\phi(d)$, the Euler phi

function that counts numbers relatively prime to d. Since $\phi(2) = 1$ this agrees with the above search.

So in a cyclic group of order divided by p, the number of elements that are killed by the prime number p is exactly $\phi(p) + 1$, the 1 coming from the fact that the identity is killed by p but already dead (and so did not count for the order-p-count in $\phi(p)$). But $\phi(p) + 1 = p$, so in a cyclic group of order divided by p there are exactly p elements killed by p if p is prime.

Now, in a product such as $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/4\mathbb{Z})$ there are now 2×2 elements killed by 2, because if a pair is killed by 2 then each component is killed by 2. And since there are 2 choices in each component, then there are 2×2 such pairs.

More generally, in a product $(\mathbb{Z}/p^{e_1}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p^{e_k}\mathbb{Z})$, p^k elements will be killed by p. So, groups with a different number of factors of the sort $\mathbb{Z}/p^e\mathbb{Z}$ cannot be isomorphic, because they have different numbers of elements that are killed by p.

If the number of such $\mathbb{Z}/p^e\mathbb{Z}$ is the same, consider the number of elements killed by p^2. In each $\mathbb{Z}/p^{e_i}$, if $e_i = 1$ there are p elements killed by p^2, but if $e_i > 1$ then there are p^2 such elements. So in $(\mathbb{Z}/p^{e_1}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p^{e_k}\mathbb{Z})$ there are $p^{\#\{e_i \geq 1\}} \cdot p^{\#\{e_i > 1\}}$ elements killed by p^2. So, groups with equal number of factors of type $\mathbb{Z}/p\mathbb{Z}$ but different numbers of factors $\mathbb{Z}/p^2\mathbb{Z}$ are not isomorphic.

In this manner one can prove by induction the last part of the theorem.

Remark VII.9. The above is relevant to the finite part of a group (the part to elementary divisors different from 0). In homework you will show that $\mathbb{Z}^m$ and $\mathbb{Z}^n$ are isomorphic exactly if $m = n$. That then finishes the last part of FTFGAG stated next.

Theorem VII.10 (FTFGAG, Part 3). *Let G be any finitely generated Abelian group, and choose generators $g_1, \ldots, g_m$. Then G has a presentation $\pi\colon \mathbb{Z}^m \twoheadrightarrow G$ with $\pi(e_i) = g_i$ where e_i is the i-th unit vector of $\mathbb{Z}^m$. Then this identifies $G = \mathbb{Z}^m/H$ as $\mathbb{Z}^m$ modulo some subgroup H of $\mathbb{Z}^m$. Here the*

One can find a matrix A whose column span is exactly H. The elementary divisors of A do not depend on the chosen presentation of G nor do they depend on the chosen matrix A. They only depend on G.

The finitely generated Abelian group G is characterized by the elementary divisors in the sense that two groups have the same elementary divisors if and only if they are isomorphic.

CHAPTER VIII

Week 8: Group actions

We have seen in two different places that one can read a group as a bunch of permutations. First, as symmetries of actual objects (like an equilateral triangle, for example) where the permutations occur at special places of the objects (the corners of the triangle). Secondly, and much more formally, we have interpreted a group element $g \in G$ as a permutation σ^g of the elements of G via left mutiplication: $\sigma^g(g') = gg'$. In this section we formalize this sort of idea and discuss some consequences.

DEFINITION VIII.1. Let X be a set and let G be a group. Under the following circumstances we shall speak of a *left action of G on X*:

(1) There should be a way of "multiplying" any element of G onto any element of X. In other words, we need a function
$$\lambda \colon G \times X \to X,$$
$$(g, x) \mapsto \lambda(g, x).$$
We then want that this action behaves well with respect to group multiplication as follows.

(2) The identity element $e = e_G$ should "fix" every element of X, so that we have
$$\lambda(e_G, x) = x$$
for all $x \in X$.

(3) Given any two group elements $g, g' \in G$ we require
$$\lambda(g, \lambda(g', x)) = \lambda(gg', x).$$

We will look exclusively at left actions, and henceforth just say "action" when we mean "left action". (Just to fill the void: a right action $\rho \colon X \times G \to X$ would want that $\rho(g', \rho(g, x)) = \rho(gg', x)$; note the reversion in the order of g, g' here).

We will often write less officially gx for the result $\lambda(g, x)$ of g acting on x. Then the two rules above become

$$\begin{aligned} eg &= g \qquad \forall x \in X, \forall g \in G, \\ g(g'x) &= (gg')x \qquad \forall g, g' \in G, \forall x \in X. \end{aligned}$$

I recommend thinking of the elements of X as physical objects ("points") that one can draw and touch, and the process $\lambda(g, -)$ as a way of moving the points in X about. Here, $\lambda(g, -)$ is the process the lets $g \in G$ act on all points of X, the $-$ is just a place holder.

I order to say interesting things about group actions, we need a few more concepts that arise naturally.

DEFINITION VIII.2. Let λ be an action of G on X and choose $x \in X$.

- The *orbit* of x is those points y in X that you can "get to from x" using multiplication of x by elements of G. In symbols, denoting the orbit of x by $\mathrm{orb}_G(x)$,
$$\mathrm{orb}_G(x) = \{y \in X | \exists g \in G \text{ with } \underbrace{gx} = \lambda(g,x) = y\}$$
or simply $\mathrm{orb}_G(x) = Gx$.
- If starting from x, the action can carry you to all other points of X, then we say that the action is *transitive*. If G acts transitively on X then it is customary to call X a *homogeneous G-space*.
- Complementary to the orbit of x is the notion of the *stabilizer* of x,
$$\mathrm{Stab}_G(x) = \{g \in G \text{ with } gx = x\},$$
the group elements that do not move x. Here we say that g *moves* x if $gx \neq x$.
- If no element of G moves x, that is when $\mathrm{Stab}_G(x) = G$, we call x a *fixed point of G*. If g does not move x, we say that x is a *fixed point for g*, or that g *fixes* x. We write $\mathrm{Fix}_X(g)$ for the points $x \in X$ for which $gx = x$.

REMARK VIII.3. You will show in homework that $\mathrm{Stab}_G(x)$ is a subgroup of G.

We consider some examples, concrete and abstract.

EXAMPLE VIII.4. Let $G = \mathrm{Sym}(\triangle)$ and let X consist of the vertices of the triangle. As we said many times, G can also be interpreted as S_X, the permutation group on the elements of X.

Let x be the A-vertex. Then $\mathrm{Stab}_G(x)$ consists of the identity e and the A-flip a, since the other 4 elements b, c, ℓ, r of G all move x.

Similarly, the stabilizer of C is $\{e, c\}$ and that of B is $\{e, b\}$.

The rotations ℓ, r have no fixed points, and the fixed points of e are al points of X. The reflections a, b, c have only one fixed point each.

The action is transitive, since already the rotations are capable to carry any point to any other point.

EXAMPLE VIII.5. Let $\overline{G}$ be the symmetris of a cube, and let G be the rigid symmetry group of a cube. cube. (This is the subgroup of all symmetries of the cube consisiting of just the rigid motions that are cube symmetries). We found that $|G| = 48$, 24 rotations from G, plus 24 non-rigid motions that are a composition of a rotation and the antipodal map (which sends each vertex to the one diametrically across).

Let X be the vertices of the cube and study the action of G (or $\overline{G}$) on X. If x is the upper left front vertex, there are 3 rigid motions that stabilize it (the 3 rotations that fix the big diagonal on which x lies) and then 3 more non-rigid motions that combine the antipodal map with the 3 rotations that exchange x with its antipode. So $|\mathrm{Stab}_G(x)| = 3$ and $\mathrm{Stab}_{\overline{G}}(x)| = 6$.

Both actions are transitive. (Since $G \subseteq \overline{G}$, it is enough to check that for G, but we know that one can rotate any vertex of the cube into any other).

Most elements of G have no fixed point in X. Note that if a motion fixes a vertex, it must also fix the antipodal point of that vertex. The 2×4 non-trivial rotations that fix a big diagonal have 2 fixed points. The identity of G has 8 fixed points.

The 3×4 motions from $\overline{G}$ that combine the antipodal map with a rotation about one of the big diagonals followed by a reflection about the plane perpendicular to this diagonal also have two fixed points.

EXAMPLE VIII.6. Let G be any group and H a subgroup. We do not require H to be normal. Let G/H be the set of all cosets gH relative to H. We take X to be G/H and act on it by left multiplication:

$$\lambda(g, g'H) = gg'H \text{ for all } g, g' \in G.$$

It is straightforward to check the group action rules: $\lambda(e_G, gH) = e_G gH = gH$, and $\lambda(g, \lambda(g', g"H)) = gg'g"H = \lambda(gg', g"H)$ because of associativity of multiplication in H.

The stabilizer of a coset gH is the set of all $a \in G$ with $agH = gH$, which says $g^{-1}agH = H$ and that is equivalent to $g^{-1}ag \in H$. For example, if $g = e$ and so $gH = eH = H$, the condition becomes $a \in H$, so the stabilizer of the "point" eH in X is exactly H. In general, the the equation $agH = gH$ means that for every $h \in H$ the expression agh should be of the form gh' for some $h' \in H$. That means $ag = gh'h^{-1}$ and so $a = gh'h^{-1}g^{-1}$. Since the product $h'h^{-1}$ is again in H, we find that a must be ing gHg^{-1}. On the other hand, $(gHg^{-1})(gH) = gH(gg^{-1})H = gHH = gH$ so that the stabilizer of gH is exactly the set gHg^{-1}. This says that the stabilizers of gH are always conjugate subgroups of H. In particular, if H happens to be normal (but only then), each stabilizer is equal to H.

If gH wants to be a fixed point for multiplication by g' then we need $g'gH = gH$, which amounts to $g^{-1}g'gH = H$. This forces $g^{-1}g'g$ to be in H, so there should be an element $h \in H$ with $g^{-1}g'g = h$, or $g' = ghg^{-1}$. So, gH is a fixed point for g' precisely if g' is in the conjugate subgroup gHg^{-1}.

In reverse, given g' then gH is fixed under multilication with g' precisely when gHg^{-1} contains g'. Note that belonging to gHg^{-1} may not be very easy for g'. For example, if H is normal then the condition "g' should belong to some conjugate subgroup of H" just boils down to "g' must be in H". Specifically, this applies in an Abelian group Gm as then all subgroups H are normal.

We are interested in counting. That usualy means, G and X should be finite.

THEOREM VIII.7 (Stabilizer–Orbit Theorem). *If G acts transitively on X and both are finite, then*

$$|G| = |\operatorname{orb}_G(x)| \cdot |\operatorname{Stab}_G(x)|$$

for every point x of X. If the action is transitive, so that there is only one orbit X, this becomes

$$|G| = |X| \cdot |\operatorname{Stab}_G(x).$$

I won't prove this formally, but give some ideas.

If $x, y \in X$ are in the same orbit, then $\operatorname{Stab}_G(x)$ and $\operatorname{Stab}_G(y)$ are conjugate subgroups as you show in homework. So in particular, they have the same size. This explains why in the theorem it is not important which x one takes: all stabilizers are conjugate to one another.

Next you cluster the elements of G in such a way that g, g' are in the same cluster if and only if $gx = g'x$. Note that $g \cdot \operatorname{Stab}_G(x)$ all end up in the same cluster since they all send x to gx. Note that these sets are just the left cosets relative to $H := \operatorname{Stab}_G(x)$. One then checks (easy but detailed) that g, g' belonging to different H-costets rules out the possibility of $gx = g'x$. So, the clusters all are of the same

size as $\mathrm{Stab}_G(x)$. So, G is partitioned into clusters of size $|\mathrm{Stab}_G(x)|$, and elements g, g' in different clusters produce different output $gx \neq g'x$ when multiplied against x. But the collection of all outputs Gx is just the orbit of x. So, as the theorem claims, $|G| = |\mathrm{orb}_G(x)| \cdot |\mathrm{Stab}_G(x)|$.

(This all should remind you much of Lagrange's Theorem and its proof. In fact, this proof here is the proof for Lagrange's Theorem if you take X to be the coset space for the subgroup H as in the example above. Then the Stabilizer-Orbit Theorem becomes Langrange's: "$|G| = |G/H| \cdot |H|$". In reverse, this theorem and its proof is simply Lagrange applied to G and its subgroup $H := \mathrm{Stab}_G(x)$).

Finally, if G acts transitively, there is only one orbit, and so $\mathrm{orb}_G(x) = X$.

EXAMPLE VIII.8. Let G be the rigid symmetries of a cube, choose as X the vertices of the cube, and let x be the upper front left vertex. The orbit of x is X since the action is transitive, and $|X| = 8$. The stabilizer of x has 3 elements (the big diagonal rotations that fix x) as discussed above. And indeed, $3 \cdot 8 = 24 = |G|$.

Now we discuss fixed point counts. Recall that for $g \in G$, the set $\mathrm{Fix}_X(g)$ is the points of X that are unmoved by g, so $gx = x$. Let us alo write X/G for the *orbit space* of X under G. This is just the set of all orbits, the notation sugesting that X/G arises from X by clustering elements of X where clusters are orbits. The following theorem addresses the question of counting the number of orbits.

THEOREM VIII.9 (Burnside). *If G acts on X and both are finite, then the size of the orbit space is*

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

Again, I won't give a very formal proof but the main ideas. Let us count $\sum_{g \in G} |\mathrm{Fix}_X(g)|$ as follows. Look at the collection of pairs (g, x) in the Cartesian product $G \times X$ for which $gx = x$. Let F be the collection of all such pairs. We can sort them by the separate g, or the separate x. If we sort them by g then we get clusters $\mathrm{Fix}_X(g)$ and so the number of all such pairs is precisely $\sum_{g \in G} |\mathrm{Fix}_X(g)|$. But if we cluster by x, then each cluster has the form $\{g \in G | gx = x\}$ and that is exactly $\mathrm{Stab}_G(x)$. So, if we now sum this over all x we get $\sum_{x \in X} |\mathrm{Stab}_G(x)|$. Of course, these two counts must agree:

$$\sum_{x \in X} |\mathrm{Stab}_G(x)| = \sum_{g \in G} |\mathrm{Fix}_X(g)|.$$

We now need to interpret the sum on the left a bit differently. From the Stabilizer–Orbit Theorem, if we let G just act on the orbit Gx of x, we know that $|G| = |\mathrm{Stab}_G(x)| \cdot |\mathrm{orb}_G(x)|$. So, restricting the sum to the orbit of x, we get $\sum_{x \in \mathrm{orb}_G(x)} |\mathrm{Stab}_G(x)| = \sum_{x \in \mathrm{orb}_G(x)} |G|/|\mathrm{orb}_G(x)| = |G| \sum_{x \in \mathrm{orb}_G(x)} 1/|\mathrm{orb}_G(x)| = |G|$.

So, orbit by orbit, the expression $\sum_{x \in X} |\mathrm{Stab}_G(x)|$ contributes one copy of $|G|$. If you sum over all orbits, this is $|G|$ times the sum of the number of orbits. The latter is $|X/G|$, and so we find that $|G| \cdot |X/G| = \sum_{x \in X} |\mathrm{Stab}_G(x)|$. Combined with the equation $|X/G| = \frac{1}{|G|} \sum_{g \in G} |\mathrm{Fix}_X(g)|$ this shows the Burnside Theorem. Note that there is very little "power" in this proof, it relies on 2 ways of counting the same thing.

EXAMPLE VIII.10. How many different dice can one make with labels 1 through 6 on them? It turns out, this question is made for Mr Burnside.

First off, if the die can't move, there are $720 = 6!$ ways to paint numbers on the faces of the cube. The problem is that dice *can* move, and so many of the seemingly different dice will turn out to be the same.

Let us write X for the 720 different dice that we painted. Let G be the symmetry group of the cube, it moves the dice around and has $|G| = 24$ elements.

If 2 dice are truly differently labeled, they would not look the same under any symmetry. So they would not be in the same G-orbit. In other words, we want to count the size of the orbit space X/G.

If we plan to use the Burnside Theorem, we need to study the fixed points of all motions. Note that a "fixed point" is now a labeling of the cube that looks the same no matter what we do with that cube. But it is clear that every rigid motion of the cube will move a face and in fact several. So there are no g witha fixed point. Unless, of course, you took g to be the identity motion, which has every labeling as a fixed point. So, in the Burnside formula there is exactly one summand that contributes anything, namely the one that belongs to $g = e$. And the summand for $g = e$ is $|\operatorname{Fix}_X(e)| = |X| = 720$. All other summands belong to a g without fixed points and contribute 0. So the formula says $|X/G| = \frac{1}{24}(720 + 0 + \ldots + 0) = 30$.

The example makes clear a special case of the Burnside Theorem:

COROLLARY VIII.11. *If X acts on G and no element $e \neq g \in G$ has any fixed point, then $|X/G| = |X|/|G|$.*

Review

- Week 1
  - induction, well ordering
  - modular arithmetic
  - primes and irreducibles in a domain
  - Euclidean algorithm in $\mathbb{Z}$, gcd, lcm, relative prime (coprime)
- Week 2
  - symmetries of an object and composition of symmetries
  - group (axioms), and Cayley table
  - cancellation property in groups
  - exampls: symmetry groups, KV_4, $GL(n, \mathbb{R})$, vector spaces, $(\mathbb{Z}/n\mathbb{Z}, +)$, $U(n)$, C_n, free groups, $\mathbb{Z}^n$
  - Abelian groups, cyclic groups
  - order of a group, and of elements in a group
  - subgroup
  - product group $G \times H$
  - $\mathrm{Aut}(G)$, the relabelings of G that preserve the Cayley table, a group with composition
- Week 3
  - the Euler ϕ-function
  - the number of elements in the cyclic group C_n that have order d (distinction for $d|n$ and $d \not| n$)
  - the number of subgroups of a given size in C_n
  - the number of generators for C_n
  - $\phi(\mathbb{Z}/pq\mathbb{Z}) = \phi(\mathbb{Z}/p\mathbb{Z}) \cdot \phi(\mathbb{Z}/q\mathbb{Z})$ if $\gcd(p, q) = 1$
  - if $a = a_1 \cdot a_k$ then $C_{a_1} \times C_{a_2} \times \cdots C_{a_k} = C_a$ provided that the a_i are pairwise coprime
  - solving $x \bmod n = a \bmod n, a \bmod m = b \bmod m$ when m, n coprime
  - $U(mn) = U(m) \times U(n)$ if coprime
  - $|U(p^k)| = p^{k-1}(p-1)$, and why
  - $\phi\colon \mathbb{Z}/mn\mathbb{Z} \to (\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$ via $1 + mn\mathbb{Z} \mapsto (1 + m\mathbb{Z}, 1 + n\mathbb{Z})$ is isomorphism provided m, n coprime
- Week 4
  - left and right cosets of G relative to the subgroup H; coset space G/H
  - morphisms $\psi\colon G \to G'$ and a list of examples
  - know how to test whether multiplication by $k \in \mathbb{N}$ gives a morphism $\mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$
  - conjugation by a, $g \mapsto aga^{-1}$

  - inner automorphisms $\phi_a \colon G \to G$, sending $g \mapsto aga^{-1}$
  - properties of cosets: either equal or disjoint; union is G; all same size
  - Lagrange: $\operatorname{ord}(g) | \operatorname{ord}(G)$; $|G| = |H| \cdot |G/H$
  - if $|G|$ prime then G cyclic
  - normal subgroup (stable under conjugation)
  - kernels of morphisms are normal
  - G Abelian, then every subgroup normal
  - index 2 subgroups are normal
- Week 5
  - the symmetric group S_n
  - 3 notations: output notation, standard notation, cycle notation; know how to convert one into the other and how to make cycles disjoint
  - tranposition = 2-cycle
  - disorder as number of switches
  - odd/even: parity
  - sign of a permutation
  - the alternating group as the kernel of $\sigma \mapsto \operatorname{sign}(\sigma)$, a group morphism from S_n to $(\pm 1, \cdot)$.
  - every group is a subgroup of a pemutation group
- Week 6
  - suppose here H is normal. Then G/H can be made a group, $(g_1 H) \cdot (g)2H) = (g_1 g_2 H)$. That this works is precisely because H is normal.
  - the kernel of a morphism is a normal subgroup
  - if $\phi \colon G \to G'$ is a morphism, denote H the kernel. Then G/H is isomorphic to the image of ϕ (this is a subgroup of G')
- Week 7
  - $A \in \mathbb{Z}^{m,n}$ has a Smith normal form, computable via standard row and column reduction steps
  - the diagonal elements of the Smith normal form are the elementary divisors of A, independent of pivot choices in the reduction
  - elementary divisors $d_1, \ldots, d_m$ divide one another, $d_i | d_{i+1}$
  - If $G = \mathbb{Z}^m / H$ where H is the column span of $A \in \mathbb{Z}^{m,n}$ with $m \leq n$, then the elementary divisors $d_1 | \ldots | d_m$ of A characterize G: if you discard any "1" on that list, then two Abelian groups G, G give the same lists of elementary divisors if and only if G and G' are isomorphic. So, elementary divisors solve the "classification problem" for finitely generated Abelian groups

CHAPTER IX

Week 9: Introduction to rings

This begins the second part of the course, where we study structures that allow both addition and multiplication. The standard example is $\mathbb{Z}$, with $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ following closely behind.

DEFINITION IX.1. A *ring* is a set R with a binary operation $+\colon R \times R \to R$ called *addition* and a second binary operation $\cdot\colon R \times R \to R$ called *multiplication* such that

(1) $(R, +)$ is an Abelian group;
(2) multiplication is associative, $(r \cdot s) \cdot t = r \cdot (s \cdot t)$ for all $r, s, t \in R$;
(3) the distributivity law is intact: $r(s+t) = r \cdot s + r \cdot t$ and $(r+s) \cdot t = r \cdot t + s \cdot t$ for all $r, s, t \in R$;
(4) there is a neutral element for multiplication, written 1_R, with $1_R \cdot r = r = r \cdot 1_R$ for all $r \in R$.

It is perhaps useful to make some comments here.

- We denote 0_R (or just 0) the neutral element for addition in R, and write $(-a)$ for the additive inverse of $a \in R$. Note the following two facts.

$$a \cdot 0 = a \cdot (0 + 0) = (a \cdot 0) + (a \cdot 0), \text{ so } a \cdot 0 = 0;$$
$$0 = a \cdot 0 = a \cdot (1 + (-1)) = a \cdot 1 + a \cdot (-1),$$

  so that $(-1) \cdot a$ is the additive inverse of a. We usually denote it by $-a$ and write $b - a$ for $b + (-1) \cdot a$.
- We will almost exclusively look at *commutative rings*, which are those where $r \cdot s = s \cdot r$ for all $r, s \in R$. But there is a general consensus that non-commutative rings are important enough for not being disqualified from the start.
- Some people no not require the existence of 1_R. Rings without multiplicative identity are not difficult to find, but they lack key features of rings that we want to discuss in our remaining chapters.
- One thing to note is something that a ring need not have, and that is multiplicative inverses. We are not saying that invereses *must not* exist (after all, 1_R is always its own inverse!); we just concede that they *may not* exist in all cases. Do not confuse $+$ and $\cdot$; $+$ is always commutative by definition.
- However, if an element $a \in R$ does have a multiplicative inverse, this inverse is unique, because if $a', a"$ both are inverses to $a \in R$ then $a' = a' \cdot e = a' \cdot (a \cdot a") = (a' \cdot a) \cdot a" = e \cdot a" = a"$.

EXAMPLE IX.2. Here is a list of standard rings that come up all the time in mathematics. The first three are all commutative.

- The rings after which all others are modelled is $(\mathbb{Z}, +, \cdot)$, the set of integers with usual addition and multiplication.
- The three collections $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ of rational, real and complex numbers respectively are all rings as well. They are rather special rings, since in contract to $\mathbb{Z}$, every non-zero number in these three rings does have a multiplicative inverse (whereas in $\mathbb{Z}$ that is only the case for ± 1.)
- The groups $\mathbb{Z}/n\mathbb{Z}$ are also all rings, with addition and multiplication of cosets.
- A collection of non-commutative *matrix rings* arises for each number n and choice of *coefficient ring* $\mathbb{K}$ as follows. Let $M_n(\mathbb{K})$ be the set of all $n \times n$ matrices with entries in $\mathbb{K}$. Then usual matrix addition and multiplication has the usual properties, which are those listed in Definition IX.1 above. Note that in general $A \cdot B \neq B \cdot A$ so that $M_n(\mathbb{K})$ is not commutative.
- Another collection of rings are the *polynomial rings* $\mathbb{K}[x_1, \ldots, x_n]$ over a chosen coefficients field $\mathbb{K}$. These are commutative rings, and their elements are the polynomials in the variables $x_1, \ldots, x_n$ which have coefficients in $\mathbb{K}$.
- A type of ring we will not look at much is popular in analysis: the set of all real-valued functions on the interval $[0, 1]$. Addition and multiplication is pointwise, which means that $(f + g)(x)$ is declared as $f(x) + g(x)$ and likewise for multiplication.

EXAMPLE IX.3. Here is an example of an *extension ring*. Let $\mathbb{Z}[\sqrt{-1}]$ be the set of complex numbers that have both real and imaginary value integer. So, $\mathbb{Z}[\sqrt{-1}] = \{a + b\sqrt{-1} \text{ with } a, b \in \mathbb{Z}\}$. You might want to think of this as a "vector space of dimension 2 over $\mathbb{Z}$, spanned by $1 \in \mathbb{Z}$ and $\sqrt{-1}$". So, we add componentwise: $(a + b\sqrt{-1}) + (c + d\sqrt{-1}) = (a + c) + (b + d)\sqrt{-1}$. Multplication has a bit of a surprise, as it does not go componentwise, but instead like for complex numbers in general: $(a + b\sqrt{-1}) \cdot (c + d\sqrt{-1}) = (ac - bd) + (bc + ad)\sqrt{-1}$. This is the *ring of Gaussian integers*.

DEFINITION IX.4. If in a ring R we have $a, b \in R$ both nonzero, but $ab = 0$, then we call a and b *zero-divisors*.

Most of the rings listed in examples here do not have zero-divisors. The exceptions are: $\mathbb{Z}/n\mathbb{Z}$ if n is not prime; $M_n(R)$ in the case $n > 1$ and also in the case that R itself has zero-divisors; the polynomial ring $R[x_1, \ldots, x_n]$ in the case that R has zero-divisors. You might want to check these three claims explicitly by finding one example of zerodivision in each of the three scenarios.

DEFINITION IX.5. A commutative ring that has no zero-divisors is called a *domain*.

Note that if $a \in R$ has an inverse, then a cannot be a zero-divisor. Indeed, if $ab = 1$ and $ca = 0$ then $c = c \cdot 1 = cab = 0 \cdot b = 0$.

DEFINITION IX.6. Consider $1_R, 1_R + 1_R, 1_R + 1_R + 1_R, \ldots$. This sequence might or might not contain the element 0_R. If it does, there is a smallest number $c \in \mathbb{N}_+$ such that adding 1_R c times gives 0_R. We call this c the *characteristic* of R.

If this sequence never produces 0_R we say that the *characteristic of R is zero*.

LEMMA IX.7. *If R is a domain, its characteristic is a prime number or zero.*

PROOF. Suppose $\underbrace{1+1+\ldots}_{c \text{ copies}} = 0$ and c is the characteristic (the smallest positive such c). Suppose $c = mn$ factors. Then let $e_m = \underbrace{1+1+\ldots}_{m \text{ copies}}$ and $e_n = \underbrace{1+1+\ldots}_{n \text{ copies}}$. Using the distributive property, $e_m \cdot e_n$ is the sum of mn copies of 1, hence zero. Since R is supposed to be a domain, it can't have zero-divisors, and so we must have $e_m = 0$ or $e_n = 0$. But if $c = mn$ is really a factorizatin of c, m and n are strictly less than t, which makes it impossible that $e_m = 0$ or $e_n = 0$. We conclude c does not factor and is therefore a prime number. □

DEFINITION IX.8. A commutative ring that has multiplicative inverses for each nonzero element is called a *field*.

We will discuss fields in detail later.

THEOREM IX.9. *If R has finitely many elements ("R is finite"), is commutative and is a domain, then it is a field.*

PROOF. We need to show that the absence of zero-divisors forces the presence of inverses when R is finite. Take $a \in R$ nonzero. then multiplication by a gives a permutation of the elements of R. Indeed, let $r_1, \ldots, r_t$ be the complete list of nonzero elements of R. Then $ar_1, \ldots, ar_t$ is another list of elements of R. No expression ar_i can be zero, sine $a \neq 0$ and $r_i \neq 0$, and R is supposed to be a domain. Also, there is no repetition on this list since if $ar_i = ar_j$ then $a(r_i - r_j) = 0$ and $a \neq 0$ now forces $(r_i - r_j) = 0$ as otherwise we would be looking at zero-divisors which can't exist in a domain. So, the second list is a permutatin of the first list, because both list all nonzero elements of R. (This is where finiteness is used: if R were infinite we could not argue like this. For example, the multiples of 2 are not a permutation of the nonzero integers. But in a finite set, if you list as many different elements of S as the set has, you listed them all). It follows, that one of the elements on the second list is 1_R, which amounts to saying that there is $r_i \in R$ with $a \cdot r_i = 1_R$. □

REMARK IX.10. A postscript of this proof goes like this: let R have p elements. Then the nonzero elements are a group with multiplication, since the theorem assures the esitence of inverses. This group has $p-1$ elements. So Lagrange says that if a is a nonzero element of R then its mutiplicative order divides $p-1$. In particular, there *is* a power c such that $a^c = 1_R$. But then $a^{c-1} \cdot a = 1_R$ and so the inverse of a is actually a power of a.

EXAMPLE IX.11. In the same way the Gaussian integers are an extension of the integers, one can mak extensions of fields. For example, $\mathbb{Q}[\sqrt{2}]$ is the collection of all expressions $a + b\sqrt{2}$ with $a, b \in \mathbb{Q}$. One adds componentwise, $(a+b\sqrt{2}) + (c+d\sqrt{2}) = (a+c) + (b+d)\sqrt{2}$ and multiplies according to $(a+b\sqrt{2}) \cdot (c+d\sqrt{2}) = (ac+2bd) + (bc+ad)\sqrt{2}$.

Note how one computes inverses here: $(a+b\sqrt{2})^{-1} = \frac{(a-b\sqrt{2})}{(a+\sqrt{2})(a-\sqrt{2})} = \frac{a-b\sqrt{2}}{a^2-2b^2} = \frac{a}{a^2-2b^2} + \frac{b}{a^2+2b^2}\sqrt{2}$ is of the required form. Recall that we proved that there cannot be rational numbers a, b with $a^2 = 2b^2$ and so the numerator is nonzero.

EXAMPLE IX.12. One can do this also with modular numbers. Let $R = (\mathbb{Z}/3\mathbb{Z})[\sqrt{2}]$. Here, $\sqrt{2}$ stands for a symbol whose square is the coset of 2. (Note that there is no element in $\mathbb{Z}/3\mathbb{Z}$ whose square is the coset of 2, just like there was no rational number whose square was 2.)

This is a ring with 9 elements, the possible expressions of the form $a + b\sqrt{2}$ with $a, b \in \mathbb{Z}/3\mathbb{Z}$. You calculate exactly as expected, always going modulo 3.

So for example, the inverse of $\overline{2} + \overline{1}\sqrt{2}$ is $\frac{\overline{1}}{\overline{2}+\overline{1}\sqrt{2}} = \frac{\overline{2}-\overline{1}\sqrt{2}}{(\overline{2}+\overline{1}\sqrt{2})(\overline{2}-\overline{1}\sqrt{2})} = \frac{\overline{2}-\overline{1}\sqrt{2}}{\overline{4}-\overline{2}} = \overline{1} - \overline{2}\sqrt{2} = \overline{1} + \overline{1}\sqrt{2}$. And indeed, $(\overline{2} + \overline{1}\sqrt{2})(\overline{1} + \overline{1}\sqrt{2}) = \overline{1} + \overline{0}\sqrt{2}$.

CHAPTER X

Week 10: Ideals and morphisms

Recall that we insist that our rings have a (multiplicative) 1. (All rings have a neutral element for + (which we write as 0), since R with + is a group).

DEFINITION X.1. A *ring morphism* is a function $f\colon R \to R'$ from one ring to another such that it is a morphism of groups $(R,+) \to (R',+)$, and moreover it respects ring multiplication: $f(r_1 r_2) = f(r_1)f(r_2)$.

Examples of such things abound.

- the inclusions $\mathbb{Z} \hookrightarrow \mathbb{Q} \hookrightarrow \mathbb{R} \hookrightarrow CC$ and the inclusions $\mathbb{Z} \hookrightarrow \mathbb{Z}[\sqrt{-1}] \hookrightarrow \mathbb{C}$;
- the surjection $\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ sending k to $k + n\mathbb{Z}$ for any n;
- if $m|n$, the surjection $\mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ sending $k + n\mathbb{Z}$ to $k + m\mathbb{Z}$;
- complex conjugation;
- the "conjugation" $\mathbb{Z}[\sqrt{2}] \to \mathbb{Z}[\sqrt{2}]$ sending $a + b\sqrt{2}$ to $a - \sqrt{2}$ and any similar constructs;
- the polynomial map $\mathbb{C}[x,y] \to \mathbb{C}[t]$ that sends $x \mapsto t^2, y \mapsto t^3$;
- If $\mathcal{O}$ is the collection of real functions defined on the real line, then any $a \in \mathbb{R}$ induces an *evaluation morphism* $\epsilon_a\colon \mathcal{O} \to \mathbb{R}$ that sends $f(x) \in \mathcal{O}$ to the value $f(a)$ of f at a.

EXAMPLE X.2. Recall that there are rings of positive characteristic. If $\mathrm{char}(R) = p > 0$ is prime, there is the *Frobenius* morphism $\mathrm{Frob}\colon R \to R$ that sends $r \in R$ to $\mathrm{Frob}(r) = r^p$. That this is then a morphism is due to *freshman's dream in algebra*: $(x+y)^p = x^p + y^p$ in characteristic p, since by the binomial theorem every missing term of $(x+y)^p$ is a multiple of p.

DEFINITION X.3. If $f\colon R \to R'$ is a ring morphism, its kernel is the elements of R that are sent to $0 \in R'$ by f.

DEFINITION X.4. An *ideal* in a ring R is a subset $I \subseteq R$ such that

- I is a subgroup of R with respect to addition;
- For all $x \in I$ and all $r \in R$, the product xr is in I.

REMARK X.5. A standard way of producing ideals is as follows. Let $f_1, \ldots, f_k$ be elements of the ring R. Then let I be the set of all *R-linear combinations* you can make from $f_1, \ldots, f_k$. In other words, I is made precisely of all things like $r_1 f_1 + \ldots + r_k f_k$ where $r_1, \ldots, r_k$ run through all possible elements in R. Then I is an ideal: sums of such things as well as differences of such things are such things again, and multiplying any such element by an arbitrary ring element gives another thing of this type.

It is important to note that it is allowed for an ideal to have infinitely generators. Often, one can simplify such a situation to finitely many generators, but not always. The rings we consider all will have only ideals that are finitely generated, but proving this can be dicey (although we will prove it in some nice cases).

For example, the multiples of 641 are an ideal of $\mathbb{Z}$. So are the $\mathbb{C}[x,y]$-linear combinations of x^3-y^2 and x^3+y^4 in $\mathbb{C}[x,y]$.

PROPOSITION X.6. *The kernel of a ring morphism is an ideal.*

PROOF. That the kernel of a ring morphism $f\colon R \to R'$ is a subgroup of R follows straight from the fact that f is a group morphism. Now take $x \in I$ and $r \in R$. Then $f(x) = 0$ and so $f(x)f(r) = 0$ and so $f(rx) = 0$ and so $rx \in \ker(f) = I$. □

We next turn this around and use ideals to make factor rings and morphisms.

DEFINITION X.7. Let $I \subseteq R$ be an ideal. The *factor ring* R/I is the group R/I together with multiplication $(x+I)(y+i) = xy+I$. There is an induced morphism $\pi\colon R \to R/I$ that sends $r \in R$ to $r+I$.

That this construction indeed produces a ring is not difficult to see. One basically needs to check that multiplication is well-defined (this means that if $x+I = x'+I$ and $y+I = y'+I$ then $xy+I = x'y'+I$, but that is quite easy.

If $f\colon R \to R'$ is a ring morphism and I an ideal of R and J an ideal of R', then inspection shows that

- $f(I)$ may not be an ideal in R' (for example, $2\mathbb{Z}$ is an ideal in $\mathbb{Z}$ but when you inject $\mathbb{Z} \hookrightarrow \mathbb{R}$ then the even integers are no longer an ideal; make sure you believe this, it is due to the fact that products of integers and reals are often not integer).
- the *preimage* $f^{-1}(J)$ in contrast is always an ideal of R. This is seen as follows. Since f is a group morphism, the preimage is a group. Take $x \in f^{-1}(J)$ and $y \in R$. Then $f(xy) = f(x)f(y) \in J \cdot R' = J$ and so $xy \in f^{-1}(J)$ as required.
- If $\mathrm{char}(R) = n$ then there is a natural morphism $\mathbb{Z}/n\mathbb{Z} \hookrightarrow R$ induced by sending $1 \in \mathbb{Z}$ to $1 \in RR$ and using the morphism rule.

The main structure theorem for ideals says:

THEOREM X.8. *If I is an ideal of R then under the natual surjection $\pi\colon R \to R/I$ the ideals of R/I correspond to the ideals of R that contain I. More precisely, if J is an ideal of R that contains I then the quotient group J/I is an ideal of R/I. In reverse, if J/I is an ideal of R/I then the preimage $f^{-1}(J/I)$ is an ideal of R.*

For example, if $R = \mathbb{Z}$ and $I = 6\mathbb{Z}$ then R/I has 4 ideals: the whole ring $R' = \mathbb{Z}/6\mathbb{Z}$, the zero idewal $\{0+6\mathbb{Z}\}$ and two interesting ideals $J_2 = \{0+6\mathbb{Z}, 2+6\mathbb{Z}, 4+6\mathbb{Z}\}$ and $J_3 = \{0+6\mathbb{Z}, 3+6\mathbb{Z}\}$. To J_2 corresponds the ideal $2\mathbb{Z}$ of R, and it indeed contains I. To J_3 corresponds the ideal $3\mathbb{Z}$ of R, and indeed it contains I. The only ideals that contain I are $I, 2\mathbb{Z}, 3\mathbb{Z}, \mathbb{Z}$. The first of these corresponds to the zero idealin R/I and the last one to the whole of R/I.

We come now to talk about certain special types of ideals.

DEFINITION X.9. A *prime ideal* of a ring R is an ideal P such that if $a, b \in R$ with $ab \in P$ then at least one of a, b is in P.

Being a prime ideal is equivalent to saying that R/P is a domain. (There are $a, b \in R$ but not in P such that $ab \in P$ if and only if in R/P we have $(a+P)(b+P) = 0+P$ which can happen if and only if R/P is not a domain).

DEFINITION X.10. An ideal M of R is *maximal* if there is no other ideal between M and R. So, M is as large as it can be without equaling R.

REMARK X.11. If you think of an ideal as generated by some elements of R, then primeness and maximality can change withthe ring. For example, the multiples of 3 form a prime ideal in $\mathbb{Z}$, but in $\mathbb{Z}[\sqrt{-2}]$ you can factor $3 = (1+\sqrt{-2})(1-\sqrt{-2})$. Similarly, maximality can toggle: the multiples of 2 are a maximal ideal in $\mathbb{Z}$, but in $\mathbb{Q}$ the multiples of 2 are all of $\mathbb{Q}$, which does nto qualify as maximal.

By the main structure theorem on factor rings, the ideal I is maximal if and only if R/I has only two ideals, I/I and R/I.

DEFINITION X.12. A commutative ring with only two ideals is called a *field*. A non-commutative ring with only two ideals is a *skew-field*.

If a ring has only two ideals, one has to be the ideal $\langle 0\rangle$, and the other the ideal $R = \langle 1\rangle$. So, in a field it is clear what the two ideals involved are. In any ring, the zero ideal and the whole ring are considered "improper ideals". Not in the sense that they are running around naked, but in the sense that we don't want to truly (properly) call interesting ideals.

LEMMA X.13. *In a field, every nonzero element is invertible. In particular, fields are domains.*

PROOF. Let $0 \neq x \in R$ and let I be the ideal defined by x (so, I consists precisely of all the multiples of x). Since $f \neq 0$, I is not $\langle 0\rangle$. So, as we are in a field, $I = R$. This means in particular that $1_R \in I$ and so 1_R is a multiple of f. But then f must be invertible.

If a field R is not a domain, then $0 = ab$ for some $a \neq 0 \neq b$ in R. But as a field, R contains an inverse for a and then $a^{-1}ab = a^{-1}0$ leads to a contradiction. □

Examples of fields include $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ but also things like $\mathbb{Z}/p\mathbb{Z}$ with p prime:

LEMMA X.14. *If $p \in \mathbb{Z}$ is prime then $\mathbb{Z}/p\mathbb{Z}$ is a field and conversely.*

PROOF. We look at the morphism $\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ that sends 1 to $1 + p\mathbb{Z}$. Then the zero ideal in $\mathbb{Z}/p\mathbb{Z}$ corresponds to the ideal $\langle p\rangle$ of $\mathbb{Z}$ by the theorem on factor rings, and we need to show that there is no ideal of $\mathbb{Z}$ strictly between $p\mathbb{Z}$ and $\mathbb{Z}$. Suppose we have an ideal I with $p\mathbb{Z} \subseteq I$ but I is strictly greater than $p\mathbb{Z}$. Then I contains all multiples of p, and at least one number a that is not a multiple of p. The Euclidean algroithm says that $1 = \gcd(a, p)$ can be written as a linear combination $1 = ax + py$ with a, b integers. Thus, in $\mathbb{Z}/p\mathbb{Z}$, $a + p\mathbb{Z}$ and $x + p\mathbb{Z}$ are inverses, and in particular $I/p\mathbb{Z}$ contains $1 + p\mathbb{Z} = (a + p\mathbb{Z})(x + p\mathbb{Z})$. So, $I/p\mathbb{Z}$ is in fact $\mathbb{Z}/p\mathbb{Z}$ and hence $\mathbb{Z}/p\mathbb{Z}$ is a field.

On the other hand, if p is not prime and can be factored as $p = mn$ with m, n not units, then in $\mathbb{Z}/p\mathbb{Z}$ we have $(m + p\mathbb{Z})(n + p\mathbb{Z}) = 0 + p\mathbb{Z}$ and so neither factor can have an inverse. But $m + p\mathbb{Z}$ is not zero since p does nbot divide m (since n is not unit). So $\mathbb{Z}/p\mathbb{Z}$ can't be a field. □

More generally,

PROPOSITION X.15. *Let I be an ideal of R.*

(1) I is a prime ideal if and only if R/I is a domain;
(2) I is a maximal ideal if and only if R/I is a field.

PROOF. The second claim, as mentioned previously, follows directly from the structure theorem of factor rings. The proof for the first claim is analogous to the proof of the preceding lemma.Namely, if I is a prime ideal and $(a+I)(b+I) = 0+I$ in R/I then we must have $ab \in I$ and so by primeness of I one of a, b is in I, and thus one of $a + I, b + I$ is zero in R/I. If I is not prime, there are $a, b \in R$ that are not in I but with $ab \in I$. Then $(a + i)(b + I) = 0 + I$ are zerodivisors and so R/I is not a domain. □

THEOREM X.16. *Every ideal in $\mathbb{Z}$ and in $\mathbb{Z}/n\mathbb{Z}$ is generated by one element.*

PROOF. Suppose the ideal $I \subseteq \mathbb{Z}$ contains a and b. By the Euclidean algorithm, it also contains their gcd g. On the other hand, a, b are multiples of g and so we see that any ideal that contains a, b also contains $\gcd(a, b)$ and conversely.

Iterating this argument, $\langle a, b, c\rangle = \langle \gcd(a, b), c\rangle = \langle \gcd(a, b, c)\rangle$, and $\langle a, b, c, d\rangle = \langle \gcd(a, b), c, d\rangle = \langle \gcd(a, b, c), d\rangle = \langle \gcd(a, b, c, d)\rangle$, and in this way every finite generator set $a_1, \ldots, a_k$ for an ideal can be replaced by the single generator given by the gcd of all a_i.

Now imagine an infinite list $a_1, a_2, \ldots, a_n, \ldots$. We know that

$$\gcd(a_1) \geq \gcd(a_1, a_2) \geq \gcd(a_1, a_2, a_3) \ldots \geq 0.$$

It follows that this sequence of $\geq$ symbols reaches a point (say, when the index is k) from where onwards each $\geq$ is actually a $=$.

What this means is that $\gcd(a_1, \ldots, a_k)$ divides $a_{k+1}, a_{k+2}, \ldots$. But then $a_{k+1}, a_{k+2}, \ldots$ are already in the ideal generated by $a_1, \ldots, a_k$ and we can say that

$$\langle a_1, a_2, \ldots, a_n, \ldots\rangle = \langle a_1, ldots, a_k\rangle = \langle \gcd(a_1, \ldots, a_k)\rangle$$

is a generated by one element. □

DEFINITION X.17. Ideals generated by one element are called *principal*. The theorem says that $\mathbb{Z}$ has only principal ideals. Since $\mathbb{Z}$ is a domain (has no zerodivisors), it is referred to as a *principal ideal domain*.

REMARK X.18. You will prove in homework that ideals in $\mathbb{Z}/n\mathbb{Z}$ are also all principal.

CHAPTER XI

Week 11, Euclidean rings

We start with polynomial rings. As before, rings are commutative (unless expressly indicated not to be) and have a 1.

DEFINITION XI.1. Let R be any ring and x a symbol (distinct from any element of R). We let x^i, for $i \in \mathbb{N}$ be a new symbol and we postulate that the symbols $x^0, x^1, x^2, \ldots$ are linearly independent over R. (Of course, we think of them as powers of x, but what really is a power of a symbol???) Then x is an *indeterminate* over R. We abbreviate x^1 to x and identify x^0 with 1_R.

A *polynomial $f(x)$ in x with coefficients in R* is an infinite series $f(x) = \sum_{i=0}^{\infty} r_i x^i$ in which almost all coefficients r_i are zero. In other words, only finitely many coefficients are allowed to be nonzero.

The collection of all these polynomials is denoted $R[x]$ and called the *polynomial ring in x over R*.

We consider two such expressions equal,

$$\sum_{i=0}^{\infty} r_i x^i = \sum_{i0}^{\infty} r'_i x^i$$

if and only if we have $r_i = r'_i$ for all i. Note that for large i this is automatic since eventually all coefficients are zero.

Given a polynomial $\sum_{i=0}^{\infty} r_i x^i$, there is a largest index d for which r_d is nonzero, and this index d we call the *degree* $\deg(f)$ *of the polynomial*. If $d = \deg(f)$ then we usually write $r_0 + r_1 x + \ldots + r_d x^d$ instead of $\sum_{i=0}^{\infty} r_i x^i$, and call r_d the *leading coefficient* $\operatorname{lc}(f)$ of $f(x)$. .

We add polynomials degree by degree:

$$\sum_{i=0}^{\infty} r_i x^i + \sum_{i=0}^{\infty} r'_i x^i = \sum_{i=0}^{\infty} (r_i + r'_i) x^i.$$

We multiply them according to $x^i x^j = x^{i+j}$ and extend by requiring linearity.

It is easy to see that these two operations make $R[x]$ into a commutative ring, with zero element $0x^0 + 0x^1 + 0x^2 + \ldots$ and 1-element $1x^0 + 0x^1 + 0x^2 + \ldots$.

REMARK XI.2.
- $\deg(fg) = \deg(f) + \deg(g)$ and $\deg(f+g) \leq \max(\deg(f), \deg(g))$.
- $\operatorname{lc}(fg) = \operatorname{lc}(f) \cdot \operatorname{lc}(g)$. That implies that if R is a domain then also $R[x]$ is a domain since $fg = 0$ implies $\operatorname{lc}(f) \operatorname{lc}(g) = 0$.

1. Euclidean rings

DEFINITION XI.3. A domain has a *Euclidean measure* if there is a function $\delta \colon R \setminus \{0\} \to \mathbb{N}$ that satisfies

(1) $\delta(a) \leq \delta(ab)$ for all $a \neq 0 \neq b$ in R;

(2) if $a, b \in R$ are given, there is an oracle that finds $q, r \in R$ with $a = bq + r$ and either $r = 0$ or $\delta(b) > \delta(r)$.

EXAMPLE XI.4. We already know some examples like this.

- $R = \mathbb{Z}$, with Euclidean measure $\delta(n) = |n|$ the absolute value. This works because for any $a \in \mathbb{Z}$ and $0 \neq b \in \mathbb{Z}$ there is some multiple qb of b with $q \in \mathbb{Z}$ such that $|a - qb$ is less than $|a|$.
- If R is a polynomial ring over a field, we can take $\delta(f) = \deg(f)$. This works because the remainder of a by b with division leaves a rest r of degree less than $\deg(b)$.
- As you will check in HW, the ring $\mathbb{Z}[\sqrt{-1}]$ is also equipped with a Euclidean measure, $\delta(a + b\sqrt{-1}) = a^2 + b^2$.

THEOREM XI.5. *A domain R with Euclidean measure is a Euclidean ring (has a Euclidean algorithm).*

PROOF. Let δ be a Euclidean measure on the domain R and pick $a, b \in R$. If $b = 0$ there is nothing to do, since $\gcd(a, 0) = a$. If $b \neq 0$, according to the definitions, there are $q, r \in R$ with $a = bq + r$ and either $r = 0$ or $\delta(r) < \delta(b)$.

Let inductively $a_0 = a, a_1 = b, q_0 = q, r + 0 = r$. For each index i for which a_i, b_i, q_i, r_i have already been found with $b_i \neq 0$, define $a_{i+1} = b_i, b_{i+1} = r_i$ and choose q_{i+1}, r_{i+1} so that $a_{i+1} = q_{i+1}b_{i+1} + r_{i+1}$ with either $r_{i+1} = 0$ or $\delta(r_{i+1} < \delta(b_{i+1})$.

Note that this scheme is set up so that $\gcd(a_i, b_i) = \gcd(a - q_i b_i, b_i) = \gcd(r_i, b_i) = \gcd(b_{i+1}, a_{i+1})$. So, the gcs of a, b is the same as that of a_i, b_i for all i.

Since $\delta(b_i) > \delta(r_i) = \delta(b_{i+1})$, the sequence $\{\delta(b_i)\}$ is strictly descending. But they are all natural numbers (since δ can only have natural output by definition.

This seems to be a contradiction, since no eternally strictly descending chains of natural numbers can exist. The only way out is that at some point b_i was zero, since then we would not try to go another round.

Now $b_i = 0$ means $r_{i-1} = 0$ and so $a_{i-1} = q_{i-1}b_{i-1}$. But then clearly $\gcd(a_{i-1}, b_{i-1}) = b_{i-1} = r_{i-2}$ and we have found the gcd of a, b using repeatedly the oracle of the Euclidean measure. □

Note that one can now use back substitution from $a_{i-2} = q_{i-2}b_{i-2} + r_{i-2}$ to rewrite r_{i-2} as linear combination of a_{i-3}, b_{i-3} and then of a_{i-4}, b_{i-4}, etc, and finally as linear combination of a and b.

COROLLARY XI.6. *If R has a Euclidean measure then for all $a, b \in R$ one can find $x, y \in R$ such that* $\gcd(a, b) = ax + by$.

EXAMPLE XI.7. Let's find $\gcd(a := 3x^2 + 4x + 3, b := 4x^2 + 2x + 4)$ in $\mathbb{Z}/5\mathbb{Z}[x]$. (Strictly speaking, I should write bars over every number, but maybe we can liuve without that for a moment).

We have (keep in mind that $2 \cdot 3 = 1$

$$\begin{aligned} a &= \frac{4}{3}b + (3x^2 + 4x + 3 - \frac{3}{4}(4x^2 - 2x - 4)) \\ &= \frac{4}{3}b - (0x^2 - 20x + 0) \end{aligned}$$

and modulo 5, 20=0. So actually, $a = 4b/3$. This is a warning that in modular arithmetic it is not easy to see whether two polynomials are multiples of one another.

EXAMPLE XI.8. Let's do one that is a bit more thrilling. Let's compute gcd of $x^{10}-1$ and of x^6-1 in $\mathbb{Q}[x]$.

$$\begin{aligned} x^{10}-1 &= x^4(x^6-1)+(x^4-1).\\ x^6-1 &= x^2(x^4-1)+(x^2-1).\\ x^4-1 &= x^2(x^2-1)+(x^2-1),\\ &\text{and so}\\ x^4-1 &= x^2(x^2-1)+1(x^2-1)=(x^2+1)(x^2-1)+0. \end{aligned}$$

So, $\gcd(x^{10}-1,x^6-1)=\gcd(x^6-1,x^4-1)=\gcd(x^4-1,x^2-1)=\gcd(x^2-1,0)=x^2-1$.

In Euclidean rings, a lot is like in $\mathbb{Z}$.

THEOREM XI.9. *In Euclidean rings, all ideals are principal.*

PROOF. This goes parallel to the proof of Theorem X.16, where we really used only that $\mathbb{Z}$ has a Euclidean algorithm. The main idea (as shown there) is that for any sequence $a_1,a_2,\ldots$ of generators of an ideal we have $\langle a_1,a_2,\ldots\rangle = \langle\gcd(a_1,a_2),a_3,a_4,\ldots\rangle = \langle\gcd(a_1,a_2,a_3),a_4,a_5,\ldots\rangle$ and that $\delta(a_1)\geq\delta(\gcd(a_1,a_2))\geq\delta(\gcd(a_1,a_2,a_3))\geq\cdots$.

This decreasing sequence has to level off, since it can't go down indefinitely. It follows that from some index i onwards, $\delta(\gcd(a_1,\ldots,a_i))=\delta(\gcd(a_1,\ldots,a_j))$ for all $j>i$. This in turn implies that $a_{i+1},\ldots$ are all divisible by $g=\gcd(a_1,\ldots,a_i)$. But then the ideal generated by al a_k is the same as the ideal of just $a_1,\ldots,a_i$, and this is just the multiples of g. □

THEOREM XI.10. *In a Euclidean ring, "prime" and "irreducible" are the same concepts.*

PROOF. Recall that prime things are always irreducible. So we need to show that irreducible elements are prime. Let $p\in R$ be irreducible. Take a product ab that is a multiple of p. Suppose p does not divide a, and try to show it must divide b.

Let $g=\gcd(a,p)$ and find with the Euclidean algorithm $x,y\in R$ with $ax+py=g$. Since g is the end product of the Eucidean algorithm of a,p, we can say that $\delta(p)>\delta(g)$. As g does divide p we can find $h\in R$ with $gh=p$. If h is a unit, $g=ph^{-1}$ and so p would divide g, but then p would also divide a which we know to be false. So, h is not unit, but then irreducibility of p implies that g is a unit. Then $g=ax+py$ gives $b=bg^{-1}g=bg^{-1}(ax+py)=g^{-1}(abx+pby)$ is a multiple of p (since ab is). But that is what we wanted to show. □

EXAMPLE XI.11. The ring of polynomials with integer coefficients $R=\mathbb{Z}[x]$ is not Euclidean.

PROOF. On the face of it, this seems almost unprovable, since we are required to show that one cannot put a Euclidean measure on R. The strategy is therefore to say "if R were Euclidean, it should have some properties that follow from Euclideanness, and maybe we can find one such property that R does not have".

Above we proved that in Euclidean rings all ideals are principal. So if we find an ideal in $\mathbb{Z}[x]$ that is not principal, R can't be Euclidean. Let's look at the ideal

geerated by 2 and x, $I = \{2a + xb | a, b \in R\}$. Let is assume for the moment that I is principal, generated by the polynomial f. So that means that 2 is a multiple of f and also x is a multiple of f,

$$2 = fg, \qquad x = fh,$$

with $g, h \in R$. Plugging $x = 0$ into the second equation, $0 = f(0)h(0)$ and so one of $f(0)$ and $h(0)$ has to be zero. Plugging $x = 0$ into the first equation, $2 = f(0)g(0)$ and the says that $f(0)$ is not zero, hence $h(0) = 0$. But then h is a multiple of x, $h = xk$ with $k \in R$. Together then, $x = fh = fkx$ says that $1 = fk$ when dividing out x. That says that the ideal $\langle f \rangle$ of multiples of f contains 1. Since we labor under the belief that $\langle 2, x \rangle = \langle f \rangle$, 1 should be a linear combination of 2 and x, $1 = 2a + xb$ with $a, b \in R$. Then evaluation at $x = 0$ gives $1 = 2a(0)$ which is not possible since $a(0)$ is an integer.

It follows that $\langle 2, x \rangle$ is not principal and so R cannot have a Euclidean algorithm and thus cannot have a Euclidean measure. □

REMARK XI.12. A domain in which every ideal is principal is called a *principal ideal domain*, PID for short. We have seen that Euclidean rings (ER for short) are PIDs. And we have seen that in Euclidean rings the notion of primeness and of irreducibility agree. This can be used to show that in a Euclidean ring every element has a decomposition into prime factors, just like in $\mathbb{Z}$. Such rings are called *unique factorization domains*, UFD for short. As it turns out, a PID always has the UFD property, so there is a sequence of implications

$$[R \text{ is a ED}] \Rightarrow [R \text{ is a PI}] \Rightarrow [R \text{ is a UFD}] \Rightarrow [R \text{ is a domain}].$$

One can show that each implication is strict, so there are domains that are not UFDs, and there are UFDs that are not PIDs, and there are PIDs that are not ERs.

Recall that we called *norm* on a ring R a function $N \colon R \to \mathbb{N}$ for which $N(rs) = N(r) \cdot N(s)$, and $[N(r) = 0] \Leftrightarrow [r = 0]$, and $[N(r) = 1] \Leftrightarrow [r \text{ is a unit}]$.

EXAMPLE XI.13. (1) The ring $\mathbb{Z}[\sqrt{-5}]$ is a domain but not a UFD. To see this, note that it has a multiplicative norm function given by complex absolute value, squared: $N(a + b\sqrt{-5}) = a^2 + 5b^2$.

Now $2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$ are two different factrizations of the number 6. Note that $N(2) = 4, N(3) = 9, N(1 \pm \sqrt{-5}) = 6$, and we use this to show that the factors $2, 3, 1 \pm \sqrt{-5}$ are irreducible.

For example, if 2 could be factored, $2 = rs$ with $rs \in \mathbb{Z}[\sqrt{-5}]$, then $4 = N(r)N(s)$. The point of the norm is that we are now down to *integer* arithmetic only. So, either $N(r) = 1$ or $N(r) = 1$ or $N(r) = 2$. The latter case is impossible since no expression $a^2 + 5b^2$ can ever be 2 (with integer a, b). On the other hand $N(a + b\sqrt{-5}) = 1$ implies $b = 0$ and $a = \pm 1$. So the only factorization of 2 is as product of ± 1 with ± 2. So 2 is irreducible.

For $3, 1 \pm \sqrt{-5}$ the calculations are similar (see HW).

(2) The ring $\mathbb{Z}[x]$ is not a PID from the example above; we'll prove the UFD property below.

(3) The ring $\mathbb{Z}[(1 + \sqrt{-19})/2]$ is a PID but not a Euclidean ring. That this is so is a bit out of the realm of this course, you need to know a bit of what is called *number theory*.

DEFINITION XI.14. If R is a domain, then its *ring of fractions* is the ring whose elements are fractions of the form f/g with $f, g \in R$ but g nonzero. Addition and muliplication are exactly as you would think.

So, for example, the ring of fractions of the domain $\mathbb{Z}$ is the ring of rational numbers, and the ring of fractions of the polynomial ring $\mathbb{R}[x]$ is the ring of rational functions with real coefficients.

Let us note that in a ring of fractions, f/g has inverse g/f unless $f = 0$. This means that a ring of fractions of a domain is actually a field. Note also that there is an inclusion of rings of R into its ring of fractions that sends $f \in R$ to the fraction $f/1_R$. This is the natural generalization of the inclusion $\mathbb{Z} \hookrightarrow \mathbb{Q}$ via $z \mapsto z/1$.

The notion of a ring of fractions comes up in the proof of the next result.

THEOREM XI.15 (The Gauß Lemma). *If R is a domain and has unique factorization, then so does $R[x]$.*

PROOF. The idea is as follows. Let $\mathbb{K}$ be the ring of fractions of R. Then we have an inclusion $R[x] \hookrightarrow \mathbb{K}[x]$ that is a ring morphism. Given $f(x)$ a polynomial in $R[x]$, we can now also read it as a polynomial in $\mathbb{K}[x]$. But as $\mathbb{K}$ is a field, we have shown that $\mathbb{K}[x]$ is Euclidean, and therefore a UFD. So, in $\mathbb{K}[x]$ we can uniquely factor $f(x) = g_1(x) \cdots g_k(x)$ where each $g_i(x)$ is a polynomial in x with coefficients in $\mathbb{K}$, and no $g_i(x)$ can be factored further in $\mathbb{K}[x]$.

The question is how to translate this back into $R[x]$. The problems are: first off, no $g_i(x)$ might be in $R[x]$ (because of the fractions in the coefficients); secondly, if we ever manage to make a translation, why is the resulting factorization for $f(x)$ in $R[x]$ unique?

Skipping all of the details, the main part of the work consists now in showing that one can rearrange the denominators in the various $g_i(x)$ such that after the rewriting all factors have coefficients in R. In other words, if a product of polynomials with fractional coefficients only has "whole" coefficients, then one rewrite to a factorization with whole coefficients in each factor. For example, we can take x^1- in $\mathbb{Z}[x]$ and rewrite in $\mathbb{Q}[x]$ as $(2x+1)(x/2-1/2)$, but by moving around the $1/2$ we can also rewrite to $x^2 - 1 = (x+1)(x-1)$.

The official statement to be proved is:

LEMMA XI.16. *If $f \in R[x]$ can be factored as $f(x) = g(x)h(x)$ with $g, h \in R[x]$, then any prime element $p \in R$ that divides f coefficient by coefficient, must divide one of g or h coefficient by coefficient.*

The proof of the lemma proceeds by an iterated induction on the degrees of f, g, h.

With the lemma in hand one can prove that a factorization of f in $\mathbb{K}[x]$ always yields a related factorization in $R[x]$. Uniqueness is then rather easy. You might look at the proof of the Gauß Lemma in anyt textbook, if you are curious. □

CHAPTER XII

Divisibility, Field Extensions

1. Divisibility

Let R be a commutative ring with 1, and take an element $f(x)$ in the polynomial ring $R[x]$. For ever $r \in R$ there is an *evaluation morphsim*

$$\varepsilon_r \colon R[x] \to R$$

that sends $f(x)$ to the element $f(r)$ in R. It is immediately clear that if a polynomial $f(x)$ is a multiple of $x - r$ then $\varepsilon_r(f) = 0$ simply because $\varepsilon_r(x - r) = 0$. So, the kernel of ε_r contains at least all multiples of $x - r$ (that is, the ideal generated by $x - r$).

It turns out that this kernel is precisely the ideal generated by $x - r$. The argument is the following. Write $f(x) = a_0 + a_1 x + \ldots + a_d x^d$, d the degree of f, and suppose $\varepsilon_r(f) = 0$. Since $\varepsilon_f(x - r) = 0$ as well, then for arbitrary $g(x) \in R[x]$ we also have $\varepsilon_r(f(x) - g(x) \cdot (x - r)) = 0$, since we can do the plug-in process separately in the two polynomials.

Let's pick a $g_1(x)$ in such a way that $f_1(x) := f(x) - g_1(x) \cdot (x - r)$ has degree less than d. By construction, $\varepsilon_r(f) = \varepsilon_r(f_1)$. Now repeat: find $g_2(x)$ such that $f_2(x) := f_1(x) - g_2(x) \cdot (x - r)$ has degree less than $\deg(f_1)$. Keep going. At the end of the day, this must stop, because when you found a $f_k(x)$ that is constant, you can't keep the iteration going.

We have $\varepsilon_r(f) = \varepsilon_r(f_1) = \varepsilon_r(f_2) = \ldots = \varepsilon_r(f_k)$ and that $f_k(x)$ is a constant. But as a constant, pluggin in has no effect. So $\varepsilon_r(f) = f_k$. This says that the remainder that you get when you divide $f(x)$ by $x - r$ (that is hat $f_k(x)$ really is!) is precisely the value of $f(x)$ at input $x = r$.

LEMMA XII.1. *Let $f(x) \in R[x]$ and choose $r \in R$. The value $f(r)$ is the remainder of division of $f(x)$ by $x - r$.*

Going back to the kernel of our morphism ε_r, this lemma says that: $f(x) \in \ker(\varepsilon_r)$ happens if and only if $f(x)$ has remainder zero when dividing by $x - r$. But the latter statement is just a euphemism for "$f(x)$ is a multiple of $x - r$. So,

$$\ker(\varepsilon_r) = R[x] \cdot (x - r).$$

DEFINITION XII.2. If $f(x) \in \ker(\varepsilon_r)$ we call r a *root of $f(x)$ in R.*

Roots can be funny.

EXAMPLE XII.3. (1) The roots of $x^2 - 1$ in $\mathbb{Z}/12\mathbb{Z}$ are $1 + 12\mathbb{Z}.5 + 12\mathbb{Z}, 7 + 12\mathbb{Z}, 11 + 12\mathbb{Z}$. So a degree 2 polynomial can have more than 2 roots. The culprit is the fact that $\mathbb{Z}/12\mathbb{Z}$ is not a domain. Note that this is also reflected in possible factorizations: $x^2 - 1 = (x - 1)(x + 1) = (x - 5)(x - 7)$ in $\mathbb{Z}/12\mathbb{Z}$.

(2) $(x + 1)^2$ has only one root, -1, but with *multiplicity two.*

(3) The roots of x^2-1 in $\mathbb{Z}/2\mathbb{Z}$ are 1 and 1 again, since $(x-1)^2 = x^2-1$ modulo 2. So 1 is a double root.

(4) x^2+1 has no roots in $\mathbb{Q}$.

THEOREM XII.4. *If R is a domain, then any polynomial $f(x)$ has at most $\deg(f)$ roots in R, even when counting with mutiplicity.*

PROOF. If r_1 is a root of $f(x)$ then we can write $f(x) = (x-r_1)\cdot f_1(x)$ because of the lemma above. Iterate this to get that $f(x) = (x-r_1)(x-r_2)\dots(x-r_k)f_k(x)$, and we can keep going with this until $f_k(x)$ does not have any roots in R.

Now suppose that perhaps $f(x)$ has yet another root r. Then $x-r$ divides $f(x)$, and so it divides $(x-r_1)(x-r_2)\dots(x-r_k)f_k(x)$. At this point we use that polynomial rings over fields are Euclidean rings, and in particular are UFDs. The way we use this is: $x-r$ can't be a product of two polynomials (since its degree is 1) except for products of the form (unit of R) times (f divided by that unit). So, $x-r$ is irreducible. Since $R[x]$ is a UFD, $x-r$ is prime. As $x-r$ is prime and divides the product $(x-r_1)(x-r_2)\dots(x-r_k)f_k(x)$, it must divide one of the factors of this product. The factor it divides can't be $f_k(r)$ because then $f_k(x)$ should have root r, but we agreed that $f_x(x)$ has no root. So, $x-r$ divides some $x-r_i$. In other words, $r = r_i$. □

In a while, we will go and try to manufacture new fields from old. As a stepping stone we need to know when a polynomial is irredecble. We will be mainly concerned with $R = \mathbb{Z}$ and $\mathbb{Z}/p\mathbb{Z}$ with prime $p \in \mathbb{Z}$. Note that over $R = \mathbb{Z}/p\mathbb{Z}$ we can actually go and test all elements of the field on whether they are roots as there are finitly many things to test. Over $\mathbb{Q}$ that is much harder. Here is a basic test for irreducibility.

LEMMA XII.5. *Let $f(x) \in \mathbb{Z}[x]$ be given, and assume that the coefficients of $f(x)$ have no common factor. If you can find a prime number $p \in \mathbb{Z}$ such that $f(x)$ mod p is irreducible and of the same degree as f, then f is irreducible in $\mathbb{Z}[x]$ and even in $\mathbb{Q}[x]$..*

PROOF. The Gauß Lemma says that if we can show that $f(x)$ is irreducible in $\mathbb{Z}[x]$ then it is also irreducible in $\mathbb{Q}[x]$. So we focus on irreducibility on $\mathbb{Z}[]x$.

Suppose $f = gh$ with $g, h \in \mathbb{Z}[x]$. Then take this equation and reduce modulo p to get $f(x) \bmod p = (g(x) \bmod p)(h(x) \bmod p)$. Since $f(x) \bmod p$ is supposed to be irreducible, this new equation must have one of $g \bmod p, h \bmod p$ be a unit. But units must have degree zero. So between $g(x) \bmod p$ and $h(x) \bmod p$, one has degree zero. However, that means that the other factor must have degree $\deg(f \bmod p) = \deg(f)$, and so one of g or h themselves has degree $\deg(f)$. That now means that the other one of g, h has degree zero and so is an integer.

We have shows that $f(x)$ can only be factored in $\mathbb{Z}[x]$ as (integer) times (polynomial of degree $\deg(f)$). Since the coefficients of f have no common factor by hypothesis, the integer factor is a unit and we are done. □

REMARK XII.6. The lemma says that irreducibility "lifts" from $\mathbb{Z}/p\mathbb{Z}[x]$ to $\mathbb{Z}[x]$. It is not true that reducibility also lifts. Many polynomials are reducible modulo p but irreducible over $\mathbb{Z}$. There are even polynomials in $\mathbb{Z}[x]$ that are irreducible but become reducible modulo *every* prime p. You will work through one such example ($f(x) = x^4+1$) in the homeworks.

Moreover, is pretty complicated to predict whether the reduction modulo some prime p of a polynomiual $f(x)$ will give you soemthing irreducible. For example, x^2+2 is irreducible over $\mathbb{Z}$ (since $\sqrt{-2}$ is not an integer) and does not factor modulo 5, but does factor over $\mathbb{Z}/2\mathbb{Z}$ since there it is just x^2. But there are much more "non-obvious" factorizations as the above-mentioned homework problem shows.

It would be good to have a test for irreducibilty based on reduction modulo a prime where one can see right away that it will work.

THEOREM XII.7 (Eisenstein Criterion). *Let $f(x) = a_0 + \ldots + a_d x^d \in \mathbb{Z}[x]$ be of degree d with no common factor in the coefficients. If there is a prime $p \in \mathbb{Z}$ with*

(1) $p \not| a_d$,
(2) $p | a_i$ for $i = 0, \ldots, d-1$,
(3) $p^2 \not| a_0$,

then f is of degree d in $\mathbb{Z}/p\mathbb{Z}[x]$, and irreducible in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.

PROOF. Let's assume that $f = gh$ in $\mathbb{Z}[x]$ and find a contradiction.

Since $p \not| a_d$, the degree of f mod p is also d. In fact, since p divides all coefficients except the top one, $f(x)$ mod $p = x^d$ mod p. So, in $\mathbb{Z}/p\mathbb{Z}[x]$, which is a Euclidean domain, and thus a UFD, the only factorizations of f mod $p =$ (g mod p)(h mod p) are of the type $g = x^{d-k}$ mod $p, h = x^k$ mod p. So, there are $\alpha, \beta \in \mathbb{Z}[x]$ with $g = x^{n-k} + p\alpha, h = x^k + p\beta$. But this implies that the constant term of $f = gh$ is twice divisible by p, hence by p^2. And there is our contradiction.

So f is irreducible over $\mathbb{Z}[x]$. Ireducibility over $\mathbb{Q}[x]$ follows now from the Gauss Lemma. □

EXAMPLE XII.8. Let $f(x) = x^n - p$. It fits the Eisenstein conditions, so is irreducible over $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.

DEFINITION XII.9. For $n \in \mathbb{N}$ set $\Phi_n(x) \in \mathbb{Z}[x]$ be the *n-th cyclotomic polynomial*, defined as the factor of $x^n - 1$ that does not appear in $x^m - 1$ for any $m < n$.

EXAMPLE XII.10.
(1) $x^1 - 1 = (x-1)$ and $\Phi_1(x) = x - 1$.
(2) $x^2 - 1 = (x-1)(x+1)$ and $\Phi_2(x) = x + 1$.
(3) $x^3 - 1 = (x-1)(x^2+x+1)$ and $\Phi_3(x) = x^2 + x + 1$.
(4) $x^4 - 1 = (x-1)(x+1)(x^2+1)$ and $\Phi_4(x) = x^2 + 1$.
(5) $x^5 - 1 = (x-1)(x^4+x^3+x^2+x+1)$ and $\Phi_5(x) = x^4 + x^3 + x^2 + x + 1$.
(6) $x^6 - 1 = (x-1)(x+1)(x^2+x+1)(x^2-x+1)$ and $\Phi_6(x) = x^2 - x + 1$.

If p is prime, then $x^p - 1 = (x-1)(x^{p-1} + x^{p-2} + \ldots + x + 1)$ and it turns out that the second factor is irreducible, and thus

$$\Phi_p(x) = x^{p-1} + x^{p-2} + \ldots + x + 1 \qquad \text{if } p \text{ is prime.}$$

To see this, note that $(x^p - 1)/(x-1)$ becomes under $x = y + 1$ the quotient $((y+1)^p - 1)/y$, and the binomial theorem says that $(y+1)^p - 1)/y = y^{p-1} + \binom{p}{1} y^{p-2} + \ldots + \binom{p}{k} y^{p-k-1} + \ldots + \binom{p}{p-1} y^0$. We will check in a minute that $\binom{p}{k}$ is always divisible by p for $0 < k < p$. That means that $(y+1)^p - 1)/y$ satisfies the conditions of the Eisenstein test, and must be irreducible. But then $(x^p - 1)/(x-1)$ is also irreducible because you get one from the other by a linear substitution (that can be done backwards).

LEMMA XII.11. *For natural $0 < k < p$, $c_{p,k} := \binom{p}{k}$ is a multiple of p.*

PROOF. You learned in discrete math that number $c_{p,k}$ is the number of of ways to pick k things from p given ones, and that there is an explicit formula $c_{p,k} = \frac{p!}{k!(p-k)!}$. In particular, the numerator is a multiple of p. It then suffices to show that the denominator is not a multiple of p (since p is prime!). To see this, note that k and $p-k$ are both less than k, and so neither $k!$ nor $(p-k)!$ has a factor divisible by p. Since p is prime, and p divides neither factor, it also does not divide the product. □

2. Making new fields form old

Recall that a field is a ring in which every nonzeo element has an invers. Examples include $\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/p\mathbb{Z}$ for p prime, and more fancy ones like $\mathbb{R}(x)$, the ring of all rational functions in x with real coefficients. (A function is rational if it is the quotient of two polynomials.

We have seen that if $\mathbb{F}$ is a field then $\mathbb{F}[x]$ is a Euclidean domain, the Euclidean measure being degree of the polynomial.

DEFINITION XII.12. If $\mathbb{F}$ is a field and $f(x) \in \mathbb{F}[x]$ is an irredcible polynomial, then the quotient ring $\mathbb{F}[x]/\langle f\rangle$ is the *Kronecker extension* of $\mathbb{F}$ by f. We denote it $\mathrm{Kron}(f, \mathbb{F})$.

LEMMA XII.13. *If $f \in \mathbb{F}[x]$ is irreducible,* $\mathrm{Kron}(f, \mathbb{F})$ *is actually a field.*

PROOF. The point is that we need to show that every nonzero element of $\mathrm{Kron}(f, \mathbb{F})$ has an inverse. Recall that $\mathbb{F}[x]$ is Euclidean. Take $g \in \mathbb{F}[x]$. If the coset of g in $\mathrm{Kron}(f, \mathbb{F})$ is nonzero, g is not a multiple of f. If in addition f is irreducible, then $\gcd(f, g) = 1$. So in this case, we can write $af + bg = 1$ for some $a, b \in \mathbb{F}[x]$. This equation in $\mathrm{Kron}(f, \mathbb{F})$ means that b is the inverse of g in $\mathrm{Kron}(f, \mathbb{F})$. □

EXAMPLE XII.14. If $\mathbb{F} = \mathbb{R}$ and $f(x) = x^2 + 1$ then $\mathbb{F}[x]/\langle f\rangle$ is as real vector space spanned by the class of the constant polynomial 1 and the class of the polynomial x in $\mathbb{F}[x]/\langle f\rangle$. Since $x^2 + 1 = 0$ in this quotient, we have $x^2 = -1$. So, we can identify the coset $a + bx$ in $\mathbb{F}[x]/\langle f\rangle$ with the complex number $a + b\sqrt{-1}$ and this identification preserves addition and multiplication. It is thus a ring isomorphism.

Another way to make new fields is as follows.

DEFINITION XII.15. Let $\mathbb{F} \subseteq \mathbb{E}$ be fields, and pick $\beta \in \mathbb{E}$. Then $\mathbb{F}(\beta)$ is defined to be the smallest field that contains $\mathbb{F}$ and β. It is also the intersection of all fields that contain $\mathbb{F}$ and β. Clearly, $\mathbb{F}(\beta)$ is inside $\mathbb{E}$.

THEOREM XII.16 (Kronecker Extension Theorem). *Let $f \in \mathbb{F}[x]$ be an irreducible polynomial.*

The Kronecker extension $\mathrm{Kron}(\mathbb{F}, f) = \mathbb{F}[x]/\langle f\rangle$ *is always a field (provided that f is irreducible). If viewed as a vector space over $\mathbb{Q}$, its dimension is the degree of f.*

If any extension field $\mathbb{E} \supseteq \mathbb{F}$ contains a root β of $f(x)$ then the smallest field $\mathbb{F}(\beta)$ inside $\mathbb{E}$ that contains both $\mathbb{F}$ and β is isomorphic to $\mathrm{Kron}(\mathbb{F}, f)$.

PROOF. Write R for $\mathrm{Kron}(\mathbb{F}, f)$. For the first part we need to show that this ring is a domain, and that every nonzero element has an inverse.

Since $\mathbb{F}$ is a filed, $\mathbb{F}[x]$ is a Euclidean ring. In particular, it is a UFD and so "prime" is the same as "irreducible". So f is a prime element in $\mathbb{F}[x]$. Now suppose

$g, h \in \mathbb{F}[x]$ are such that $\overline{g}$ and $\overline{h}$ in R are zerodivisors in R. That means that $gh \in \langle f \rangle$, so that $gh = \alpha f$ for some $\alpha \in \mathbb{F}[x]$. But then gh is divided by f and since f is prime, f muts divide one of them, say $f|g$. But then $\overline{g} = \overline{0}$ in R, so R has no zerodivisors and is a domain.

Now take $g \in \mathbb{F}[x]$ with $\overline{g}$ nonzero in R and look for an inverse. Since $\overline{g}$ is nozero, f can't divide g. Since $\mathbb{F}[x]$ is a Euclidean ring, the gcd of f, g is a linear combination of f, g. But this gcd is 1 since f is irreducble. Thus there are polynomails $a(x), b(x)$ with $f(x)a(x) + g(x)b(x) = 1$. Read this module $\langle f \rangle$ to get $\overline{g} \cdot \overline{b} = \overline{1}$. So $\overline{b}$ is an inverse to $\overline{g}$.

From the definition, it is clear that $\mathrm{Kron}(\mathbb{F}, f)$ has a basis given by $1, x, \ldots, x^{\deg f - 1}$.

Now we prove the last claim. Let us make a ring morphism from $\mathbb{F}[x]$ to $\mathbb{F}(\beta)$ by sending x to β and any element of $\mathbb{F}$ to itself. The kernel is the polynomials $p(x)$ for which $p(\beta) = 0$. These are the multiples of the minimal polynomial of β. This minimal polynomial is a divisor of $f(x)$ and since $f(x)$ is irreducible, this minimal polynomial is $f(s)$ itself.

It follows that we can actually make a ring morphism from $\mathrm{Kron}(\mathbb{F}, f)$ to $\mathbb{F}(\beta)$ by sending the coset of x to β and all elements of $\mathbb{F}$ to themselves.

As we know, $\mathrm{Kron}(\mathbb{F}, f)$ is a field, and so is by definition $\mathbb{F}(\beta)$. So we have one field contained in another, which makes the bigger field a vector space over the smaller one. The bigger one is generated over $\mathbb{F}$ by $1, \beta, \beta^2 \ldots, \beta^{\deg f - 1}$, and each of these is in the image of the morphism, $x^i \mapsto \beta^i$. So, the morphism is a surjective inclusion, hence an isomorphism. □

Example XII.17. Let $\mathbb{F} = \mathbb{Z}/3\mathbb{Z}$ and choose $f(x) = x^2 + x - 1$. We study the Kronecker extension $R = \mathbb{F}[x]/\langle f \rangle$.

Denote α the coset of x in R. Then $\alpha^2 + \alpha - \overline{1} = 0$ in R. So, powers of α higher then the first power can be replaced by lower powers. So, R is a $\mathbb{Z}/3\mathbb{Z}$-vector space spanned by $\overline{1}$ and α. So the element of R are

$$0, \overline{1}, \overline{2}, \alpha, \overline{1} + \alpha, \overline{2} + \alpha, 2\alpha, \overline{1} + 2\alpha, \overline{2} + 2\alpha.$$

Moreover, $\alpha^2 + \alpha - \overline{1} = 0$ in R means that the polynomial $y^2 + y - 1$ has a root in R. (This is built into the construction for any Kronecker extension). One could ask "what is the other root?". Let's find it. We do long division of $y^2 + y - 1$ by $(y - \alpha)$. The answer is: $y^2 + y - 1 = (y - \alpha)(y + \alpha + 1)$. (You might want to check this: $(y - \alpha)(y + \alpha + 1) = y^2 + y(-\alpha + \alpha + 1) + (-\alpha)(\alpha + 1)$. The linear term is fine, and for the constant term observe that is equals $-\alpha^2 - \alpha$. But as $\alpha^2 + \alpha - \overline{1} = 0$, this constant term is -1.

So, if you make a Kronecker extension, a previously irreducible polynomial will have at least one root. This says that you can (somehat artificially) make bigger fields in which your favorite polynomial splits completely into linear terms. That is a topic of a future lecture. For now, more examples.

Example XII.18. Let $\mathbb{F} = \mathbb{Q}$ and choose $f(x) = x^3 - 2$. Note that we need no fancy theorems to let us know that this is an irreducible polynomial: the cubic root of 2 is not a rational number (you can carry out the same proof as for irrationality of the square root of 2) and so $f(x)$ has no linear factor. But it's cubic, so if it factors at all it must have a linear factor.

As $f(x)$ is irreducible, $R = \mathbb{F}[x]/\langle f \rangle$ is a field. Denote again the coset of x in R by α. Then in R we have a linear dependence $1 \cdot \alpha^3 + 0 \cdot \alpha^2 + 0 \cdot \alpha^1 - 2 \cdot \alpha^0 = 0$,

which allows to rewrite all powers of α in terms of just second, first, and zeroth powers. So, R is as a vector space over $\mathbb{Q}$ spanned by $\overline{1}, \alpha, \alpha^2$.

Supposedly, f has a root in R called α. Let;s find the other factor: divide y^3-2 by $y-\alpha$. We find a quotient of $y^2+\alpha y+\alpha^2$.

It is entirely reasonable to ask whether this quadric splits further. In other words, does f have a further root in R? To find out, pick an element of R; it will look like $a\alpha^2+b\alpha+c$ with rational numbers a, b, c. Now take this and plug it for y into $y^2+\alpha y+\alpha^2$. After sifting through the mess, you find that you obtained (using that $\alpha^3=2$ of course)

$$\alpha^2(b^2+2ac+b+1)+\alpha(2a^2+2bc+c)+\overline{1}(c^2+4ab+2a).$$

We are asking whether this can be zero for suitable choices of $a, b, c \in \mathbb{Q}$. This is here not totally easy, and in general (other Kronecker extensions) can be extremely hard.

Here we can argue as follows: if the displayed expression is zero, then the three expressions $b^2+2ac+b+1, 2a^2+2bc+c = 2a^2+c(2b+1), c^2+4ab+2a = c^2+2a(2b+1)$ are all zero. But then $2a^2=-c(2b+1)$ and $c^2=-2a(2b+1)$ gives $c/2a = 2a^2/c^2$. So $(2a/c)^3=2$ and as weknow there are no rational numbers a, c suchthat $(2a/c)^3=2$. It follows that the quadric does not have any further roots in R.

CHAPTER XIII

Splitting fields and extension towers

DEFINITION XIII.1. If $f \in \mathbb{F}[x]$ is a polynomial over $\mathbb{F}$ then a *splitting field* for f is any field extension $\mathbb{E} \supseteq \mathbb{F}$ such that f splits a product of linear polynomials over $\mathbb{E}$.

We let $\mathrm{Split}(f)$ stand for any splitting field of f that is minimal with this respect.

Let us prove that such things exist:

THEOREM XIII.2. *For any field $\mathbb{F}$ and any $f \in \mathbb{F}[x]$, splitting fields exist.*

PROOF. If f factors over $\mathbb{F}$, factor inasmuch as you can. Then you are left with proving that you can sp-lit any irreducible polynomial. So assume from the start that f is irreducible.

Let $\mathbb{E}$ be the Kronecker extension $\mathrm{Kron}(\mathbb{F}, f)$. Then we know that f has a root in $\mathbb{E}$, namely the coset β of x. So you can split off a linear factor from f. Do that, split what is left as far as you can in $\mathbb{E}$ and repeat the argument. Since degree goes down at each step, eventually you will arrive at an extension in which f splits completely. □

Note that this also proves that $\mathrm{Split}(f)$ exists. However, there are choices being made in the process and it is not clear right away to what extent these choices have an impact on the final result. It is a fact that no matter how you construct $\mathrm{Split}(f)$, each version is isomorphic to any other, and that any such isomorphism can be arranged to identify the copy of $\mathbb{F}$ each splitting field contains. One says that the various versions of $\mathrm{Split}(f)$ are *isomorphic over* $\mathbb{F}$.

Note also that the worst case scenario is that we need to make Kronecker extensions to polynomials of degrees $\deg(f), \deg(f) - 1, \ldots, 3, 2$.

EXAMPLE XIII.3. If $f(x) = x^2 - 2$ with $\mathbb{F} = \mathbb{Q}$, then $\mathrm{Split}(f) = \mathrm{Kron}(\mathbb{Q}, f)$. Indeed, $x^2 - 2$ does not split over $\mathbb{Q}$, so $\mathrm{Split}(f)$ is not $\mathbb{Q}$. On the other hand, $\mathrm{Kron}(\mathbb{Q}, f)$ contains one root β of f, and dividing $x^2 - 2$ by $x - \beta$ leaves a linear polynomial. So, f splits over $\mathrm{Kron}(\mathbb{Q}, f)$. It follows that we can find a copy of $\mathrm{Split}(f)$ inside $\mathrm{Kron}(\mathbb{Q}, f)$.

On the other hand, as a vector space over $\mathbb{Q}$, $\mathrm{Kron}(\mathbb{Q}, f)$ is 2-dimensional, with basis $\{1, \beta\}$. So $\mathrm{Split}(f)$, another vector space over $\mathbb{Q}$, is wedged between the one-dimensional $\mathbb{Q}$-vector space $\mathbb{Q}$ and the 2-dimensional $\mathbb{Q}$-vector space $\mathrm{Kron}(\mathbb{Q}, f)$. Since we know that $\mathrm{Split}(f)$ can't be $\mathbb{Q}$, it must be $\mathrm{Kron}(\mathbb{Q}, f)$.

This argument works of course for any base field and any irreducible quadric. Indeed, if β is a root of $x^2 + ax + b$ then $x^2 + ax + b = (x - \beta)(x + a + \beta)$ in $\mathrm{Kron}(\mathbb{F}, f)[x]$.

EXAMPLE XIII.4. As we have seen, the quadratic factor $x^2 + \beta x + \beta^2 = (x^3 - 2)/(x - \beta)$ remains irreducible over $\mathrm{Kron}(\mathbb{Q}, x^3 - 2)$. So in order to make a splitting

field for $x^3 - 2$ we need first β and then additionally a Kronecker extension that catches a root of $x^2 + \beta x + \beta^2$.

It is time for the following concept.

DEFINITION XIII.5. If $\mathbb{F} \subseteq \mathbb{E}$ is an extension of fields, $\mathbb{E}$ is a vector space over $\mathbb{F}$. We denote the vector space dimension of $\mathbb{E}$ over $\mathbb{F}$ by $[\mathbb{E} : \mathbb{F}]$ and call it the *degree of the extension*.

Clearly, the degree of a Kronecker extension $\mathrm{Kron}(\mathbb{F}, f)$ is the degree of the polynomial f, since $\mathrm{Kron}(\mathbb{F}, f)$ has the basis $1, x, x^2, \ldots, x^{\deg f - 1}$.

EXAMPLE XIII.6. $\mathrm{Kron}(\mathbb{Q}, x^3 - 2)$ is degree 3 over $\mathbb{Q}$; $\mathrm{Split}(\mathbb{Q}, x^3 - 2)$ is degree 6 over $\mathbb{Q}$; $\mathrm{Split}(\mathbb{F}, f)$ is of degree at most $(\deg(f))!$ over $\mathbb{F}$.

EXAMPLE XIII.7. It is definitely possible for a cubic polynomial to split in a degree 3 extension (within the first Kronecker extension of the iterative splitting process).

Let $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$ and choose $f = x^3 + x + 1$. Since f has no roots in $\mathbb{F}$ (check that!), it has no linear factors over $\mathbb{Z}/2\mathbb{Z}$. Since it is degree 3, it has no factors and is thus irreducible.

Let β be the Kronecker root for f in $\mathbb{K} := \mathrm{Kron}(\mathbb{F}, f)$. Then β^2 and $\beta^2 + \beta$ are also roots of f inside $\mathbb{K}$. This can be seen by stupidly plugging in: $(\beta^2)^3 + (\beta^2)^1 + 1 = \beta^6 + \beta^2 + 1 = (\beta^3 + \beta + 1)^2$ since we are in characteristic 2. But $\beta^3 + \beta + 1 = 0$.

Similarly, $(\beta^2 + \beta)^3 + (\beta^2 + \beta)^1 + 1 = \beta^6 + 3\beta^5 + 3\beta^4 + \beta^3 + \beta^2 + \beta + 1 = (\beta^6 + \beta^2 + 1) + (\beta^5 + \beta^3 + \beta^2) + (\beta^4 + \beta^2 + \beta^1)$ (remember that 2=0 here!). Each bracket is zero since it is a multiple of $\beta^3 + \beta + 1$.

You might wonder how I knew these 2 other roots in the example. One comes as follows.

LEMMA XIII.8. *If β is a root of $f(x)$ and the field has charactersitic p, then β^p is also a root.*

PROOF. In charactersitic p, freshman's dream for p-th powers holds: $f(x)^p = f(x^p)$. Plug in $x = \beta$. □

On the third root in the example: if a cubic $f(x)$ has 2 roots r_1, r_2 in some field, the third root is also in the field, and you can find it by longly dividing $f(x)$ first by $x - r_1$ and then what you got by $x - r_2$. You'll be left with $x - r_3$, and that is what I did.

1. Roots with multiplicity

Some polynomials, like $x^2 + 2x - 3 = (x + 3)(x - 1)$, have all their roots distinct. For polynomials $x^2 + px + q$ of degree 2, we know that this happens exactly when the *discriminant* $p^2 - 4q$ is nonzero. (I am assuming here that we can use the quadratic formula, which necessitates that $2 \neq 0$ in the ring). So for example, when $p = 6, q = 9$ we find that $x^2 + px + q = x^2 + 6x + 9 = (x + 3^2$ and one is prompted to say that -3 is a root of $x^2 + 6x + 9$ of multiplicity two.

For polynomials of higher degree, there is a similar discriminant test; the problem is that the formula for the discriminant gets impossibly difficult to remember. For example, for a cubic $x^3 + px^2 + qx + r$, the discriminant is $8pqr 4p^3 r + q^2 r^2 4r^3 27r^2$.

It is kind of clear that if a polynomial has a factor of the sort $(x-r)^k$, then r is a root and of multiplicity at least k. So, before one investigates multiplicity, one should perhaps split the polynomial as far as one can into irreducibles.

As one sees by examples, it is ofetn interesting to take a polynomial over one ring R and ask for its roots in a bigger ring. For eample, we know that we need to look inside $\mathbb{C}$ for roots of x^2+1.

Some strange things can happen.

EXAMPLE XIII.9. Let $\mathbb{K}=\mathbb{Z}/p\mathbb{Z}$. Then x^p-1 (which over the complex numbers has the p different roots of 1 as solutions) is equal to $(x-1)^p$ (because of freshman's dream in characteristic p). So, it has only one root, $x=1$, and that with multiplicity p.

Stranger yet, there are polynomials that are irreducible and yet have multiple roots in a suitable larger ring.

EXAMPLE XIII.10. Let $\mathbb{K}=\mathbb{Z}/p\mathbb{Z}(t)$. So, $p=0$ in our ring, t is a variable, and we are looking at the rational functions in t with coefficients in $\mathbb{Z}/p\mathbb{Z}$. (Recall that "rational function" means "quotient of two polynomials).

Now look at the polynomial x^p-t. In $\mathbb{K}$, this has no roots because the p-th root of a variable is not expressible as a quotient of 2 polynomials in that variable. We will show in a bit, that x^p-t is also irreducible. Let's make the field bigger, say $\tilde{\mathbb{K}}=\mathbb{Z}/p\mathbb{Z}(\sqrt[p]{t})$ the rational functions with $\mathbb{Z}/p\mathbb{Z}$ coefficients in the symbol p-th root of t.

If $p=2$, we have $(x-\sqrt[2]{t})(x-\sqrt[2]{t})=x^2-2x\sqrt[2]{t}+t=x^2+t=x^2-t$ since $2=0$.

For $p=3$ we have $(x-\sqrt[3]{t})^3=x^3-3x^3\sqrt[3]{t}+3x\sqrt[3]{t}^2-t=x^3-t$ since $3=0$.

In general, $(x-\sqrt[p]{t})^p=x^p+p(\text{stuff})-t$ where the middle part is the stuff that comes from the binomial theorem for $i=1.\ldots,p-1$. In all cases then, $x^p-t=(x-\sqrt[p]{t})^p$ has a p-fold root in $\tilde{\mathbb{K}}$ while it was irreducible over $\mathbb{K}$.

Here is a way for testing whether a polynomial can ever have multiple roots. The prime in the theorem denotes taking the derivative according to the rules of calculus: product rule and power rule. (You might ask "What other rules might I possible want to use for a derivativ, isn't that a stupid thing to say?". You are sort of right. There are no other rules one should ever use. But the fact is that in some environments, calculus seems like a dubious activity to engage in. For example, in $\mathbb{Z}/3\mathbb{Z}[x]$, what could "differentiation" mean? Normally, a derivative is a limit, but in $\mathbb{Z}/p\mathbb{Z}$ there are only finitely many "numbers", so limits are very limited in their nature...)

THEOREM XIII.11. *Let $\mathbb{F}$ be a field and $f(x)\in\mathbb{F}[x]$. Then $f(x)$ has a double root in some (perhaps mysterious) extension field $\mathbb{E}\supseteq\mathbb{F}$ if and only if* $\gcd(f,f')$ *is not 1.*

In other words, if f,f' are coprime then f has single roots in any field.

PROOF. If in some extension field $\mathbb{E}$ we have $(x-r)^2|f(x)$ (so $r\in\mathbb{E}$ is a multiple root) write $f(x)=(x-r)^2\cdot g(x)$. Then taking derivatives, we have $f'(x)=2(x-r)g(x)+(x-r)^2g(x)=(x-r)[2g(x)+(x-r)\cdot g'(x)]$ is a multiple of $x-r$. Of course, so is f itself, and so $x-r$ divides both f,f' and hence must divide their gcd. This means that a multiple root in an extension field prevents the gcd of f,f' being 1.

Now suppose the gcd of f, f' is not 1, or in other words, some $g(x)$ of positive degree divides both f and f'. Then let $\mathbb{E}$ be the Kronecker extension on $\mathbb{F}$ for any irreducible factor of $g(x)$. In $\mathbb{E}$, $g(x)$ has the Kronecker root β, and so $g(x)$ is a multiple of $x - \beta$ and also $f(x)$ is a multiple of $x - \beta$. So we can write $f(x) = h(x)(x - \beta)$. Then the derivative of f is $f'(x) = h'(x)(x - \beta) + h(x)$. Now plug in $x \mapsto \beta$. We know that $(x-\beta)|g(x)|f'(x)$, so $f'(\beta) = 0$. But then $h(\beta)$ must also be zero. That says that $(x - \beta)$ divides $h(x)$, and so $f(x) = (x - \beta)h(x)$ has $x - \beta$ twice as factor. So β is a double root of f in $\mathbb{E}$. □

Remark XIII.12. In characteristic zero (when $\mathbb{K}$ contains $\mathbb{Q}$) an irreducible polynomial is relatively prime to its own derivative, because the derivative is a *nonzero* polynomial of lower degree, and so cannot have a common divisor with the irreducible f.

In prime characteristic, the derivative $f'(x)$ can be zero without f being a constant. For example, the polynomial $x^p - t$ from the example above has derivative zero, since $(x^p)' = px^{p-1}$ and $p = 0$. (Note that we take x-derivatives, so $(t)' = 0$ as t and x do not relate in that example!) In that case, then, we have $\gcd(fmf') = \gcd(f, 0) = f$.

Definition XIII.13. A polynomial $f(x)$ with coefficients in the field $\mathbb{F}$ is *separable* if f does not have multiple roots in any extension field of $\mathbb{F}$. Any other polynomial is *inseparable*.

The choice of "separable" indicates that separable polynomials have their roots "separated out" in any extension: the roots never equal one another. In characteristic zero, "irreducible" implies "separable". But in characteristic p, separability is an actual condition. It is a fact that over a finite field, "irreducible" still implies "separable", but in infinite fields of characteristic p one needs to be careful.

CHAPTER XIV

Week 14: Minimal polynomials and finite fields

1. Minimal Polynomials

Recall that a field extension $\mathbb{F} \subseteq \mathbb{E}$ makes $\mathbb{E}$ a vector space over $\mathbb{F}$. (Think of $\mathbb{R} \subseteq \mathbb{C}$). The start of our investigations is based on

DEFINITION XIV.1. If $\mathbb{F} \subseteq \mathbb{E}$ is a field extension it is called *algebraic* if for any $\alpha \in \mathbb{E}$ the powers $1, \alpha, \alpha^2, \ldots$ of α are linearly dependent over $\mathbb{F}$.

There are many field extensions that are not algebraic. For example, $\mathbb{Q} \subseteq \mathbb{R}$ is not finite, because the powers $1, \pi, \pi^2, \ldots$ of π have no $\mathbb{Q}$-linear dependence. Another way is to say that π does not occur as a root of a polynomial in $\mathbb{Q}[x]$, π is *transcendental*. Finiteness of a field extension indicates that the two fields are not too far from one another, in a sense to be discussed this week and next.

For now we note the obvious

THEOREM XIV.2. *If $\mathbb{F} \subseteq \mathbb{E}$ is algebraic, then for any $\alpha \in \mathbb{E}$ there is a monic irreducible polynomial $f(x) \in \mathbb{F}[x]$ such that $f(\alpha) = 0$.*

Note that previously we started with a polynomial and looked for roots; this is now the other way round.

PROOF. If the powers of α are dependent, there is an expression $\sum_{i=0}^{k} r_i \alpha^i$ that equates to zero. The polynomial $\sum_{i=0}^{k} r_i x^i$ is what we are looking for. It can be made monic (have lead coefficient 1) by dividing out the actual lead coefficient. Note that this division does not affect the vanishing of the polynomial at $x = \alpha$. $\square$

If we have several polynomials that vanish for $x = \alpha$, then their gcd has the same property. So, the gcd of all polynomials that vanish at $x = \alpha$ is one such as well, and clearly the one of lowest degree. (Note: if I is the ideal of all polynomials vanishing at $x = r$ then the generator for this ideal—principal since $\mathbb{F}[x]$ is Euclidean—is the one we want.)

DEFINITION XIV.3. If $\mathbb{F} \subseteq \mathbb{E}$ is a field extension, and if the powers of $\alpha \in \mathbb{F}$ are linearly dependent over $\mathbb{F}$ then the monic polynomial $f(x) \in \mathbb{F}[x]$ of minimal degree with $f(\alpha) = 0$ is the *mimimal polynomial of α over $\mathbb{F}$* and denoted $\mathrm{minpol}_{\mathbb{F}}(\alpha)$.

Note that one really needs to know $\mathbb{F}$ in this definition: $\mathrm{minpol}_{\mathbb{R}}(\sqrt{-1}) = x^2+1$, but $\mathrm{minpol}_{\mathbb{C}}(\sqrt{-1}) = x - \sqrt{-1}$. We note for future purposes that the second (complex) minimum polynomial divides the first (real).

DEFINITION XIV.4. If $\mathbb{F} \subseteq \mathbb{E}$ is a field extension and if the powers of $\alpha \in \mathbb{E}$ are linearly dependent over $\mathbb{F}$ then α *is algebraic over* $\mathbb{F}$. Elsewise we call α *transcendental over* $\mathbb{F}$.

DEFINITION XIV.5. If $\mathbb{F} \subseteq \mathbb{E}$ is a field extension, it is called *finite* if $\mathbb{E}$ is a finite-dimensional vector space over $\mathbb{F}$.

LEMMA XIV.6. *A field extension $\mathbb{F} \subseteq \mathbb{E}$ that is finite is also algebraic.*

PROOF. If $\mathbb{E}$ is a finite-dimensional vector space, then any infinite collection of elements in $\mathbb{E}$ must be linearly dependent, since the size of any linearly independent set is a lower bound for the size of any basis (and the size of a basis is the dimension of the vector space). In particular, the infinitely many powers of α must be dependent. □

It is time to introduce some notation regarding field and ring extensions.

DEFINITION XIV.7. If $R \subseteq S$ are ring and α some element of S then $R[\alpha]$ denotes the smallest ring that contains α and all of R.

If $\mathbb{F} \subseteq \mathbb{K}$ are fields, and $\alpha \in \mathbb{K}$, then $\mathbb{F}(\alpha)$ is the smallest field that contains $\mathbb{F}$ and α. (This may be considerably larger than $\mathbb{F}[\alpha]$ since it also must contain all inverses to the elements of $\mathbb{F}[\alpha]$).

Many, but not all, algebraic extensions are finite. For example, if $\mathbb{F}$ is any field and α is a root to any polynomial $f(x) \in \mathbb{F}[x]$ then $\mathbb{F}(\alpha)$ (the smallest field that contains $\mathbb{F}$ and α) is finite over $\mathbb{F}$. This is simply because $\mathbb{F}(\alpha)$ is the Kronecker extension $\mathrm{Kron}(\alpha, \mathbb{F}) = \mathbb{F}[x]/\langle f \rangle$ which is a vector space of dimension $\deg(f)$ over $\mathbb{F}$m spanned by $1, x, \ldots, x^{\deg(f)-1}$.

In fact, since $\deg(f)$ is finite, so is its factorial, and it follows that $\mathrm{Split}(f, \mathbb{F})$ is a finite extension of $\mathbb{F}$.

On the other hand, it is not true that any algebraic extension is also finite. For example, the field that you get when you start with $\mathbb{Q}$ and then throw in all n-th roots of 2 ($n = 2, 3, \ldots$) is algebraic but not finite. That is is not finite is kind of believable since whatever finite basis this field might have over $\mathbb{Q}$, this basis can't involve all roots of 2. It is more difficult to believe that this extension is algebraic, because while of course the n-th root of 2 fits the equation $x^n = 2$ (and thus is algebraic over $\mathbb{Q}$), it is far less clear that unpleasantries such as

$$\frac{33\ \sqrt[46]{2} + 112\ \sqrt[17]{2} - 641\ \sqrt[666]{2}}{6\sqrt[4]{2} - 3352295\ \sqrt[532]{2}}$$

fit into a polynomial with rational coefficients. It turns out that they indeed do, and the reason is the following.

Recall that we defined the vector space dimension of $\mathbb{E}$ over $\mathbb{F}$ as the degree of the extension, and wrote $[\mathbb{E} : \mathbb{F}]$. If one iterates extensions, $\mathbb{F} \subseteq \mathbb{F}' \subseteq \mathbb{F}$", we have a formula

$$[\mathbb{F}'' : \mathbb{F}] = [\mathbb{F}'' : \mathbb{F}'] \cdot [\mathbb{F}' : \mathbb{F}].$$

This is kind of clear from linear algebra: if $\mathbb{F}''$ looks like $(\mathbb{F}')^r$ and $\mathbb{F}'$ looks like $\mathbb{F}^s$ then $\mathbb{F}''$ looks like $(\mathbb{F}^s)^r = \mathbb{F}^{rs}$.

This formula implies that any extension of $\mathbb{Q}$ of the form $\mathbb{Q}[\sqrt[2]{2}, \sqrt[3]{2}, \ldots, \sqrt[k]{2}]$ is still finite over $\mathbb{Q}$. So the powers of any element in this extension are still algebraic over $\mathbb{F}$, even if it is often very difficult to find a polynomial that they fit into. In particular, the monster in the display above has some minimal polynomial. (My guess is that it has degree equal to a number of about 40 digits).

EXAMPLE XIV.8. If $\mathbb{F} = \mathbb{Q}$ and $f = x^3 - 2$, then we can build splitting fields one step at a time. $\mathbb{F}' := \mathrm{Kron}(\mathbb{F}, f) = \mathbb{Q}[x]/\langle f \rangle$ contains at least the Kronecker

root β. So in $\mathbb{F}'[x]$ we can factor $x^3-2=(x-\beta)(x^2+x\beta+\beta^2)$. This first extension is degree 3, because we adjoined the root of a cubic.

If we picture β as the real third root of 2, it is clear that $\mathbb{F}'$ is still inside the field of real numbers, and in particular can't contain all three third roots of 2 because the other two are complex and not real.

Thus, $x^2+x\beta+\beta^2$ has no roots in $\mathbb{F}'$ and hence does not factor in $\mathbb{F}'[x]$. A second Kronecker extension $F'' := \mathrm{Kron}(\mathbb{F}', x^2+x\beta+\beta^2)$ is a degree two extension of $\mathbb{F}'$ and therefore a degree $2\cdot 3=6$ extension of $\mathbb{F}$. In that field, f splits completely. Hence $\mathbb{F}'' = \mathrm{Split}(\mathbb{F}, f)$.

2. Finite Fields

EXAMPLE XIV.9. Let $\mathbb{F}=\mathbb{Z}/2\mathbb{Z}$ and take $f(x)=x^3+x+1$, $g(x)=x^3+x^2+1$. It is easy to check that neither f nor g have a root in $\mathbb{F}$, and so (as cubics) are irreducible.

$\mathrm{Kron}(\mathbb{F}, f)=\mathbb{F}[x]/\langle f\rangle$ has $8=2^3$ elements $\{\overline{0}, \overline{1}, \overline{x}, \overline{x}+\overline{1}, \overline{x}^2, \overline{x}^2+\overline{1}, \overline{x}^2+\overline{x}, \overline{x}^2+\overline{x}+\overline{1}\}$. The same is true for $\mathrm{Kron}(\mathbb{F}, g)=\mathbb{F}[y]\langle g\rangle$, but we must not confuse elements in the different extensions because in the first, we go modulo f and in the other we go modulo g. I intentionally write different variables x, y here.

Let α be a the Kroenecker root for f, so $\alpha = x \bmod \langle f\rangle$. Write β for the Kronecker root of g, so $\beta = x \bmod \langle g\rangle$.

So α is a root of f, who else? $f(x):(x-\alpha)=x^2+\alpha x+(\alpha^2+1)$, which we call $f_2(x)$. Then if you plug α^2 into $f_2(x)$, you get zero, so α^2 is a root of f_2 and also then of $f(x)$. The remaining root can be found as $\alpha^2+\alpha$. (As a test, if you multiply out $(x-\alpha)(x-\alpha^2)(x-\alpha^2-\alpha)$ you get $f(x)$ back, using that $f(\alpha)=0$.)

So $\mathrm{Kron}(\mathbb{F}, f)$ is actually the splitting field of f over $\mathbb{F}$.

Now plug $x-1$ into $f(x)$. You get $(x-1)^3+(x-1)+1=x^3+x^2+1=g(x)$. So, g has roots equal to thos of f shifted up by 1. They are $\alpha+1, \alpha^2+1, \alpha^2+\alpha+1$.

In particular, $\mathrm{Kron}(\mathbb{F}, f)$ is also the splitting field for $g(x)$. So there is no real difference between $\alpha+1$ in $\mathrm{Kron}(\mathbb{F}, f)$ and $\beta\in\mathrm{Kron}(\mathbb{F}, g)$. There is only one degree 3 extension of $\mathbb{F}$.

REMARK XIV.10. Here is an amusing computation that explains the previous example. Any degree 3 extension of $\mathbb{Z}/2\mathbb{Z}$ will be a vector space of dimension 3 over $\mathbb{Z}/2\mathbb{Z}$. As such, it contains 2^3 elements, of which 7 are nonzero. Since in a field all nonzero elements have inverses, these 7 elements form a group with multiplication. So, all group elements have order dividing 7, by Lagrange. That translates to: "all nonzero elements satisfy $a^7=1$", and so all elements satisfy $a^8-a=0$. We can factor $x^8-x=x(x+1)(x^3+x+1)(x^3+x^2+1)$ and we find here the irreducibles f and g as factors of x^8-x.

So, any field of 8 elements contains all roots to f, and all roots to g. There is only one field with 8 elements.

THEOREM XIV.11. *Let $\mathbb{F}=\mathbb{Z}/p\mathbb{Z}$ be the field with p elements, p a prime number.*

Choose $e\in\mathbb{N}_{|geq1}$. Then there exists a field $\mathbb{GF}(p,e)$ with p^e elements. All elements $a\in\mathbb{GF}(p,e)$ are roots of the polynomial $x^{p^e}-x$, and $x^{p^e}-x$ completely splits over $\mathbb{GF}(p,e)$. In other words, $\mathbb{GF}(p,e)$ is the splitting field of the polynomial $x^{p^e}-x$.

Every element of $\mathbb{GF}(p,e)$ *is equal to its* p^e*-th power, and so every element has a* p^e*-th root in* $\mathbb{GF}(p,e)$.

One has $\mathbb{GF}(p,1) = \mathbb{Z}/p\mathbb{Z}$.

The degree of the field extension $[\mathbb{GF}(p,e) : \mathbb{GF}(p,1)]$ *is* e. *In consequence,* $\mathbb{GF}(p,e)$ *is the Kronecker extension* $\mathrm{Kron}(\mathbb{F}, g(x))$ *for every irreducible polynomial of degree* e.

SKETCH. Existence Take any splitting field $\mathbb{K}$ for $f(x) := x^{p^e} - x$ (for example, a suitable field inside an iteration of Kronecker extensions). Denote this *set* by $\mathbb{GF}(p,e)$.

Note that if a, b are both roots of $f(x)$ then that is also true for $a \pm b$ and ab and a/b provided that $b \neq 0$. (Why? For $\pm$ the binomial theorem gives you p-divisible coefficients in $(a \pm b)^{p^e}$ in all but the first and last term. For ab and a/b this is very easy.) It follows that $\mathbb{GF}(p,e)$ is closed under $+$ and $-$, and under multiplication and division. So, this set of roots is a *field* (This is really weird and only happens over finite fields. For example, the field $\mathbb{Q}[\sqrt{2}]$ has lots and lots of elements that are not roots of $x^2 - 2\ldots$)

Splitting By construction, f has all its roots in $\mathbb{GF}(p,e)$, so $\mathbb{GF}(p,e)$ is the smallest field over which f splits. It follows that $\mathbb{GF}(p,e)$ *is the splitting field.*

Size The gcd of $f(x)$ and $f'(x) = p^e x^{p^e - 1} - 1$ is 1, since $p^e = 0$ and so $f'(x) = -1$. So, f has no multiple roots in any extension, and in particular not in $\mathbb{GF}(p,e)$. So, $\mathbb{GF}(p,e)$ is full of single roots of $f(x)$ and so must have p^e elements (it is a plitting field!).

Uniqueness Any field with p^e elements has $a^{p^e} - a = 0$ by the argument of the remark above, so any field with p^e elements is the splitting field of f.

(1) Since $f(a) = 0$ for all $a \in \mathbb{GF}(p,e)$, each element agrees with its own p^e-th power.

(2) If $e = 1$, we want the splitting field over $\mathbb{Z}/p\mathbb{Z}$ of $x^p - x$. But Little Fermat says that $a^p = a$ for each $a \in \mathbb{Z}/p\mathbb{Z}$. So all roots of $x^p - x$ are in $\mathbb{Z}/p\mathbb{Z}$ and we need no extension.

(3) A vector space with p^e elements over a field of p elements has to have e basis vectors. Let g be an irreducible polynomial of degree e over $\mathbb{Z}/p\mathbb{Z}$. Its Kronecker extension is a field extension of order e, so has p^e elements, and so must be a splitting field for $x^{p^e} - x$. So $\mathrm{Kron}(\mathbb{F}, g) = \mathbb{GF}(p,e)$

□

COROLLARY XIV.12. *The nonzero elements* $U(p,e)$ *of* $\mathbb{GF}(p,e)$ *form an Abelian group with respect to multiplications. This group is cyclic.*

PROOF. The first sentence is clear since fields have commutative multiplication and every nonzero element in a field has an inverse.

As Abelian group, $U(p,e)$ can be writen as $\mathbb{Z}/a_1\mathbb{Z} \times \cdots \times \mathbb{Z}/a_k\mathbb{Z}$ with $a_1 | a_2 | \cdots | a_k$, by FTFGAG. If q is an element of this product group, it has order a_k. So the elements of $U(p,e)$ have their a_k-th power equal to the identity. That means, they are roots of $x^{a_k} - 1$. So elle elements of $\mathbb{GF}(p,e)$ are roots to $x^{a_k+1} - x$. But such a polynomial can have only $a_k + 1$ roots, and we know $\mathbb{GF}(p,e)$ is the set of these roots, p^e in number. So $p^e = a_k + 1$. So $a_k = p^e - 1$. But $a_1 \cdot a_2 \cdots a_k$ shouldbe $p^e - 1 = |U(p,e)|$, and that means that $k = 1$ and so $U(p,e)$ is cyclic. □

EXAMPLE XIV.13. Let $p = 2$ and take $f(x) = x^4 + x + 1$. Since $f(\overline{0}) = f(\overline{1}) = \overline{1} \in \mathbb{Z}/2\mathbb{Z}$, f has no linear factors.

If f were to factor, then, it should factor as the product of 2 quadrics. But over $\mathbb{Z}/2\mathbb{Z}$ there is (easy check!) only one irreducible quadric, x^2+x+1. And x^2+x+1 doesn ot divde $f(x)$. So $f(x)$ is irreducible and $\mathrm{Kron}(\mathbb{Z}/2\mathbb{Z}, f)$ has $2^4 = 16$ elements.

By the theorem, $\mathrm{Kron}(\mathbb{Z}/2\mathbb{Z}, f) = \mathbb{GF}(2,4)$.

Let α be the Kronecker root and compute explicitly: $(\alpha^2+\alpha+1)^2+(\alpha^2+\alpha+1)+1 = 0$ in $\mathrm{Kron}(\mathbb{Z}/2\mathbb{Z}, f)$. This says that $\alpha^2+\alpha+1$ is a root in $\mathbb{GF}(2,4)$ of the irreducible quadric x^2+x^1. In particular, $\mathbb{GF}(2,4)$ contains $\mathrm{Kron}(\mathbb{Z}/2\mathbb{Z}, x^2+x+1) = \mathbb{GF}(2,2)$.

Example XIV.14. Let $p=2$ and take $f(x) = x^3+x+1$, an irreducible cubic in $\mathbb{Z}/2\mathbb{Z}[x]$. Take $\mathbb{GF}(p,3)$ to be its splitting field $\mathrm{Kron}(\mathbb{Z}/2\mathbb{Z}, f)$. The 8 elements of $\mathbb{GF}(2,3)$ are precisely the roots of $x^{2^3} - x = x(x-1)(x^3+x+1)(x^2+x^2+1)$.

Let us look for a copy of $\mathbb{GF}(2,2)$ in here. Since elements of $\mathbb{GF}(2,2)$ are characterized by satisfying $x^{2^2} - x = 0$, we need the 4 roots to $x(x-1)(x^2+x+1)$ to be in $\mathbb{GF}(2,3)$. But $\gcd(x^2+x+1, x^8-x) = 1$, so any element that makes both of these polynomials to zero should also make the polynomial 1 to zero. That being preposterous, nobody can be in $\mathbb{GF}(2,3)$ and $\mathbb{GF}(2,2)$ simultanously.

It is natural to ask when finite fields sit inside one another.

Theorem XIV.15. *$\mathbb{GF}(p,e)$ sits inside $\mathbb{GF}(p',e')$ if and only if $p=p'$ and $e|e'$.*

Proof. In $\mathbb{GF}(p,e)$ we have that $1+\ldots+1$ (e copies) gives 0. In $\mathbb{GF}(p',e')$, this is so with p' copies. If $p \neq p'$ then $\gcd(p,p') = 1$ copies of 1 aso amount to zero. We conclude $p=p'$ is necessary.

Suppose $\mathbb{GF}(p,e)$ sits inside $\mathbb{GF}(p,e')$. Then $\mathbb{GF}(p,e')$ is a vector space over $\mathbb{GF}(p,e)$. Since one has p^e elements, and the other $p^{e'}$ elements, $p^{e'}$ should be a power of p^e, and that means that e' should be a multiple of e.

Now suppose $e|e'$ and look for $\mathbb{GF}(p,e)$ inside $\mathbb{GF}(p,e')$. The main step is to see that $e|e'$ gives you that $x^{p^e} - x$ divides $x^{p^{e'}} - x$. To see this, write $e' = de$ and calculate

$$x^{p^{e'}} - x = (x^e - x)(x^{(d-1)e} + x^{(d-2)e} + \ldots + x^{1e} + x^0.$$

But then the splitting field of $x^{p^{e'}} - x$ must contain the splitting field of $x^{p^e} - x$ as we wanted to show. □

It is a natural question to ask "if $\mathbb{GF}(p,e)$ is a subfield of $\mathbb{GF}(p,e')$, how do we best identify the smaller field? (So far we only know that it must be in there somewhere).

Corollary XIV.16. *If $e|e'$, the subfield $\mathbb{GF}(p,e)$ inside $\mathbb{GF}(p,e')$ consists of exactly those elements that satisfy $x^{p^e} = x$.*

One can obtain elements in $\mathbb{GF}(p,e)$ by raising any element of $\mathbb{GF}(p,e')$ to the power $(p^{e'}-1)/(p^e-1)$.

Proof. That $\mathbb{GF}(p,e)$ is inside $\mathbb{GF}(p,e')$ comes from the preceding theorem. Since eny element in any version of $\mathbb{GF}(p,e)$ is a root of $x^{p^e} - x$, selecting the ones that do this is the right strategy.

Since $U(p,e')$ is cyclic of order $p^{e'}-1$ and since $U(p,e)$ of size p^e-1 sits inside $U(p,e')$ it must be so that $U(p,e)$ is the $(p^{e'}-1)/(p^e-1)$-powers of the elements of $U(p,e')$. □

EXAMPLE XIV.17. Let's look at the finite fields inside the field of size 2^{24}. They are the fields of sizes $2^{12}, 2^8, 2^6, 2^4, 2^3, 2^2, 2^1$. The containment relations are

$$\mathbb{GF}(2,1) \subseteq \mathbb{GF}(2,2) \subseteq \mathbb{GF}(2,4) \subseteq \mathbb{GF}(2,8) \subseteq \mathbb{GF}(2,24),$$

$$\mathbb{GF}(2,1) \subseteq \mathbb{GF}(2,3) \subseteq \mathbb{GF}(2,6) \subseteq \mathbb{GF}(2,12) \subseteq \mathbb{GF}(2,24),$$

and additional containments $\mathbb{GF}(2,2) \subseteq \mathbb{GF}(2,6)$ and $\mathbb{GF}(2,4) \subseteq \mathbb{GF}(2,12)$.

CHAPTER XV

Galois

1. The Frobenius

In a ring of characteristic $p > 0$ (such as in $\mathbb{GF}(p, e)$ or indeed a ring containing $\mathbb{Z}/p\mathbb{Z}$), we have

$$(a + b)^p = a^p + p^p$$

since the intermediate terms arising from the binomial theorem all are multiples of p, hence zero. It follows that

$$\begin{array}{rccc} \text{Frob}\colon \mathbb{GF}(p, e) & \to & \mathbb{GF}(p, e), \\ \gamma & \mapsto & \gamma^p \end{array}$$

is a morphism of additive groups. Since clearly $1^p = 1$ and $(\gamma\gamma')^p = \gamma^p(\gamma')^p$, the Frobenius also respects the multiplicative structure. It is therefore a ring moorphism.

THEOREM XV.1. *The p-Frobenius (p-th power map) is a field isomorphism*

$$\text{Frob}\colon \mathbb{GF}(p, e) \to \mathbb{GF}(p, e)$$

for any e.

If $e'|e$ and $\mathbb{GF}(p, e')$ therefore sits inside $\mathbb{GF}(p, e)$ then the Frobenius sends elements of this subfield into the subfield.

The e-fold iteration of the Frobenius on $\mathbb{GF}(p, e)$ is the identity map. One can interpret this as

> *the group $\mathbb{Z}/e\mathbb{Z}$ acts on $\mathbb{GF}(p, e)$ by sending the coset of t to the t-fold iteration of* Frob.

The elements of $\mathbb{GF}(p, e)$ that are fixed by Frob *are exactly the elements of $\mathbb{GF}(p, 1) = \mathbb{Z}/p\mathbb{Z}$. More generally, the elements of $\mathbb{GF}(p, e)$ that are fixed under the k-fold iteration of the p-th power map are precisely the elements of $\mathbb{GF}(p, \gcd(e, k))$.*

Suppose $\alpha \in \mathbb{GF}(p, e)$ is the root of a polynomial $f(x)$ with coefficients in $\mathbb{Z}/p\mathbb{Z}$. Then α^p is a root of $f(x)$ as well. In fact, iterating the p-th power map will produce all other roots in $\mathbb{GF}(p, e)$ of $f(x)$. In other words,

> *The orbit of $\alpha \in \mathbb{GF}(p, e)$ under the action of $\mathbb{Z}/e\mathbb{Z}$ above is the set of all roots of the minimal polynomial of α over $\mathbb{Z}/p\mathbb{Z}$.*

If $k \in \mathbb{N}$ one can define an action of $\mathbb{Z}/e\mathbb{Z}$ via $\lambda(t, \alpha) = (\alpha^p)^t$. The orbits under this action are the roots of the minimal polynomial of α over $\mathbb{GF}(p, \gcd(e, k))$.

PROOF. Let α be in the kernel of Frob. Then $\alpha^p = 0$ and since a field has no zerodivisors, $\alpha = 0$. So Frob is injective. By the isomorphism theorem, the image of Frob is isomorphic to $\mathbb{GF}(p, e)$. But that means it has p^e elements, and hence fills out the target field. So, Frob is bijective and hence an isomorphism. It permutes the elements of $\mathbb{GF}(p, e)$.

All elements in $\mathbb{GF}(p,e)$ satisfy $\alpha^{p^e}=\alpha$. If $e'|e$, the elements of $\mathbb{GF}(p,e')$ inside $\mathbb{GF}(p,e)$ are characterized by being those elements for which $\alpha^{p^{e'}}=\alpha$ already. Take such α and raise it to the p-th power. Then note that $(\alpha^p)^{p^{e'}}=\alpha^{p\cdot p^{e'}}=(\alpha^{p^{e'}})^p=\alpha^p$. In other words, $\mathrm{Frob}(\alpha)$ belongs to $\mathbb{GF}(p,e')$ again. So the isomorphisms that the Frobenius induces on the various fields $\mathbb{GF}(p,-)$ are compatible with inclusions.

The e-fold iteration of Frob sends $\alpha\in\mathbb{GF}(p,e)$ to $\alpha^{p^e}=\alpha$, so it is the identity on $\mathbb{GF}(p,e)$. It follows that we can read Frob as a group actiion of $\mathbb{Z}/e\mathbb{Z}$ on the elements of $\mathbb{GF}(p,e)$ via $\lambda(t \bmod e\mathbb{Z},\alpha)\mapsto \mathrm{Frob}^t(\alpha)$, which is nothing but α^{p^t}.

Since Frob^e is the identity on $\mathbb{GF}(p,e)$, Frob^k acts the same way as $\mathrm{Frob}^{\gcd(e,k)}$. (You should make sure you believe this before going on. It can be seen via the Euclidean algorithm: $\mathrm{Frob}^k(\alpha)=\mathrm{Frob}^{k-e}(\alpha)$; now iterate). If $\alpha^p=\alpha$ then α is a root of x^p-x, and there are exactly p of those, the elements of $\mathbb{Z}/p\mathbb{Z}=\mathbb{GF}(p,1)$. If $\mathrm{Frob}^k(\alpha)=\alpha$ then α is a root of $x^{p^{\gcd(e,k}}-x$ and therefore belongs to $\mathbb{GF}(p,\gcd(k,e))$.

Suppose $e'|e$, so $\mathbb{GF}(p,e')$ sits inside $\mathbb{GF}(p,e)$. If $f(\alpha)=0$ and the coefficients of f come from a field $\mathbb{GF}(p,e')$ then the coefficients c_i satisfy $\mathrm{Frob}^{e'}(c_i)=c_i$. Thus, $0=f(\alpha)=\sum c_i\alpha^i$ produces under e'-fold Frobenius that $0=\sum c_i(\alpha^{p^{e'}})^i$. In other words, $\mathrm{Frob}^{e'}(\alpha)$ is a root to the same polynomial as α. Since the degree of α over $\mathbb{GF}(p,1)$ is the product of the degree of α over $\mathbb{GF}(p,e')$ with e/e', it follows that the degree of the minimal polynomial of α over $\mathbb{GF}(p,e')$ is e/e'. This implies that iterating $\mathrm{Frob}^{e'}$ on α makes it circle throuugh all the roots of f. (If it did not move through all roots, one could take the roots it moves thorugh and construct a minimal polynomial of lower degree, which cannot be). □

Example XV.2. Let $p=3$ and $e=4$. There are 81 elements in $\mathbb{GF}(3,4)$. An irreducible polynomial of degree 4 is x^4-x^3-1, so we can view $\mathbb{GF}(3,4)$ as $\mathrm{Kron}(\mathbb{Z}/3\mathbb{Z},x^4-x^3-1)=\mathbb{Z}/3\mathbb{Z}[x]/\langle\langle x^4-x^3-1\rangle$.

A slightly horrendous calculation shows that $x^{3^4}-x$ factors as $(x)*(x-1)*(x+1)$ times $(x^2+1)*(x^2-x-1)*(x^2+x-1)$ times

$$\begin{aligned}&(x^4-x-1)*(x^4+x-1)*(x^4-x^2-1)*(x^4+x^2-1)*(x^4+x^2-x+1)*\\&\quad*(x^4+x^2+x+1)*(x^4-x^3-1)*(x^4-x^3+x+1)*(x^4-x^3-x^2+x-1)*\\&\quad*(x^4-x^3+x^2+1)*(x^4-x^3+x^2-x+1)*(x^4-x^3+x^2+x-1)*(x^4+x^3-1)*\\&\quad*(x^4+x^3-x+1)*(x^4+x^3-x^2-x-1)*(x^4+x^3+x^2+1)*\\&\quad*(x^4+x^3+x^2-x-1)*(x^4+x^3+x^2+x+1).\end{aligned}$$

In particular, there are 3 irreducible linear polynomials, 3 irreducible quadrics, and 18 irreducible quartics over $\mathbb{Z}/3\mathbb{Z}$. (We learn nothing about cubics, because cubics make field extension of degree 3, and right now we are looking at an extension of degree 4 and 3 does not divide 4, so no $\mathbb{GF}(3,3)$ is inside $\mathbb{GF}(3,4)$.)

Note that $3\cdot 1+3\cdot 2+18\cdot 4=3+6+72=81$ as it should.

So the roots of $x^{81}-x$, which are precisely the elements of $\mathbb{GF}(3,4)$, come in 3 types:

- elements of $\mathbb{GF}(3,1)$: as the roots to $x=0,x-1=0,x+1=0$;
- elements of $\mathbb{GF}(3,2)$ that are not in $\mathbb{GF}(3,1)$: they come in pairs of the roots of $x^2+1=0,x^2-x-1=0,x^2+x-1=0$;

- elements in $\mathbb{GF}(3,4)$ that are not in $\mathbb{GF}(3,2)$: these come in quadruplets as the roots of the 18 irreducible quartics.

Let us take the irreducible quadric x^2+1, and let α be the Kronecker root of $\mathbb{GF}(3,4) = \mathrm{Kron}(\mathbb{Z}/3\mathbb{Z}, x^4-x^3-1)$ for $f(x) = x^4-x^3-1$. In other words, $\alpha = \overline{x}$. Let's try to find a copy of $\mathbb{GF}(3,2)$ inside this field. This would require, for example, finding the roots to x^2+1 (one of the three irreducible quadrics above). We calculate

$$\begin{aligned}(\alpha^3+\alpha^2+1)^2+1 &= \alpha^6+2\alpha^5+\alpha^4+2\alpha^3+2\alpha^2+1+1\\ &= \alpha^2\cdot(\alpha^3+1)+2\alpha^5+\alpha^4+2\alpha^3+2\alpha^2+2\\ &= 3\alpha^5+\alpha^4+2\alpha^3+3\alpha^2+2\\ &= \alpha^4-\alpha^3-1=0.\end{aligned}$$

It follows that $\alpha^3+\alpha^2+1$ is a root of x^2+1. (The other root is $2(\alpha^3+\alpha^2+1)$.)

So, inside $\mathbb{GF}(3,4)$ the copy of $\mathbb{GF}(3,2)$ consists of the $\mathbb{Z}/3\mathbb{Z}$-linear combinations of 1 and $\beta := \alpha^3+\alpha^2+1$. These are the 9 elements

$$0, 1, 2, \beta, \beta+1, \beta+2, 2\beta, 2\beta+1, 2\beta+2.$$

Now look at what the Frobenius (third power map) does to them:

$$\begin{aligned}0^3 &= 0,\\ 1^3 &= 1,\\ 2^3 &= 8=2,\\ (\alpha^3+\alpha^2+1)^3 &= \alpha^9+\alpha^6+1 = \ldots = 2\alpha^3+2\alpha^2+2,\\ ((\alpha^3+\alpha^2+1)+1)^3 &= \ldots = 2\alpha^3+2\alpha^2+0,\\ ((\alpha^3+\alpha^2+1)+2)^3 &= \ldots = 2\alpha^3+2\alpha^2+1,\\ (2(\alpha^3+\alpha^2+1))^3 &= \ldots = \alpha^3+\alpha^2+1,\\ (2(\alpha^3+\alpha^2+1)+1)^3 &= \ldots = \alpha^3+\alpha^2+2,\\ (2(\alpha^3+\alpha^2+1)+2)^3 &= \ldots = \alpha^3+\alpha^2+0\end{aligned}$$

So, the third-power map flips them about in pairs. The 3 pairs correspond to the roots of the 3 irreducible quadrics above.

Now let us look what the Frobenius does to general elements of $\mathbb{GF}(3,4)$, those that do not live in smaller fields. As a starter, we look at what happens to α itself under iterates of Frob. It is clear that $\mathrm{Frob}(\alpha) = \alpha^3$ and we leave it like that since we can't rewrite polynomials of degree less than four.

Then $\mathrm{Frob}(\mathrm{Frob}(\alpha)) = \alpha^9$ and that can be rewritten (with labor) as $\alpha^3+\alpha^2+2\alpha$. The third power of this is $\alpha^3+2\alpha^2+1$, and the Frobenius sends this last guy to α. So the Frobenius action circles

$$\alpha \mapsto \alpha^3 \mapsto \alpha^3+\alpha^2+2\alpha \mapsto \alpha^3+2\alpha^2+1 \mapsto \alpha.$$

These 4 elements are the roots of x^4+x^3+1, since we took one such root, and applied Frobenius. (Frobenius takes the equation $x^4-x^3-1=0$ and turns it into $(x^3)^4-(x^3)^3-1=0$, so that if you "Frobenius a root" then you get a root back).

The same sort of thing happens to the roots of the other 17 irreducible quadrics: the Frobenius circles them within their lucky clover leaf, preserving that they are roots to whatever quadric they are roots of.

So, $\mathbb{Z}/4\mathbb{Z}$ (the 4 is because $e=4$ and the 4-th power of the Frobenius is the identity) acts on the 81 elements of $\mathbb{GF}(3,4)$. Three elements are fixed points, there

are thre orbits of size 2 (pairing the roots of the quadrics) and there are 18 orbits of size 4 (the 18 quadruplets that occur as roots of the irreducible quartics)

2. Symmetries

3. Applications

Stuff for later

3.1. Zerodivisors. When we listed the arithmetic operations that we can perform with cosets, we did not list division. There are good reasons for that. First off, we don't really expect division to work in general since even for usual integers division is problematic (try dividing 3 by 2, for example). But it is stranger than that. Even if dividing one integer by another would be just fine (let's say you planned to divide 12 by 6) it is not clear that in the modulo world this is still going as expected.

EXAMPLE XV.3. To get a feeling, let's try to divide 12 by 6 but modulo 8. The quotient, let's call it $\overline{a}$, should live in $\mathbb{Z}/8\mathbb{Z}$ and have the property that $\overline{a} \cdot \overline{6} = \overline{12}$. Of course, $\overline{a}$ could be $\overline{2}$.

But if you list the multiples of $\overline{6}$ you find:

$$\begin{array}{llll} \overline{0} \cdot \overline{6} = \overline{0}, & \overline{1} \cdot \overline{6} = \overline{6}, & \overline{2} \cdot \overline{6} = \overline{4}, & \overline{3} \cdot \overline{6} = \overline{2}, \\ \overline{4} \cdot \overline{6} = \overline{0}, & \overline{5} \cdot \overline{6} = \overline{6}, & \overline{6} \cdot \overline{6} = \overline{4}, & \overline{7} \cdot \overline{6} = \overline{2}. \end{array}$$

So we see that there are actually *two* different cosets that compete for being a quotient $\overline{12}/\overline{6}$, namely $\overline{2}$ and $\overline{6}$. This comes from the fact that we can think of $\overline{12}$ also as $\overline{4}$.

Moreover, one can see that the people in $\mathbb{Z}/8\mathbb{Z}$ split into two classes, the cosets that are multiples of $\overline{6}$ and those that are not, where each coset that shows up at all as multiple of $\overline{6}$ shows exactly twice. ◇

EXAMPLE XV.4. This time, let's try to divide 7 by 5. Usually that would not seem like a good idea (at least if you hope for integer answers), but let's do this again modulo 8. Writing out the multiples of $\overline{5}$ in $\mathbb{Z}/8\mathbb{Z}$ we find

$$\begin{array}{llll} \overline{0} \cdot \overline{5} = \overline{0}, & \overline{1} \cdot \overline{5} = \overline{5}, & \overline{2} \cdot \overline{5} = \overline{2}, & \overline{3} \cdot \overline{5} = \overline{7}, \\ \overline{4} \cdot \overline{5} = \overline{4}, & \overline{5} \cdot \overline{5} = \overline{1}, & \overline{6} \cdot \overline{5} = \overline{6}, & \overline{7} \cdot \overline{5} = \overline{3}. \end{array}$$

So, quite against expectations, $\overline{7}/\overline{5}$ can be found in $\mathbb{Z}/8\mathbb{Z}$, and there is exactly one answer: $\overline{3}$. In fact, as one can see, any coset in $\mathbb{Z}/8\mathbb{Z}$ can be divided by $\overline{5}$ in exactly one way.

In this section we will try to understand and predict this kind of behavior. ◇

The coset of 0 in $\mathbb{Z}/n\mathbb{Z}$ is "the zero" in this new system of numbers, since adding it to any coset does not change the coset. As seen in Example XV.3 above, it is possible that this new zero shows up as a product of nonzero inputs, a phenomenon not encountered in the integers.

DEFINITION XV.5. If $\overline{a}, \overline{b}$ are in $\mathbb{Z}/n\mathbb{Z}$, with neither a nor b divisible by n, then they are called *zerodivisors* if $\overline{a}\overline{b} = \overline{0}$.

This ability to multiply to zero in $\mathbb{Z}/n\mathbb{Z}$ of course comes from the fact that we equate (every multiple of) n with zero. So, a composite n will allow for products to be zero (that is, multiples of n) in several ways. We try to understand by way of an example.

EXAMPLE XV.6. Let $n = 6$; then $\overline{2} \cdot \overline{3} = \overline{0}$.

Indeed, in order to prepare what is to come in a bit, let's list all multiples of $\overline{2}$:

$$\overline{2} \cdot \overline{0} = \overline{0}, \quad \overline{2} \cdot \overline{1} = \overline{2}, \quad \overline{2} \cdot \overline{2} = \overline{4}, \quad \overline{2} \cdot \overline{3} = \overline{0}, \quad \overline{2} \cdot \overline{4} = \overline{2}, \quad \overline{2} \cdot \overline{5} = \overline{4}.$$

◇

The reason that $\overline{2}$ was capable to yield $\overline{0}$ when multiplied with a nonzero coset was of course that 2 has an interesting common factor with 6. In the general case, suppose $\overline{a}$ is a coset in $\mathbb{Z}/n\mathbb{Z}$ and we look for another element $\overline{b} \in \mathbb{Z}/n\mathbb{Z}$ such that $\overline{a}\overline{b} = \overline{0}$. If we set $\gcd(a, n) = d$ and if d happens to be greater than 1, then we can write $n = d \cdot e$ and so $\overline{a} \cdot \overline{e}$ is a multiple of $\overline{d} \cdot \overline{e} = \overline{(de)} = \overline{n} = \overline{0}$. But a multiple of $\overline{0}$ must be $\overline{0}$ itself.

On the other hand, pick now an a such that $\gcd(a, n) = 1$. This means by Proposition I.18 that there are integers α, β with $a\alpha + n\beta = 1$. Reading this "modulo n", we get $\overline{a} \cdot \overline{\alpha} + \overline{(n\beta)} = \overline{1}$. Naturally, $\overline{(n\beta)} = \overline{n} \cdot \overline{\beta} = \overline{0}$. So, $\overline{a} \cdot \overline{\alpha} = \overline{1}$. It follows that for any $b \in \mathbb{Z}$, $\overline{a} \cdot \overline{(b\alpha)} = \overline{b}$. This says that every single coset in $\mathbb{Z}/n\mathbb{Z}$ is the result of some coset being multiplied by a.

Let's try to understand what this means. There are n cosets in $\mathbb{Z}/n\mathbb{Z}$, each of which you can multiply with a. The process of multiplication produces all n of these (provided $\gcd(a, n) = 1$). It follows there is exactly one coset that when multiplied by a gives you any given coset $\overline{b}$. In particular, there is only one coset that when multiplied gives $\overline{0}$ (and of course this one coset is $\overline{0}$ itself).

Putting it all together, we have proved most of the following theorem:

THEOREM XV.7. *If $\gcd(a, n) = 1$ then multiplication by a is a bijection on $\mathbb{Z}/n\mathbb{Z}$. In other words, for each $\overline{b} \in \mathbb{Z}/n\mathbb{Z}$ there is exactly one $\overline{x} \in \mathbb{Z}/n\mathbb{Z}$ such that $a \cdot \overline{x} = \overline{b}$. Yet in other words, $\overline{a}$ becomes a unit in $\mathbb{Z}/n\mathbb{Z}$.*

Conversely, if $\gcd(a, n) = d > 1$ then multiplication by a is neither surjective nor injective. There are exactly n/d different cosets that arise through multiplication by a, and each is d times output of such a multiplication. In this case, $\overline{a}$ is a zerodivisor in $\mathbb{Z}/n\mathbb{Z}$.

EXAMPLE XV.8. Let $n = 6$. The numbers a that have $\gcd(a, n) = 1$ are living in the cosets $\overline{1}$ and $\overline{5}$. Everyone is a multiple of $\overline{1}$ for obvious reasons, and everyone is also a multiple of $\overline{5}$ because $\overline{5} = \overline{-1}$.

The multiples of $\overline{2}$ are $\{\overline{0}, \overline{2}, \overline{4}\}$, and these are also exactly the multiples of $\overline{4}$. Note that each one of $\{\overline{0}, \overline{2}, \overline{4}\}$ is a multiple of both $\overline{2}$ and $\overline{4}$ in $2 = \gcd(6, 2) = \gcd(6, 4)$ ways. For example, $\overline{4} = \overline{4} \times \overline{1} = \overline{4} \times \overline{4}$ and also $\overline{4} = \overline{2} \times \overline{2} = \overline{2} \times \overline{5}$.

The multiples of $\overline{3}$ are $\overline{3}$ and $\overline{0}$, and each of $\{\overline{0}, \overline{3}\}$ arises $3 = \gcd(3, 6)$ times as multiple. For example, $\overline{3} = \overline{3} \times \overline{1} = \overline{3} \times \overline{3} = \overline{3} \times \overline{5}$. ◇

EXERCISE XV.9. For $n = 10$ and $a = 1, 2, \ldots, 9$ determine

(1) which cosets in $\mathbb{Z}/10\mathbb{Z}$ are multiples of $\overline{a}$;

(2) how many cosets in $\mathbb{Z}/10\mathbb{Z}$ are multiples of $\overline{a}$ and express these numbers in terms of a and 10.

◇

Theorem XV.7 implies that the units of $\mathbb{Z}/n\mathbb{Z}$ are exactly the cosets of those numbers between 1 and $n-1$ inclusive that are relatively prime to n. All other cosets exhibit ambiguity (at best) or impossibility (at worst) when trying to divide by them. Which case happens depends on the two cosets to be divided. For example, in $\mathbb{Z}/4\mathbb{Z}$, trying to divide by $\overline{2}$ one fails when the input is $\overline{1}$ or $\overline{3}$ while one gets too many suggestions when one divides $\overline{2}$ by $\overline{2}$ (namely, $\overline{1}$ and $\overline{3}$) or when one divides $\overline{0}$ by $\overline{2}$ (namely, $\overline{0}$ and $\overline{2}$).

In order to explain this behavior, we shall need the following observation:

EXERCISE XV.10. Prove that $\operatorname{lcm}(a,b)\cdot\gcd(a,b) = ab$. ◇

Now suppose $\overline{a}\cdot\overline{x} = \overline{b}$ has at least one solution, so $ax-b$ is a multiple of n. If you added $c = n/\gcd(a,n)$ to x then we calculate: $\overline{a}\overline{(x+c)} = \overline{ax} + \overline{ac} = \overline{b} + \overline{(an/\gcd(a,n))} = \overline{b} + \overline{\operatorname{lcm}(a,n)} = \overline{b}$ since $\overline{\operatorname{lcm}(a,n)} = \overline{0}$ (like any other multiple of n) represents the coset of zero. It follows that besides $\overline{x}$ all expressions $\overline{x+i\cdot c}$ are also solutions to $\overline{ax} = \overline{b}$.

How many such are there? On the face of it, infinitely many but recall that $x+i\cdot c$ and $x+j\cdot c$ are in the same coset of $\mathbb{Z}/n\mathbb{Z}$ as soon as $(x+i\cdot c)-(x+j\cdot c) = (i-j)c$ is a multiple of n. That of course happens exactly if $i-j$ is a multiple of n/c. So, there are n/c different cosets $\overline{x}$, $\overline{x+c}$, …, $\overline{x+((n/c)-1)c}$ that all solve $\overline{ax} = \overline{b}$. (Of course, $n/c = \gcd(a,n)$ by definition of c).

EXERCISE XV.11. Group the elements of $\mathbb{Z}/24\mathbb{Z}$ in such a way that two cosets $\overline{a}$, $\overline{b}$ are in the same group exactly when their sets of multiples $\{\overline{1a}, \overline{2a}, \overline{3a}, \ldots\}$ and $\{\overline{1b}, \overline{2b}, \overline{3b}, \ldots\}$ agree as sets (perhaps after reordering). Describe in words each group. ◇

REMARK XV.12. The Euclidean algorithm can also be carried out in the polynomial ring $\mathbb{R}[x]$; the idea of size (absolute value for integers) in the Archimedean principle is then taken over by the degree of the polynomial. The relevant statement is then:

For all polynomials $a(x), b(x)$ in $\mathbb{R}[x]$ there are $q(x), r(x) \in \mathbb{R}[x]$ such that $a(x) = b(x)q(x)+r(x)$ and $0 \le \deg(r) \le \deg(b)-1$.

The polynomials $q(x)$ and $r(x)$ are furnished by the method of (polynomial) long division. Exactly as for integers, one can work this division process into an algorithm to compute the gcd between polynomials. ◇

EXERCISE XV.13. Compute the gcd between
(1) x^3+1 and x^1+1;
(2) x^3+1 and x^2+1;
(3) x^3+1 and x^4+1;
(4) x^3+1 and x^5+1;
(5) x^3+1 and x^6+1;
(6) x^3+1 and x^n+1 for any natural number n (this will require to consider cases depending on the remainder of division of n by 6.

◇

3.2. Cartesian Products, Euler's ϕ-function, Chinese Remainder. We wish to find a formula for the number of cosets in $\mathbb{Z}/n\mathbb{Z}$ that are units. By Theorem XV.7, we need to count the numbers on the list $1, \ldots, n-1$ that are coprime to n. For this, recall the Euler ϕ-function from Definition I.32.

If p is a prime number, it is clear that $\phi(p) = p - 1$. So, $\mathbb{Z}/p\mathbb{Z}$ has $p - 1$ units whenever p is prime.

EXERCISE XV.14. (1) Determine for $n = 4, 8, 9, 16$ the value of $\phi(n)$ by listing explicitly the units in $\mathbb{Z}/n\mathbb{Z}$.

(2) Suppose $n = p^k$ is a power of a prime number p. Prove that $\phi(n)$ is $n - n/p$. ◇

If ϕ is composite, the question becomes more interesting. Below, we will discuss that if n is factored into relatively prime factors $ab = n$ there is an easy formula: if $\gcd(a, b) = 1$ then $\phi(ab) = \phi(a) \cdot \phi(b)$. For example, $\phi(12) = \phi(4) \times \phi(3) = 2 \cdot 2 = 4$. (It is important to note that the gcd-condition is crucial: $\phi(16)$ is not $\phi(4) \cdot \phi(4)$, compare Exercise XV.14 above).

In order to understand why for coprime a, b the ϕ-function should be multiplicative, we take the following approach. Let $n = ab$ and assume $\gcd(a, b) = 1$. For simplicity of notation, write $\Phi(n)$ for the numbers on the list $\{0, 1, \ldots, n - 1\}$ that are coprime to n, and for any two numbers $r, s \in \mathbb{Z}$ write $r\%s$ for the remainder of division of r by s coming from Euclid's Theorem. Note that $|\Phi(n)| = \phi(n)$.

Now pick $i \in \Phi(n)$. Then surely $\gcd(i, a) = \gcd(i, b) = 1$, and so $i\%a$ is in $\Phi(a)$ while $i\%b$ is in $\Phi(b)$. So one could make up a function that takes inputs in $\Phi(n)$ and outputs pairs whose first component is in $\Phi(a)$ and whose second component is in $\Phi(b)$; the function would just send i to $(i\%a, i\%b)$. If we could show that this function is reversible (that is, one could construct for each pair with first coordinate in $\Phi(a)$ and with second coordinate in $\Phi(b)$ an $i \in \Phi(n)$ that produces this pair via the function) then $\phi(n)$ should be $\phi(a) \cdot \phi(b)$.

Let's look at an example.

EXAMPLE XV.15. For $n = 12$, $a = 4$ and $b = 3$ we have $\Phi(12) = \{1, 5, 7, 11\}$, $\Phi(4) = \{1, 3\}$ and $\Phi(3) = \{1, 2\}$. The function discussed above sends: 1 mod 12 to (1 mod 4, 1 mod 3); 5 mod 12 to (1 mod 4, 2 mod 3); 7 mod 12 to (3 mod 4, 1 mod 3); 11 mod 12 to (3 mod 4, 2 mod 3).

Unfortunately, if you multiply the Φ-sets directly, you get $\{1, 2\} \cdot \{1, 3\} = \{1, 2, 3, 6\}$ which are mostly not units in $\mathbb{Z}/12\mathbb{Z}$. So while $\Phi(12) = \{1, 5, 7, 11\}$ is not $\Phi(4) \cdot \Phi(3) = \{1, 2, 3, 6\}$, we do have at least $\phi(12) = \phi(4) \cdot \phi(3)$. ◇

The example teaches that one should not multiply units in $\mathbb{Z}/a\mathbb{Z}$ with units in $\mathbb{Z}/b\mathbb{Z}$ and hope to get units in $\mathbb{Z}/n\mathbb{Z}$. Indeed, what we need to do is: given $i \in \mathbb{Z}/a\mathbb{Z}$ and $j \in \mathbb{Z}/b\mathbb{Z}$, find $x \in \mathbb{Z}$ such that

$$(x \bmod a) = (r \bmod a) \text{ and } (x \bmod b) = (s \bmod b).$$

Before we go and look for x, note that changing a solution x by a multiple of $n = ab$ does not change the solution property: if x is a solution then so are $\ldots, x - 2ab, x - ab, x, x + ab, x + 2ab, \ldots$. Conversely, if x and x' are both solutions, then $a|(x - x')$ and $b|(x - x')$. Of course, this means that $x - x'$ is a simultaneous multiple of both a and b (and hence of their lcm), and since $\gcd(a, b) = 1$ is assumed, $x - x'$ is a multiple of $\operatorname{lcm}(a, b) = ab/\gcd(a, b) = ab = n$. Therefore, the solutions x, if they exist at all, form precisely one coset of $\mathbb{Z}/n\mathbb{Z}$.

So the whole question boils down to: if you take a pair of cosets modulo a and b respectively, can you find a coset modulo n that "gives birth" to the given cosets by going modulo a and b respectively. Let's look at an example.

EXAMPLE XV.16. Let $n = 36$, factored as $36 = 4 \cdot 9$. Choose $r = 2$ and $s = 7$. Is there $x + 36\mathbb{Z}$ such that $x + 4\mathbb{Z} = 2 + 4\mathbb{Z}$ while $x + 9\mathbb{Z} = 7 + 9\mathbb{Z}$?

Some experimentation reveals that, yes, there is such x; anything of the form $34 + k \cdot 36$ will do, so that $x + 36\mathbb{Z} = 34 + 36\mathbb{Z}$. But how can one go about this systematically? Here is how.

If x leaves rest 2 when divided by 4 then x mod 36 must look like one of the numbers $2 + 4k$, $0 \leq k \leq 8$. Similarly, if x leaves rest 7 when divided by 9, then x mod 36 must look like $7+9\ell$, with $0 \leq \ell \leq 3$. In other words, we want k and ℓ such that $(7+9\ell)$ and $(2+4k)$ differ by a multiple of 36: $(7+9\ell)+36\mathbb{Z} = (2+4k)+36\mathbb{Z}$

Now go back to $\mathbb{Z}/4\mathbb{Z}$ where this reads $(7 + 9\ell) + 4\mathbb{Z} = (2 + 4k) + 4\mathbb{Z}$ or $3 + 1 \cdot \ell + 4\mathbb{Z} = 2 + 4\mathbb{Z}$. This is fancy speak for: $1 + 1 \cdot \ell$ is a multiple of 4. Pick $\ell = 3$. It follows that $x = 7 + 9\ell = 7 + 9 \cdot 3 = 34$ mod 36.

One could also have gone the other way, and reduce modulo 9: one learns from $(7+9\ell)+36\mathbb{Z} = (2+4k)+36\mathbb{Z}$ that one should also have $(7+9\ell)+9\mathbb{Z} = (2+4k)+9\mathbb{Z}$ which boils down to: $5 - 4k$ is a multiple of 9. So we'd like to solve $5 = 4k$ mod 9. (This doesn't seem so easy as before. We were rather lucky earlier, because 9 mod 4 is 1 and so the coefficient in "$1 + 1 \cdot \ell$ is a multiple of 4" was 1, making it easy to solve for ℓ). Since 4 is coprime to 9 (by hypothesis, not by accident!) there is actually such a k. Tests show that $4 \cdot 8 = 32 = 5$ mod 9. So, $k = 8$ would work. We see then that $x + 36\mathbb{Z}$ should be $2 + 4 \cdot 8 + 36\mathbb{Z} = 34 + 36\mathbb{Z}$, as we already found twice. ◇

In the following theorem, we state formally what exactly happened in the computation above, and how to accomplish it.

THEOREM XV.17 (Chinese Remainder Theorem). *Suppose* $\gcd(a, b) = 1$ *and set* $ab = n$. *Choose integers* r, s. *The set of integers* x *for which*

$$x + a\mathbb{Z} = r + a\mathbb{Z} \text{ and } x + b\mathbb{Z} = s + b\mathbb{Z}$$

fills exactly one coset inside $\mathbb{Z}/n\mathbb{Z}$.

This coset can be found as follows. Find integers i, j *such that* $i \cdot b = 1$ *mod* a *and* $j \cdot a = 1$ *mod* b. *Let* $x = r \cdot i \cdot b + s \cdot j \cdot a$. *Then* $x + n\mathbb{Z}$ *is the sought after coset.*

In the above example, $a = 4, b = 9, i = 1, j = 7, r = 2, s = 7$. Hence $x = 2 \cdot 1 \cdot 9 + 7 \cdot 7 \cdot 4 = 18 + 196 = 214 = 34$ mod 36.

EXAMPLE XV.18. Let's solve

$$x = 13 \text{ mod } 29 \text{ and } x = 8 \text{ mod } 12.$$

Matching letters, we have $a = 29$, $b = 12$, $r = 13$, $s = 8$. We need to find i, j with $i \cdot 12 = 1$ mod 29 and $j \cdot 29 = 1$ mod 12. Both will come out of the Euclidean algorithm when applied to 12 and 29:

$$29 = 2 \cdot 12 + 5; \quad 12 = 2 \cdot 5 + 2; \quad 5 = 2 \cdot 2 + 1.$$

We derive (going backwards):

$$1 = 5 - 2 \cdot 2 = 5 - 2 \cdot (12 - 2 \cdot 5) = 5 \cdot 5 - 2 \cdot 12 = 5 \cdot (29 - 2 \cdot 12) - 2 \cdot 12 = 5 \cdot 29 - 12 \cdot 12.$$

It follows that we can take $i = -12$ and $j = 5$. This gives $x = r \cdot i \cdot b + s \cdot j \cdot a = -712$. One can test easily that this is correct. ◇

If one needs to solve three simultaneous equations,

$$x = r \text{ mod } a; \quad x = s \text{ mod } b; \quad x = t \text{ mod } c,$$

with $\gcd(a,b) = \gcd(b,c) = \gcd(c,a) = 1$, one has two options. Either, take the souped-up version of the Chinese Remainder Theorem which we state below. Or, one first solves $x = r \bmod a$ and $x = s \bmod b$ as above and then $y = x \bmod ab$ and $y = t \bmod c$ again as above.

Here is the multiverse formulation of the Chinese Remainder Theorem; its proof is in parallel to its little brother above.

THEOREM XV.19 (Chinese Remainder Theorem). *Let $n_1, n_2, \ldots, n_t$ be pairwise coprime numbers. Choose values $a_1, \ldots, a_t$. Then the set of integers x which leave remainder a_i when divided by n_i for all i are the elements in the coset $x + n_1 \cdots n_t\mathbb{Z}$ determined as follows. Let $N = n_1 \cdots n_t$ and set $N_i = N/n_i$. Find, for each i, a solution x_i to the equation $N_i \cdot x_i = 1 \bmod n_i$. Then*

$$x = \sum_{i=1}^{t} x_i N_i a_i.$$

EXERCISE XV.20. Solve the simultaneous equations
(1) $x = 1 \bmod 3$, $x = 2 \bmod 5$, $x = 3 \bmod 7$.
(2) $x = 2 \bmod 3$, $x = 2 \bmod 5$, $x = 3 \bmod 7$.
(3) $x = 1 \bmod 9$, $x = 2 \bmod 5$, $x = 3 \bmod 7$.

◇

It now follows from the Chinese Remainder theorem that:

COROLLARY XV.21. *If you factor $n = a_1 \cdots a_k$ into pairwise prime numbers $a_1, \ldots, a_k$ then $\phi(n) = \phi(a_1) \cdot \ldots \cdot \phi(a_k)$.*

In particular, if $n = p_1^{a_1} \cdot p_2^{a_2} \cdots p_k^{a_k}$ and all p_i are prime and distinct and all a_i are positive, then

$$\phi(n) = n \cdot (1 - \frac{1}{p_1})(1 - \frac{1}{p_2}) \cdots (1 - \frac{1}{p_k}) = \prod \left(p_i^{a_i} - p_i^{a_i - 1}\right).$$

EXERCISE XV.22. (1) Determine $\phi(666)$.
(2) Determine $\phi(720)$.
(3) Is there an $n \neq 29$ with $\phi(n) = 28$?
(4) Is there an n with $\phi(n) = 24$?
(5) Is there an n with $\phi(n) = 14$?
(6) Prove that $\phi(n)$ is always even if $n > 2$.

◇

3.3. Fermat's little theorem.

THEOREM XV.23. *For any prime number p and any $a \in \mathbb{N}$ with $gcd(a,p) = 1$, $a^p = a \bmod p$.*

In fact, unless p divides a, one has $a^{p-1} = 1 \bmod p$.

PROOF. If p divides a then the first equation is obviously true, so assume p does not divide a. Then $\gcd(p,a) = 1$ since p is a prime number.

The numbers $1, 2, \ldots, p-1$ are coprime to p and so their cosets modulo p are units in $\mathbb{Z}/p\mathbb{Z}$. In particular, $\overline{a}$ is a unit in $\mathbb{Z}/p\mathbb{Z}$ and so multiplication by $\overline{a}$ is a bijection on the elements of $\mathbb{Z}/p\mathbb{Z}$, since for units multiplication is an invertible operation. That means that the sets $\{\overline{1}, \overline{2}, \ldots, \overline{p-1}\}$ and $\{\overline{a}, \overline{2a}, \ldots, \overline{(p-1)a}\}$ are the same, up to reordering.

Equal sets have equal products:

$$\prod_{i=1}^{p-1} \overline{i} = \prod_{i=1}^{p-1} \overline{ai} = \overline{a}^{p-1} \cdot \prod_{i=1}^{p-1} \overline{i}.$$

The product of units is a unit, so one can cancel out the term $\prod_{i=1}^{p-1} \overline{i}$ and obtain $1 = a^{p-1} \bmod p$. □

Exercise XV.24. By looking at the smallest non-prime number, show that for composite numbers n the equation $a^p = a \bmod n$ may fail. ◇

Exercise XV.25. (Not easy). Formulate and prove a theorem like Fermat's little theorem in the case where p is not prime. ◇

Exercise XV.26. Show that $(n-1)n(n+1)$ is a multiple of 24 if n is a prime number greater than 2. Is "prime" really needed? ◇

Exercise XV.27. Suppose n is a number with k digits $a_k, \ldots, a_0$: $n = a_k \cdot 10^k + \ldots + a_1 \cdot 10 + a_0$. Show that n is divisible by 7 if and only if the expression

$$\cdots + 5a_5 + 4a_4 + 6a_3 + 2a_2 + 3a_1 + 1 \cdot a_0$$

is divisible by 7. Here, the dots towards the left mean that the coefficient pattern $5, 4, 6, 2, 3, 1$ that appears should be repeated. So, a_6 gets coefficient 1 again (like a_0), a_7 gets 3 again (like a_1) and so on: the coefficient of a_{i+6} is the same as of a_i. ◇