

Lecture notes on Abstract Algebra

Uli Walther

©2021

Version of Fall 2025

Contents

Basic notions	7
0.1. How to use these notes	7
0.2. Set lingo	7
0.3. Size of sets	8
0.4. Finite vs infinite	10
0.5. Inclusion/Exclusion	10
Chapter I. Week 1: Introduction	13
1. Induction	13
1.1. Setup	13
1.2. The idea	13
1.3. Archimedean property and well-order	15
2. Arithmetic	18
2.1. The Euclidean algorithm	19
2.2. Primes and irreducibles	21
2.3. Some famous theorems and open problems on prime numbers	23
3. Modular arithmetic	26
3.1. Computing “modulo”: $\mathbb{Z}/n\mathbb{Z}$	26
3.2. Divisibility tests	27
Chapter II. Week 2: Groups	29
1. Symmetries	29
2. Groups	31
3. Cyclic and Abelian groups	33
4. Automorphisms	35
5. Free groups	37
Chapter III. Week 3: $\mathbb{Z}/n\mathbb{Z}$ and cyclic groups	39
1. Subgroups of cyclic groups	39
2. Products and simultaneous modular equations	42
3. $U(n)$: Automorphisms of $\mathbb{Z}/n\mathbb{Z}$ and the Euler ϕ function	43
Chapter IV. Week 4: Cosets and morphisms	47
1. Equivalence relations	47
2. Morphisms	48
3. Cosets for subgroups and Lagrange	50
4. Kernels and normal subgroups	52
Chapter V. Week 5: Permutations and the symmetric group	55
Chapter VI. Week 6: Quotients and the Isomorphism Theorem	61

1. Making quotients	61
2. The isomorphism theorem	64
Chapter VII. Week 7: Finitely generated Abelian groups	69
1. Row reduced echelon form over the integers	69
2. Generating groups	72
Chapter VIII. Week 8: Group actions	77
Review	83
Chapter IX. Week 9: Introduction to rings	87
Chapter X. Week 9/10: Ideals and morphisms	91
Chapter XI. Week 10, Euclidean rings	95
1. Euclidean rings	96
Chapter XII. Week 11/12: Divisibility, Field Extensions	103
1. Divisibility	103
2. Making new fields from old	107
Chapter XIII. Week 12/13: Splitting fields and extension towers	111
1. Splitting fields	111
2. Roots with multiplicity	113
Chapter XIV. Week 13/14: Minimal polynomials and finite fields	117
1. Minimal Polynomials	117
2. Finite Fields	120
Chapter XV. Week 14: Galois	125
1. The Frobenius	125
2. Review	128
Applications	131
3. Geometric Constructions	131
3.1. The Delian Problem	131
3.2. Trisecting angles	132
3.3. Regular n -gons	133
4. Solving equations by radicals	134
Various thoughts	137
4.1. Zerodivisors	137
4.2. Cartesian Products, Euler's ϕ -function, Chinese Remainder	139
Index	143
Bibliography	145

Expected progress:

- Week 1: Archimedes, Factorization
- Week 2: Symmetries, Groups, Subgroups, Order, $\text{Aut}(G)$
- 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, Group Actions
- Week 8: Group Actions, Review, Rings
- Week 9: Rings, Midterm, Ideals
- Week 10: Morphisms, Euclidean Rings
- Week 11: PID, UFD, Fields
- Week 12: Extensions, Eisenstein, Kronecker
- Week 13: Multiplicity, Splitting Fields, Minimal Polynomial
- Week 14: Finite Fields, Frobenius, Separability
- 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 a common 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.

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

NOTATION .2. The following symbols will be used throughout to denote number sets:

- the *natural numbers* $0, 1, 2, 3, \dots$ are denoted by \mathbb{N} ;
- the *integer numbers* $\dots, -3, -2, -1, 0, 1, 2, 3, \dots$ 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} .

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 both 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 *empty set* (denoted \emptyset) and, despite its humble content, one of the most important sets. One uses the notation $S = \{s_1, \dots, s_n, \dots\}$ 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).

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 $\{a_1, \dots, a_n\}$ and an *ordered n -tuple* (a_1, \dots, 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, \dots, a_n has no emphasis on order, but it can have repeated entries.

A *function* $\phi: A \rightarrow B$ 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 $\phi: \mathbb{Z} \rightarrow \mathbb{Z}$ 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 .3. The function $\phi: A \rightarrow B$ 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 .4. We write $|S| \leq |S'|$ if there is an injective map $\phi: S \hookrightarrow S'$.

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

EXAMPLE .5.

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

EXERCISE .6. Show that $|\mathbb{Z}| \leq |\mathbb{N}|$. ◇

EXAMPLE .7. • $|\mathbb{Z}| \leq |\mathbb{Q}|$ 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:

$p \backslash q$	1	2	3	4	...
1	1/1	1/2	1/3	1/4	...
2	2/1	2/2	2/3	2/4	...
3	3/1	3/2	3/3	3/4	...
4	4/1	4/2	4/3	4/4	...
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots

Clearly all positive rationals \mathbb{Q}_+ 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 $(0, 1) = \{r_0, r_1, r_2, \dots\}$ 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 $0.r_{i,1}r_{i,2}r_{i,3}\dots$. Now write the real numbers in the interval $(0, 1)$ into a two-way infinite table, listing their digits ($r_{i,j}$ is the j -th digit of r_i):

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

(We assume that no infinite sequence $999\dots$ appears, to make this presentation unique). 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

¹All 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

either self-contradictory or allows unanswerable questions. Cohen showed that, in particular, the continuum hypothesis cannot be decided with Zermelo–Fraenkel’s axioms). \diamond

0.4. Finite vs infinite. Some sets S allow injections into themselves that are not surjective. For example, one can make a function $\phi: \mathbb{N} \rightarrow \mathbb{N}$ 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 $\phi: S \rightarrow S$ 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 $|S| \leq |\{1, 2, \dots, n\}|$ ”. 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 .8. 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 \emptyset 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 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. \diamond

EXERCISE .9. 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?) \diamond

REMARK .10. If S is finite of size n and $k \in \mathbb{N}$ satisfies $0 \leq k \leq n$ then the collection of sets in S that have size k is denoted by $\binom{S}{k}$. The size of this collection is the familiar number $\binom{|S|}{k}$. \diamond

EXERCISE .11. 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 exactly one element called s , then the options for (A, B) are: (\emptyset, \emptyset) , $(\emptyset, \{s\})$ and $(\{s\}, \emptyset)$. \diamond

EXERCISE .12. 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. \diamond

0.5. Inclusion/Exclusion.

NOTATION .13. 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, \dots, 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 .14 (Inclusion/Exclusion Formula). *For any n finite sets A_1, \dots, A_n ,*

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

REMARK .15. 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. \diamond

EXERCISE .16. Let $S = \{1, 2, \dots, 666\}$. 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.

\diamond

CHAPTER I

Week 1: Introduction

NOTATION I.1. 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 $1 + 2 + \dots + n$ 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 $1 + 2 + \dots + n$ equals $n(n+1)/2$ ” from above).

The task at hand is to prove that *all* statements $P(0), P(1), \dots$ 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.2. 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 $P(n+1)$. \diamond

EXAMPLE I.3. We consider the question from the start of the section: show that $0 + 1 + \dots + n = n(n+1)/2$. So, for $n \in \mathbb{N}$ we let the statement $P(n)$ be “ $0 + 1 + \dots + n = n(n+1)/2$ ”.

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 equations

$$P(n) : \quad 1 + 2 + \dots + i = i(i+1)/2$$

for $0 \leq i \leq n$, then one should also believe in the equation

$$P(n+1) : \quad 1 + 2 + \dots + n + (n+1) = (n+1)(n+1+1)/2.$$

(Often one relies only on the truth of $P(n)$ to show that of $P(n+1)$, but there are some exceptions. Relying on all $0 \leq i \leq n$ is sometimes called “strong induction”, but one can prove exactly the same statements with strong and with “regular” induction). 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. \diamond

REMARK I.4. 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 $1 + 2 + \dots + n = n(n+1)/2 + 6$. Let’s write $P'(n)$ to be the statement “ $1 + 2 + \dots + n = n(n+1)/2 + 6$ ”. Now argue as follows: suppose that for some $n \in \mathbb{N}$, $P'(n)$ is true: $1 + 2 + \dots + n = n(n+1)/2 + 6$. Add $n+1$ on both sides to obtain $1 + 2 + \dots + n + (n+1) = n(n+1)/2 + 6 + n + 1 = [n(n+1) + 2(n+1)]/2 + 6 = (n+1)(n+2)/2 + 6$. 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.3 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! \diamond

REMARK I.5. 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, \dots, H_{n+1} . Let T_1 stand for the size n group of the first n horses, $T_1 = \{H_1, \dots, H_n\}$. Let T_2 stand for the last n horses, $T_2 = \{H_2, \dots, H_{n+1}\}$. 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, \dots, 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 $0, \dots, n$. 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, \dots, 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. \diamond

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.

This is a notable property because S might have infinitely many elements. Any *finite* set has a minimum for 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.6. The example above 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, $P(0), P(1), \dots, P(b - 1)$ 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. \diamond

REMARK I.7. 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. \diamond

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

THEOREM I.8 (Archimedean property). *Choose $a, b \in \mathbb{N}$ with $0 < b$. Then the sequence $b, b + b, b + b + b, \dots$ contains an element that exceeds a , so that $a - (b + \dots + b) < 0$. In other words, $\exists k \in \mathbb{N}$ with $kb > a$.* \square

I call this a theorem because if one wrote down the axiomatic definition of \mathbb{N} then this property is one that one needs to prove from the axioms. This axiomatic definition, 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.9. 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 $s - 1, s - 2, s - 3, \dots$ 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.10. 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 $(0/1)^2$ and $(1/1)^2$ are less than 2 while $(m/1)^2 > 2$ 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. \diamond

- 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 $\frac{1}{1 \cdot 2 \cdot 3} + \frac{1}{2 \cdot 3 \cdot 4} + \cdots + \frac{1}{n \cdot (n+1) \cdot (n+2)} = \frac{n(n+3)}{4 \cdot (n+1) \cdot (n+2)}$ 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 $1 + \cdots + 1 = n$ and $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$. Now show that $1 \cdot 2 + 2 \cdot 3 + \cdots + n \cdot (n+1) = \frac{n(n+1)(n+2)}{3}$.
- (8) Show that $1 \cdot 2 \cdot 3 + 2 \cdot 3 \cdot 4 + \cdots + n \cdot (n+1) \cdot (n+2) = \frac{n(n+1)(n+2)(n+3)}{4}$.
- (9) Generalize the two previous exercises to products of any length.
- (10) Show that $1 + 3 + 5 + \cdots + (2n-1) = n^2$ 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 $\sum_{\emptyset \neq S \subseteq \{1, \dots, n\}} \frac{1}{\prod_{\sigma \in S} \sigma} = n$. (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 $1, 2, 3, \dots, 2N$ 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 a *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: $r \cdot (s + t) = r \cdot s + r \cdot t$ 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 $0 \leq r < |b|$.

(Find the smallest integer q such that qb beats a and let r be $a - qb$). 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. 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 $a \nmid b$.

DEFINITION I.17. Let $a, b \in \mathbb{Z}$. Then $\gcd(a, b) = \max\{d \in \mathbb{N} \text{ with } d|a \text{ and } d|b\}$ and $\text{lcm}(a, b) = \min\{m \in \mathbb{N} \text{ with } a|m \text{ and } b|m\}$. 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

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

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.18 (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 $0 \leq r_i \leq |c_i| - 1$ using the Archimedean property.
- Set $c_{i+1} := r_i$.
- Replace i by $i + 1$.

Until:

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

Output: $\gcd(a, b) = c_i$. \diamond

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

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

EXAMPLE I.20. 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, $7 = 21 - 14 = 21 - (35 - 21) = 2 \cdot 21 - 35 = 2(56 - 35) - 35 = 2 \cdot 56 - 3 \cdot 35$. \diamond

EXERCISE I.21. 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) $(a, b) = (34, 13)$.

\diamond

We recap what we have learned:

THEOREM I.22. *Given $a, b \in \mathbb{Z}$, there are $x, y \in \mathbb{Z}$ such that $\gcd(a, b)$ can be written as a \mathbb{Z} -linear combination $\gcd(a, b) = ax + by$ of a and b . Moreover, if $g = ax' + by'$ is an arbitrary \mathbb{Z} -linear combination of a and b , then g is always a multiple of $\gcd(a, b)$.* \square

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

DEFINITION I.23. A *unit* of a ring R is a number to which there exists an inverse in the given ring.

Note that divisibility and being a unit is a relative notion: $2|1$ in the ring \mathbb{R} since $1/2 \in \mathbb{R}$, but $2 \nmid 1$ in the ring \mathbb{Z} . Thus, 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.24. The element p in the ring R is *prime* if $p|ab$ always 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.25. 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 $n \in \mathbb{N}$ 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 and hence $q = 1$ and so $n = g$ is the \gcd of a and b , and so in particular divides a .

On the other hand, if $g = 1$ 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 $n | ab$, 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. \square

THEOREM I.26. 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 $2 = 1 \cdot 2$. So we focus on the crank. So let $2 \leq n \in \mathbb{N}$ and assume that the numbers $1, 2, \dots, n$ 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 factored 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 by Theorem I.25 we know 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 $1, 2, \dots, n$ about which we already know that they can all be factored. So, factor $a = 1 \cdot a_1 \cdots a_k$ and $b = 1 \cdot b_1 \cdots b_\ell$ into primes, so that $n + 1 = a \cdot b = 1 \cdot a_1 \cdots a_k \cdot b_1 \cdots b_\ell$ 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 $c_1 \cdot a_1 \cdots a_k = n = c_2 \cdot b_1 \cdots b_\ell$ 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 negative 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 $q_1 = 1$ and $a_1 = b_t$.

Divide out $a_1 = b_t$ to get $c_1 \cdot a_2 \cdots a_k = n/a_1 = c_2 \cdot b_1 \cdots \hat{b}_t \cdots b_\ell$, 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. \square

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

EXAMPLE I.27. 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. \diamond

EXERCISE I.28. (1) On $R = \mathbb{Z}[\sqrt{-5}]$ define a *norm* function

$$N: a + b\sqrt{-5} \mapsto N(a + b\sqrt{-5}) := a^2 + 5b^2 \in \mathbb{Z}.$$

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 $|2|, |3|$ and $|1 \pm \sqrt{-5}|$.

◇

2.3. Some famous theorems and open problems on prime numbers.

The proofs of several theorems in this section are rather beyond us, but if interested you might look at [Sil12] for further pointers.

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

THEOREM I.29. *There are infinitely many prime numbers.*

PROOF. Suppose p_1, \dots, p_k are prime numbers. Then $M = p_1 \cdot \dots \cdot p_k + 1$ 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, $M = c \cdot q_1 \cdot \dots \cdot q_t$ with c a unit and all q_i prime. Since no p_i divides M , none of the q_i is on the list of primes p_1, \dots, p_k . In other words, any finite list of primes is missing at least one other prime. □

Recall now the *harmonic series*

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

EXERCISE I.30. Show that the harmonic series diverges (has no finite value). (Hint: what can you say about each a_i ?). ◇

The previous exercise implies that the sum of the reciprocals of the even numbers is also infinite, and more generally the sum of the inverses of all multiples of some fixed number $k \in \mathbb{N}$ is infinite. On the other hand, we know (from calculus) that the sum of the inverses of all squares, or more generally the sums $\sum_{i=1}^{\infty} \frac{1}{n^t}$, converge for $t > 1$. But suppose you only took the terms that are inverses of prime numbers,

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

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

EXERCISE I.31. Show:

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

◇

THEOREM I.32. *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 prime numbers you get different end products, pun intended.

Consider now the product

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

where p is some prime number.

If you actually multiply this out, the resulting mess contains, for each choice of finitely many powers $p_1^{a_1}, \dots, p_k^{a_k}$ of distinct primes bounded by p , the quantity $1/(p_1^{a_1} \cdots p_k^{a_k})$. 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 $p \rightarrow \infty$, one might try to believe in an equation

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

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, 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 sub-sum 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.33 (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 $k \rightarrow \infty$.*

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

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

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

DEFINITION I.34. For $a, b \in \mathbb{Z}$ we call them *relatively prime* or *coprime* if $\gcd(a, b) = 1$.

The Euler ϕ -function attaches to each natural number n the number of natural numbers less than or equal to 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. (Unless $n = 1$, $\gcd(n, n)$ is of course not 1. The definition is phrased this way to make $\phi(1) = 1$. One can debate whether this is useful).

EXERCISE I.35. (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 $n - n/p$. \diamond

THEOREM I.36. Suppose a, n are natural coprime numbers. 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 by setting $a = 0, n = 1$.

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

CONJECTURE I.37. There are infinitely many twin primes.

Not much is known except that if you looked at the sub-sum 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.38.

- (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.39. 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.31. 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.40 (Riemann hypothesis). Apart from negative even integers, all other complex values s where $\zeta(s) = 0$ satisfy: s has real part $1/2$.

This question 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 20-th century (although lots of mathematics was made that didn't relate 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.41. (1) For any integer n , write $n\mathbb{Z}$ for the set of all integer multiples of n .

(2) 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 $a+n\mathbb{Z} = \{\dots, a-2n, a-n, a, a+n, a+2n, \dots\}$. 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 " $a \bmod n$ ". If the value of n is understood from the context, we may write \bar{a} for $a+n\mathbb{Z}$.

(3) Finally, write $\mathbb{Z}/n\mathbb{Z}$ ("zee modulo enn zee") for the collection of all cosets modulo n .

Here are some explanatory examples. The even integers can be written as $2\mathbb{Z}$, and more generally $n\mathbb{Z}$ stands for $\{\dots, -3n, -2n, -n, 0, n, 2n, \dots\}$.

If $n = 12$, and if $a = 3$, then $\bar{3} = 3 \bmod 12 = 3+12\mathbb{Z} = \{\dots, -21, -9, 3, 15, 27, \dots\}$. This is the set of all times on an absolute clock at which a usual clock shows "3 o'clock".

If $n = 12$, $\mathbb{Z}/n\mathbb{Z}$ is the set of all possible full hours, clustered by what a 12-hour-clock would show.

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

EXAMPLE I.42. Let $n = 4$. there are four cosets modulo 4, namely

$$(I.3.1) \quad 0 + 4\mathbb{Z} = \{\dots, -8, -4, 0, 4, 8, \dots\}, \quad 1 + 4\mathbb{Z} = \{\dots, -7, -3, 1, 5, 9, \dots\},$$

$$(I.3.2) \quad 2 + 4\mathbb{Z} = \{\dots, -6, -2, 2, 6, 10, \dots\}, \quad 3 + 4\mathbb{Z} = \{\dots, -5, -1, 3, 7, 11, \dots\}.$$

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

REMARK I.43. 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. \diamond

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: we set

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

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 $\bar{2}$ and representative $24 = 3 + 3 \cdot 7$ for $\bar{3}$. 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 $\bar{5}$ and $\bar{19}$ 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.44. 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).

\diamond

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 $n \bmod 3$ 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:

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

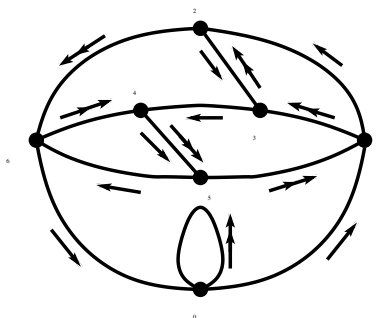
Hence, 1234567 leaves rest 1 when divided by 3.

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

EXERCISE I.45. Prove that 11 divides n if and only if it divides $a_0 - a_1 + a_2 - a_3 + \cdots + (-1)^k a_k$. \diamond

EXAMPLE I.46. Here is a test on divisibility by 7 by way of a picture. Take the digital representation of your number. Start at 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? (To get an idea, inspect/compare what happens when you start with 1 or with 10).

CHAPTER II

Week 2: Groups

1. Symmetries

EXAMPLE II.1. Let T be an equilateral triangle \triangle . We imagine its vertices to carry labels A, B, C , attached counter-clockwise, 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 $0^\circ =: e, 120^\circ =: \ell, 240^\circ =: r$. So, for example ℓ is the motion that moves the A -corner into the old B -position, the B -corner into the old C -position, and the C -corner into the old A -position. (This should not be confused with “write B where there was A , write C where there was B , write A where there was C ”. That is the “inverse” motion!)

Further, 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—explain why!). So $\text{Sym}(\triangle) = \{e, r, \ell, a, b, c\}$.

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 ra):

	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)”. I recommend checking explicitly a few of the product claims that are contained in the table.

NOTATION II.2. Whenever we speak of the collection of symmetries of some object, we denote by e the symmetry that “does nothing”.


DEFINITION II.3. 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 (by multiples of $2\pi/n$, including rotation by 0° .) and n reflections (for $n \geq 3$). Of these reflections, in the n even case, $n/2$ have as symmetry axis the lines that pass through opposite vertices, and $n/2$ that pass through opposite centers of edges. For n odd, vertices lie across centers of edges, and the n corresponding lines are the symmetry axes of the n reflections in this case.

By convention, D_2 is the symmetries of a line segment, which is just $\{e, f\}$ where f is the flip exchanging the ends. And D_1 is the symmetries of a point, so

just $\{e\}$. The two composition tables are $\begin{array}{c|c} & e & f \\ \hline e & e & f \\ f & f & e \end{array}$ and $\begin{array}{c|c} & e \\ \hline e & e \end{array}$.

REMARK II.4. 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 $\begin{array}{c|c} e & f \\ \hline f & e \end{array}$.

EXAMPLE II.5. 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. It is like a clock with 4 hours).

EXAMPLE II.6. Now we look at the symmetries of the letter H. It has 4 elements: the identity e , the rotation \curvearrowright by 180° , the left-right flip \leftrightarrow , and the up-down flip \updownarrow . The table is

e	\leftrightarrow	\updownarrow	\curvearrowright
\leftrightarrow	e	\curvearrowright	\updownarrow
\updownarrow	\curvearrowright	e	\leftrightarrow
\curvearrowright	\updownarrow	\leftrightarrow	e

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

This set of symmetries is called the *Klein 4-group* and denoted KV_4 . Felix Klein was the superstar of symmetry investigations. Note that $\text{Sym}(\text{H}) \subseteq \text{Sym}(\square)$ in a natural way since drawing a box \square around the H does not change the symmetries, but removing the H from \square increases the number of possible symmetries. More generally, if a figure F is contained exactly once in a figure G then we have $\text{Sym}(G) \subseteq \text{Sym}(F)$, and equality happens precisely when each symmetry of F extends to one of G .

EXERCISE II.7. Find an example of a containment of figures $F \subseteq G$ such that $\text{Sym}(G)$ is not contained in $\text{Sym}(F)$. (By the above, you’ll need 2 or more copies of F inside G .)

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 letters in both tables, as it is the unique 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.8. 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, $\cdot : G \times G \rightarrow G$). 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 $g \cdot \tilde{g} = \tilde{g} \cdot g = e$.

REMARK II.9. (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 procedures 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. (Here, c is “dress”, and c^{-1} is “undress” your feet).

(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 sort of behavior 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* $\text{Gl}(n, \mathbb{R})$.

(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 copies. (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.15 below).

THEOREM II.10. *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). \square

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.11. 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) $(\mathbb{R} \setminus \{0\}, *)$, which contains the previous group and uses the same operation, identity and inverse;
- (7) $(\text{Gl}(n, \mathbb{R}), *)$ and $(\text{Gl}(n, \mathbb{C}), *)$ as previously mentioned;

EXAMPLE II.12. 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 $|G| = 3$, $G = \{e, a, b\}$. From the Cancellation Property it follows that $ab \neq ae = a$ and $ab \neq eb = b$. Consequently, $ab = e$ for lack of alternatives. Then we are forced to concede $aa = b$ and $bb = a$ for lack of alternatives, and so $a^3 = e$. So the Cayley 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: $ab = e$. 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, since each column contains each letter once. 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 .

One can proceed similarly if $ac = e$ or if $bc = e$. We thus arrive at the second case, where all non- e elements are inverse to themselves: $aa = bb = cc = e$. Then ab and ba aren't e , and they can be neither a nor b in any case; so they must both be c . Similarly, one finds $cb = bc = a$ and $ca = ac = b$. The resulting multiplication

table is

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

. This is, up to relabeling a to \leftrightarrow , and b to \updownarrow , and c to \curvearrowright ,

the same table as that of KV_4 .

DEFINITION II.13. 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 $G \cong G'$.

If ϕ is the pairing, then the preservation of the Cayley table by ϕ boils down to the identity

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

for all $g_1, g_2 \in G$.

For example, the symmetries $\text{Sym}(S)$ on the letter **S** are isomorphic to the symmetries $\text{Sym}(A)$ on the letter **A** (and also isomorphic to D_2) although the actual motions that carry **S** to **S** (the rotation \curvearrowright) 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 encoded in the table

e	x
x	e

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

A fundamental problem in algebra is to classify all groups up to isomorphism. This means to produce a list of groups such that no two groups on the list are isomorphic, and any group in the world is isomorphic to one of those on the list. Through a monumental effort of many mathematicians spanning several decades, there is now an atlas for all *finite* groups. See https://en.wikipedia.org/wiki/Classification_of_finite_simple_groups.

3. Cyclic and Abelian groups

DEFINITION II.14. A group (G, \cdot) is called *cyclic* if 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.15. *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 $g^i g^j = g^j g^i$ since they are both the product of $i + j$ copies of g . \square

DEFINITION II.16. If in a group (G, \cdot) it is true that $gh = hg$ for all $g, h \in G$ then G is called *Abelian*.

Cyclic groups are Abelian, but lots of groups are not, such as $\text{Sym}(\triangle)$. (This is not cyclic: a, b, c only have two different powers, e and themselves, and ℓ, r only have the three powers e, ℓ, r). And it is not Abelian because $\ell a \neq a\ell$.) Also, $\text{Sym}(H)$ is not cyclic as one sees easily from the squares, but it is Abelian as one sees from the multiplication table.

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

DEFINITION II.17. 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* $\text{ord}(g)$. (There might not be such a k (like for $3 \in (\mathbb{Z}, +)$ for example. We then say $\text{ord}(g) = 0$ or $\text{ord}(g) = \infty$.)

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

Inside $\text{Sym}(\triangle)$, both the powers of ℓ and the powers of a form what we call a subgroup.

DEFINITION II.18. 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.

Note that if we pick a random subset $H \subseteq G$, there is no reason why $H \cdot H$ should be a subset of H . For example, the odd integers are not closed under addition. So, H being closed under the group operation is already special. Even more special is that it must contain the inverse to each of its elements. For example, the positive integers are closed under addition, but do not have their additive inverse ($-n$ for given $n > 0$) also positive. So neither the odd integers nor the positive integers are subgroups of \mathbb{Z} .

On the other hand, as mentioned, $H_1 = \{e, \ell, r\}$ and $H_2 = \{e, a\}$ are subgroups of $\text{Sym}(\triangle)$. The Cayley table of a subgroup is simply the appropriate sub-table of the Cayley table for G .

Note that G counts as a subgroup of G , but the 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.19 (Subgroup criterion).

(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.20. 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, because 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 $\{e\}, \{e, \leftrightarrow\}, \{e, \updownarrow\}, \{e, \curvearrowright\}, KV_4$. 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.21. 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.22. 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 $1 = \text{lcm}(1, 1)$; (e, b) and (e, b^2) have order $3 = \text{lcm}(1, 3)$; (a, e) has order $2 = \text{lcm}(2, 1)$; and (a, b) and (a, b^2) have order $6 = \text{lcm}(2, 3)$.

(We explain the lcm statements: in general one has $\text{ord}(x, y) = \text{lcm}(\text{ord}(x), \text{ord}(y))$. Why? Surely, the lcm is a power that sends (x, y) to (e, e) . We always have $x^{\text{ord}(x)} = e$ and $y^{\text{ord}(y)} = e$. So $y^k = e$ implies $y^{m \cdot \text{ord}(y) - nk}$ for any integer m, n . By the Euclidean algorithm, $y^{\text{gcd}(\text{ord}(y), k)} = e$. Now, the gcd can't be bigger than $\text{ord}(y)$ 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 $\text{ord}(x)$ and $\text{ord}(y)$, while being as small as possible. That is exactly the lcm.

4. Automorphisms

DEFINITION II.23. An *automorphism* of a group G is a bijection $\psi: G \rightarrow G$ that respects the group multiplication:

$$\psi(g \cdot Gg') = \psi(g) \cdot \psi(g')$$

for all $g, g' \in G$; we call this equation the *morphism property*. Let $\text{Aut}(G)$ be the collection of all these automorphisms of G

EXAMPLE II.24. Let C_3 be 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 rotation with right rotation in the rotational symmetries of an equilateral triangle).

On the other hand, the assignment ψ with


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

is not an automorphism, since $\alpha(eb) = \alpha(b) = b$ but $\alpha(e)\alpha(b) = ab = e$.

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. (If G has 60 elements, not so many for a group, then the relabelings are as numerous as the elementary particles in the universe).

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: multiplication of two automorphisms is just doing one relabeling after the other; the inverse of an automorphism is the relabeling done backwards. The relabeling that does nothing is the identity in the automorphism group.

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 $\text{Aut}(C_3)$ is basically the group with table as in Example II.12 part (2).

EXAMPLE II.25. 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 interesting automorphism since we must send e to e (the only element of order one) and b to b (the only element of order two). There is, of course, the (not so interesting) identity on C_4 leaving everything fixed. So, $\text{Aut}(C_4)$ has two elements and as such is the “same” group as $\text{Aut}(C_3)$, both isomorphic to C_2 , sameness in the sense that their Cayley tables are the same after renaming.

EXAMPLE II.26. The automorphisms of KV_4 are more interesting. We worked out the multiplication table of this group in Example II.6. With letter-symbol identifications as in Example II.12, suppose some automorphism fixes 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 $a \mapsto a, b \mapsto c, c \mapsto b$ would be an option. One can check that this relabeling has the morphism property.

By symmetry, there are two more automorphisms that fix either b or c and which interchange the other two non- e variables.

However, we can also try not to fix any non- e variable. Sending $a \mapsto b$ we can try $b \mapsto c$ which forces $c \mapsto a$. Again, one can check that this assignment has the morphism property. And then there is one more, where $a \mapsto c, c \mapsto b, b \mapsto a$.

There can be no other automorphism since e must go to e and the other three elements allow at most $3!$ permutations.

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 $\text{Sym}(\triangle)$. Indeed, if you align ψ_x in $\text{Aut}(KV_4)$ with $x \in \text{Sym}(\triangle)$ then you find this to be an isomorphism (see Definition IV.9 below): it is a one-to-one correspondence between the elements of $\text{Sym}(\triangle)$ 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 ψ_c , and so $\psi_\ell\psi_a = \psi_c$. If we compare to the Cayley table of $\text{Sym}(\triangle)$ then we also have correspondingly $\ell a = c$. Checking the entire list of products, we see $\text{Aut}(KV_4) = \text{Sym}(\triangle)$. \diamond

5. Free groups

DEFINITION II.27. A group is *free* (on the elements g_1, \dots, g_k) if, for some $k \in \mathbb{N}$, it is isomorphic to the group F_k of all words in the letter set $L_k = \{e, x_1, \dots, x_k, y_1, \dots, y_k\}$ 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 and simplifying (things such as gg^{-1} to e , or eg to g).

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, $F_1 = \{\dots, y_1^2, y_1, e, x_1, x_1^2, \dots\}$ is isomorphic to the Abelian group $(\mathbb{Z}, +)$ via the identification $x_1^k \leftrightarrow k \in \mathbb{Z}, y_1^k \leftrightarrow -k \in \mathbb{Z}$.

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 isomorphic to F_3, F_4, \dots . 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.28. 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.29. Let $G = \mathbb{Z} \times \mathbb{Z}$, generated by $x_1 := (1, 0), x_2 := (0, 1)$, which then forces $y_1 = (-1, 0), y_2 = (0, -1)$. 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 $y_1 y_2 = y_2 y_1$, since G is Abelian..

It is a general fact that one can view any group as “a free group on the elements of G , with additional rules” (and these rules exactly portray the Cayley table of G).

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 the group $\mathbb{Z}/n\mathbb{Z}$ is also n .

All groups $\mathbb{Z}/n\mathbb{Z}$ are Abelian, because \mathbb{Z} is Abelian and we just install additional rules in order to make $\mathbb{Z}/n\mathbb{Z}$ from \mathbb{Z} (namely, the rule $n = 0$).

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 $G = \mathbb{Z}/12\mathbb{Z}$.

(1) Let us sort elements by order.

(a) $\text{ord}(1 + 12\mathbb{Z}) = \text{ord}(5 + 12\mathbb{Z}) = \text{ord}(7 + 12\mathbb{Z}) = \text{ord}(11 + 12\mathbb{Z}) = 12$ since the first time that a sum $\underbrace{(1 + 12\mathbb{Z}) + \dots + (1 + 12\mathbb{Z})}_{k \text{ copies}}$ is the coset $0 + 12\mathbb{Z}$ is for $k = 12$.

A similar argument applies to $5 + 12\mathbb{Z}$, $7 + 12\mathbb{Z}$ and $11 + 12\mathbb{Z}$.

(b) $\text{ord}(2 + 12\mathbb{Z}) = \text{ord}(10 + 12\mathbb{Z}) = 6$.

(c) $\text{ord}(3 + 12\mathbb{Z}) = \text{ord}(9 + 12\mathbb{Z}) = 4$.

(d) $\text{ord}(4 + 12\mathbb{Z}) = \text{ord}(8 + 12\mathbb{Z}) = 3$.

(e) $\text{ord}(6 + 12\mathbb{Z}) = 2$.

(2) This time 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 $\bar{7}$ are $\{\bar{7}, \bar{2}, \bar{9}, \bar{4}, \bar{11}, \bar{6}, \bar{1}, \bar{8}, \bar{3}, \bar{10}, \bar{5}, \bar{0}\}$ in that sequence.
- $2 + 12\mathbb{Z}, 10 + 12\mathbb{Z}$ 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.
- $0 + 12\mathbb{Z}$ generates 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 have grouped in the same item the elements that generate the same group).

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. In any group G , if $\text{ord}(g) = n > 0$ 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 $n \mid (i - j)$.

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 $\text{gcd}(n, i)$ 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, $\text{gcd}(n, i) = n$ and so $n \mid i$.

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

DEFINITION III.3. If $g \in G$ we write $\langle g \rangle$ for the group of all 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 $i + n\mathbb{Z}$ 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 $\underbrace{(1 + n\mathbb{Z}) + \dots + (1 + n\mathbb{Z})}_{n \text{ copies}} = 0 + n\mathbb{Z}$. \square

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 g^k has order $n/\text{gcd}(n, k)$. Moreover, the group $\langle g^k \rangle$ generated by g^k is the same group as the group $\langle g^{\text{gcd}(n, k)} \rangle$ that is generated by $g^{\text{gcd}(n, k)}$. Moreover, abstractly this group is the same as the cyclic group $\mathbb{Z}/m\mathbb{Z}$ where $m = n/\text{gcd}(n, k)$.

PROOF. By the proof of Lemma III.2, $\langle g^k \rangle$ contains $g^{\text{gcd}(n, k)}$, and so also all its powers. Conversely, $\text{gcd}(n, k)$ divides k and so of course $\langle g^{\text{gcd}(n, k)} \rangle$ contains g^k and all its powers. So, the groups $\langle g^k \rangle$ and $\langle g^{\text{gcd}(n, k)} \rangle$ are contained one in the other in both directions and hence equal.

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

We can now complete a table from above on subgroups of $\mathbb{Z}/12\mathbb{Z}$, see Example III.1. So we have $n = 12$, $g = k + 12\mathbb{Z}$ the k -th power of the generator $1 + 12\mathbb{Z}$.

$g := k \bmod 12\mathbb{Z}$	size of $\langle g \rangle$	$\text{gcd}(n, k)$	$n/\text{gcd}(k, n) = \text{ord}(k + 12\mathbb{Z})$
1, 5, 7, 11	12	1	12
2, 10	6	2	6
3, 9	4	3	4
4, 8	3	4	3
6	2	6	2
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 $\bar{1}, \bar{5}, \bar{7}, \bar{11}$. These are the numbers that are *coprime* to 12. (It is a good moment to realize here that any representative in the coset \bar{k} is coprime to n when k is coprime to n . For example, $5 + 127 \cdot 12$ lives in the same coset as 5, and so from $\gcd(5, 12) = 1$ follows that $\gcd(5 + 127 \cdot 12, 12) = 1$).

The magic therefore lies in coprimeness.

DEFINITION III.6. For $n \in \mathbb{Z}$ let $\phi(n)$ be the *Euler ϕ -function* 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 $\text{ord}(g^k) = n$. But $\text{ord}(g^k) = n / \gcd(n, k)$ by Lemma III.2, 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 made of numbers coprime to n . \square

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 $\langle g^j \rangle$. Exchanging i, j gives the reverse containment, hence an equality.

We have now seen almost all parts of

THEOREM III.8. Let g be an element of order n . So $\langle g \rangle$ is the cyclic group $\mathbb{Z}/n\mathbb{Z}$ up to relabeling.

- (1) Subgroups of cyclic groups are always cyclic.
- (2) For all $i \in \mathbb{Z}$, $\text{ord}(g^i)$ 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 comprised exactly the set of (n/k) -th powers $\langle g^{n/k} \rangle$ inside $\langle g \rangle$.
- (4) If $k|n$ then the number of elements of order k inside $\langle g \rangle$ is equal to $\phi(k)$.
If $k \nmid n$, no elements have order k .
- (5) We have $n = \sum_{d|n} \phi(d)$.

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

$$12 = \underbrace{|\{\bar{1}, \bar{5}, \bar{7}, \bar{11}\}|}_{4=\phi(12)} + \underbrace{|\{\bar{2}, \bar{10}\}|}_{2=\phi(6)} + \underbrace{|\{\bar{3}, \bar{9}\}|}_{2=\phi(4)} + \underbrace{|\{\bar{4}, \bar{8}\}|}_{2=\phi(3)} + \underbrace{|\{\bar{6}\}|}_{1=\phi(2)} + \underbrace{|\{\bar{0}\}|}_{1=\phi(1)}.$$

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. $G := \mathbb{Z}/n_1\mathbb{Z} \times \mathbb{Z}/n_2\mathbb{Z} \times \dots \times \mathbb{Z}/n_k\mathbb{Z}$ 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 $g = (\bar{1}, \dots, \bar{1})$. 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 $\text{ord}(g) = n_1 \cdots n_k$ and so g generates G .

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

In particular, this says that a product $(\mathbb{Z}/(p_1)^{e_1}\mathbb{Z}) \times (\mathbb{Z}/(p_2)^{e_2}\mathbb{Z}) \times \dots \times (\mathbb{Z}/(p_k)^{e_k}\mathbb{Z})$ for *distinct* primes $p_1 < p_2 < \dots < p_k$ is cyclic. Note that our first example in this section showed that distinctness is crucial.

REMARK III.11. If $m, n \in \mathbb{N}$ and $m|n$ then there is a function $\pi: \mathbb{Z}/n\mathbb{Z} \rightarrow \mathbb{Z}/m\mathbb{Z}$ given by $(a + n\mathbb{Z}) \mapsto (a + m\mathbb{Z})$. That this is really a function follows from "if $n|(a - a')$ then $m|(a - a')$ ". In other words, our function does not destroy cosets. This assignment also has the morphism property $\pi((a + n\mathbb{Z}) +_n (a' + n\mathbb{Z})) = \pi(a + n\mathbb{Z}) +_m \pi(a' + n\mathbb{Z})$ where $+_n, +_m$ are addition in $\mathbb{Z}/n\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z}$ respectively; this follows from basic rules of arithmetic.

EXAMPLE III.12. 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 $3 + 35\mathbb{Z}, 8 + 35\mathbb{Z}, 13 + 35\mathbb{Z}, 18 + 35\mathbb{Z}, 23 + 35\mathbb{Z}, 28 + 35\mathbb{Z}, 33 + 35\mathbb{Z}$ in $\mathbb{Z}/35\mathbb{Z}$ 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 in $\mathbb{Z}/25\mathbb{Z}$, but in $\mathbb{Z}/6\mathbb{Z}$ they do not. Destroying cosets is not 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. (By Lemma III.10 $(\mathbb{Z}/5\mathbb{Z}) \times (\mathbb{Z}/7\mathbb{Z})$ is cyclic. But $\mathbb{Z}/35\mathbb{Z}$ is also cyclic, of the same size, and hence they must be isomorphic groups). 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.13. *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 $1 = mx + ny$. \square

REMARK III.14. 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.

EXAMPLE III.15. Here, we establish the equation $\mathbb{Z}/6\mathbb{Z} = (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$. To go from $\mathbb{Z}/6\mathbb{Z}$ to the product, associate to $c + 6\mathbb{Z}$ the element $(c + 2\mathbb{Z}, c + 3\mathbb{Z})$ in the product. To go back, find a Euclidean equation for 2 and 3. For example, $\gcd(2, 3) = 1 = (-1) \cdot 2 + (1) \cdot 3$. In other words, $x = -1, y = 1$. Then $a \cot y \cdot 3 + b \cdot x \cdot 2 = 3a - 2b$. So, the element $c + 6\mathbb{Z}$ that maps to $(a + 2\mathbb{Z}, b + 3\mathbb{Z})$ should be $(3a - 2b) + 6\mathbb{Z}$.

And indeed, $(3a - 2b) + 2\mathbb{Z} = 3a + 2\mathbb{Z} = a + 2\mathbb{Z}$, and $(3a - 2b) + 3\mathbb{Z} = -2b + 3\mathbb{Z} = b + 3\mathbb{Z}$.

3. $U(n)$: Automorphisms of $\mathbb{Z}/n\mathbb{Z}$ and the Euler ϕ function

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 $\underbrace{(1 + \dots + 1 + 5\mathbb{Z})}_{a \text{ copies}}$ must correspond to $\underbrace{(3 + \dots + 3 + 5\mathbb{Z})}_{a \text{ copies}}$.

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.16. 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 inside $U(n)$ is an operation $\cdot : U(n) \times U(n) \rightarrow U(n)$ (the above gcd considerations show that a product of things coprime to n is again coprime to n) that is associative (because multiplication of integers is associative) and there is an identity element for this multiplication process (namely the coset $1 + n\mathbb{Z}$). An interesting observation 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; see Theorem I.22) and also that $(a + n\mathbb{Z}) \cdot (x + n\mathbb{Z}) = (ax + n\mathbb{Z}) = ((1 - ny) + \mathbb{Z}) = 1 + n\mathbb{Z}$. So $x + n\mathbb{Z}$ is an inverse for $a + n\mathbb{Z}$.

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

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

You can think of making $U(n)$ from $\mathbb{Z}/n\mathbb{Z}$ by asking “if I want to make a multiplication 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 $(a + n\mathbb{Z})(b + n\mathbb{Z}) = 0 + n\mathbb{Z} = (0 + n\mathbb{Z})(b + n\mathbb{Z})$ would contradict the cancellation property. So all $a + n\mathbb{Z}$ with $\gcd(a, n) > 1$ must also be kicked out. Left over are the ones with $\gcd(n, a) = 1$.

Here are some examples.

EXAMPLE III.17. (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 isomorphic to 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, exactly $p - 1$ 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 course $U(p)$ has $p - 1$ elements. 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 $U(12)$ 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.18. For example, $U(6)$ is $\mathbb{Z}/2\mathbb{Z}$. Let's try to understand that. Recall from Example III.15 that we proved that there is an assignment $\psi: \mathbb{Z}/6\mathbb{Z} \rightarrow \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ that sends the coset $a + 6\mathbb{Z}$ to the coset pair $(a + 2\mathbb{Z}, a + 3\mathbb{Z})$ and that this map respects addition, and that it is bijective. It must then also respect multiplication since in all cases involved multiplication is inherited from multiplication in \mathbb{Z} , which is based on iterated addition. 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.

The assignments $c + 6\mathbb{Z} \mapsto (c + 2\mathbb{Z}, c + 3\mathbb{Z})$ and $(a + 2\mathbb{Z}, b + 3\mathbb{Z}) \mapsto (3a - 2b) + 6\mathbb{Z}$ are inverses that have the morphism property. If c is coprime to 6 then it is also coprime to 2 and 3. So, $U(6)$ is mapped to elements of $U(2) \times U(3)$ here. Conversely, if c has a common factor with 6, then this factor is 2 or 3, and so c cannot be coprime to both 2 and 3. In other words, the assignments preserve “being in U ”. So $U(6) = U(2) \times U(3)$.

Explicitly, this correspondence relates $1 + 6\mathbb{Z}$ with $(1 + 2\mathbb{Z}, 1 + 3\mathbb{Z})$ 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.19. *If m, n have $\gcd(m, n) = 1$ then $U(mn) = U(m) \times U(n)$.*

This needs coprimeness!!! For example, $U(4)$ has two elements, and that cannot be the square of the number of elements of $U(2)$, even though $2 \cdot 2 = 4$.

More generally, if you have 3 or more coprime numbers, one gets a corresponding product result on products of unit groups.

EXAMPLE III.20. How many elements does $U(750)$ have?

The bad way is to write them all out. The enlightened MA453 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.21. Recall the Euler ϕ -function that counts for $n \in \mathbb{N}$ how many numbers from $1, \dots, n$ 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, to be a generator of $\mathbb{Z}/n\mathbb{Z}$ means that 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 the Euler ϕ -function satisfies

$$\phi(mn) = \phi(m)\phi(n)$$

because of the theorem above.

Warning: this needs coprimeness!!! (Check that $\phi(4) \neq \phi(2) \times \phi(2)$, for example).

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);
- $[a \simeq b \text{ and } b \simeq c] \Rightarrow [a \simeq c]$ 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 modulo calculation (this is really the relation $i \simeq j$ on \mathbb{Z} whenever $n|(i - j)$).

An example of a relation 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, \dots , 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 G_n be the collection of all group elements $a \in G$ that have order exactly n . Then $|G_n|$ is a multiple of $\phi(n)$.*

PROOF. If $g \in G_n$ is of order n , then $\langle g \rangle$ is a cyclic group of order n . By last week's results, $\langle g \rangle$ contains exactly $\phi(n)$ elements whose order is exactly n , and these $\phi(n)$ things of order n are exactly the various generators of $\langle g \rangle$.

Now make an equivalence relation on G_n where $x \simeq y$ if and only if $\langle x \rangle = \langle y \rangle$. Since x and y are both generators of the group they both generate, our equivalence relation puts into one basket exactly those elements that satisfy $x \in \langle y \rangle$ AND $y \in \langle x \rangle$. In particular, this is indeed an equivalence relation on G_n .

Each equivalence class has size $\phi(n)$ where n is the size of the group $\langle x \rangle$, two equivalence classes either are equal or do not meet, and the union of all equivalence classes is G_n . So $\phi(n) \cdot (\# \text{ of cosets}) = |G_n|$. \square

In the special case where G is cyclic and $n = |G|$, then $G = \langle g \rangle$ and $|A| = \phi(n)$ by last week.

EXAMPLE IV.4. Consider D_6 made of 6 rotations and 6 reflections. Of course, the identity has order 1 and is the only one of that type, and $\phi(1) = 1$ divides

1. The 6 reflections all have order 2, and so does the rotation by 180; note that $\phi(2) = 1$ divides 7. There are two elements of order 3, namely the rotations by 120 and 240; as one should expect, $\phi(3) = 2$ divides 2. There are no elements of orders 4 or 5. Two rotations (by 60 and 300 respectively) have order 6, and $2 = \phi(6)$ indeed divides 2.

EXAMPLE IV.5. Let G be the physical symmetries of a regular octahedron. (By “physical” we mean the ones that you could actually carry out with a wooden octahedron; it does not allow reflections as reflections “turn inside out”).

How big is G ? The following count is instructive. Let’s say the vertices of the octahedron are labeled a, \dots, f . Vertex a needs to be moved to some vertex, so there are 6 choices. Having done that, there are 4 possibilities in each case, because whatever symmetry we have we can follow it by a rotation by 0, 90, 180 or 270 about the main axis on which vertex a lies now. (Recall that we allow no reflections). So $|G| = 6 \cdot 4 = 24$.

One element of order 1 exists, the identity, and $\phi(1) = 1$ divides 1. There are 3 rotations by 180 that spin around one of the 3 main axes, and then there are $6=12/2$ more 180 rotations that rotate about the line that links the centers of 2 opposite edges. And $\phi(2) = 1$ divides 9. There are 4 pairs of opposite triangles, and we can rotate about the 4 axes linking the midpoints of these opposite triangles by 120 or 240; these 8 symmetries all have order 3 and $\phi(3) = 2$ divides 8. There are $6 = 2 \times 3$ rotations by ± 90 around the 3 main axes; these symmetries have order 4 and $\phi(4) = 2$ divides 6. There are no elements of other orders since we already have 24 elements listed in orders 1, 2, 3 or 4.

2. Morphisms

Let G, G' be two groups.

DEFINITION IV.6. A *morphism* (or *homomorphism*) is any function $\psi: G \rightarrow G'$ from one group to another that respects the group operations:

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

for all $g_1, g_2 \in G$. (The subscript for the dot indicates in which group multiplication is happening)

You have seen many examples already.

DEFINITION IV.7. 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: $42(a + b) = (42a) + (42b)$.
- $G = GL(2, \mathbb{R})$ = the real 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 $(\mathbb{R}, +) \rightarrow (\mathbb{R}^\times, \cdot)$, since $\exp(x + y) = \exp(x) \cdot \exp(y)$.
- The logarithm function $\ln: (\mathbb{R}_{>0}, \cdot) \rightarrow (\mathbb{R}, +)$, since $\ln(a \cdot b) = \ln(a) + \ln(b)$.
- The square root function $\sqrt{\cdot}: (\mathbb{R}_{>0}, \cdot) \rightarrow (\mathbb{R}_{>0}, \cdot)$, since $\sqrt{a \cdot b} = \sqrt{a} \cdot \sqrt{b}$.
- The third power map $(-)^3: (\mathbb{R}^\times, \cdot) \rightarrow (\mathbb{R}^\times, \cdot)$, since $(a \cdot b)^3 = a^3 \cdot b^3$.
- The third power map $(-)^3: U(7) \rightarrow U(7)$ since $((a + 7\mathbb{Z})(b + 7\mathbb{Z}))^3 = (a + 7\mathbb{Z})^3 \cdot (b + 7\mathbb{Z})^3$.

EXAMPLE IV.8. Suppose we want to make a morphism $k: \mathbb{Z}/m\mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z}$ 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 be a multiple of n . Suppose $n|km$, say $sn = km$. Then $(a + tn)$ is sent by k to $ka + ktm = ka + tsn$. So, all elements of the form $a + m\mathbb{Z}$ are sent to $ka + km\mathbb{Z} \subseteq ka + \mathbb{Z}n$. We conclude:

$$(IV.2.1) \quad [\text{Multiplication } k: \mathbb{Z}/m\mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z} \text{ is a morphism}] \iff [n \text{ divides } mk].$$

Recall now:

DEFINITION IV.9. A morphism ψ is a *isomorphism* if it is bijective (= injective + surjective = into + onto). It is an *automorphism* if it is an *isomorphism* where $G' = G$.

REMARK IV.10. Let $\phi: G \rightarrow G'$ be an isomorphism.

(1) Then $\psi(e_G) = e_{G'}$, 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 $e_{G'}$.

(2) We have $\text{ord}_G(g) = k$ if and only if $\text{ord}_{G'}(\phi(g)) = k$. (For any function that satisfies the morphism law, $g^k = e_G$ implies $(\phi(g))^k = e_{G'}$. In particular, $\text{ord}_G(g)$ is an upper bound for $\text{ord}_{G'}(\phi(g))$. But for bijective maps, if ψ is the inverse map from G' to G , one can check that it also satisfies the morphism law. And so one also has $\text{ord}_G(g) \leq \text{ord}_{G'}(\phi(g))$).

So, isomorphisms (and thus also automorphisms) send elements of order k to elements of order k . One can use this property to simplify searches for isomorphisms.

An automorphism is a relabeling of G that preserves the group structure. An isomorphism is a way of linking in twin 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, $\exp: (\mathbb{R}, +) \rightarrow (\mathbb{R}^\times, \cdot)$ is an isomorphism, but not an automorphism. The map that 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.

- $|G| = |G'|$ since ϕ is bijective;
- both G and G' are cyclic, or neither one is cyclic (and generators are sent to generators);
- 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.11. If G, G' are both cyclic of the same order n then they are isomorphic. Namely, if $G = \langle g \rangle$ and $G' = \langle g' \rangle$, make a morphism sending g^i to $(g')^i$ for all i . Checking that this satisfies the morphism law is easy. It is clearly bijective.

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

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

EXAMPLE IV.12. While $U(12)$ and $U(10)$ both have 4 elements, they are not isomorphic. Indeed, $U(10)$ is cyclic generated by $3 + 10\mathbb{Z}$ (check that!), but $U(12)$ has no element of order 4 and so cannot be cyclic. We also saw last week: $U(10) = U(2) \times U(5) = U(5) = \mathbb{Z}/4\mathbb{Z}$, while $U(12) = U(4) \times U(3) = \{1 + 4\mathbb{Z}, 3 + 4\mathbb{Z}\} \times \{1 + 3\mathbb{Z}, 2 + 3\mathbb{Z}\}$.

One natural way of making automorphisms is the following:

DEFINITION IV.13. Let $a \in G$ be a group element. Define a map $\psi_a: G \rightarrow G$ by setting

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

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

REMARK IV.14. (1) Note first that this indeed respects multiplication: $\psi_a(g_1)\psi_a(g_2) = ag_1a^{-1}ag_2a^{-1} = ag_1g_2a^{-1} = \psi_a(g_1g_2)$.

(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.

(3) The composition morphism $\psi_a \circ \psi_{a^{-1}}$ sends g to $a(a^{-1}g(a^{-1})^{-1})a^{-1} = g$ and is therefore the trivial automorphism. It follows that ψ_a and $\psi_{a^{-1}}$ are inverse to one another in the automorphism group.

(4) More generally, if ψ_x, ψ_y are inner automorphisms of G , they can be composed: $\psi_x(\psi_y(g)) = xygy^{-1}x^{-1} = \psi_{xy}(g)$.

EXAMPLE IV.15. If $G = \text{Sym}(\triangle)$, then the 6 inner automorphisms are as follows:

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

In the remark above we basically proved:

LEMMA IV.16. *The assignment $a \mapsto \psi_a$ is a morphism inn_G from the group G to the group of its inner automorphisms $\text{Inn}(G)$.*

REMARK IV.17. Since $\text{Inn}(\text{Sym}(\triangle))$ has 6 different elements just like $\text{Sym}(\triangle)$ itself, then the conclusion is that inn_G is actually an isomorphism, so that as abstract groups there is no difference between $\text{Sym}(\triangle)$ and $\text{Inn}(\text{Sym}(\triangle))$. This is pretty unusual. For many groups, the inner automorphisms are different from the set of all automorphisms, and resemble the original group not at all. ($\text{Aut}(G)$ can be much simpler than G , but also much more complicated).

3. Cosets for subgroups and Lagrange

We now take a group G with subgroup H and work towards a construction that looks like $\mathbb{Z}/n\mathbb{Z}$ constructed from the pair $\mathbb{Z} = G, n\mathbb{Z} = H$.

DEFINITION IV.18. 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.19. Let $G = \text{Sym}(\triangle)$ and $H = \{e, a\}$. Then $e \cdot H = a \cdot H = H = \{e, a\}$, $r \cdot H = b \cdot H = \{b, r\}$, 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.20. *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 if $a \in H$ (since $ah = h'$ gives $a = h'h^{-1} \in H$).
- (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 $h \rightarrow ah$ is injective, and so is the map $ah \rightarrow ah h^{-1} = a$).
- (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.21. *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 $|H|$ by the lemma. \square

REMARK IV.22. In the situation of the theorem, even if left and right cosets of G by H do not agree, their number is the same, given by $|G|/|H|$.

DEFINITION IV.23. The quotient $|G|/|H|$ from the theorem is denoted $[G : H]$ and called the *index* of H in G .

COROLLARY IV.24 (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 $\text{ord}(g)$ elements and by the theorem this number divides $|G|$. \square

COROLLARY IV.25. *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 $\text{ord}(g) = |G|$ or $\text{ord}(g) = 1$. In the second case, $g = e$ must be the identity. So, take an element g that is not the identity. Now $\text{ord}(g)$ cannot be 1, so it must be $|G|$. 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 . \square

THEOREM IV.26 (Fermat's little theorem). *If p is a prime number then $a^{p-1} - 1$ is divisible by p for all $a \in \mathbb{Z}$ that are coprime to p . In particular, $a^p + p\mathbb{Z} = a + p\mathbb{Z}$ for all $a \in \mathbb{N}$.*

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 $a \in \mathbb{Z}$ coprime to p . The second statement follows from multiplication by a in the case $\gcd(a, p) = 1$, and it is obviously correct also if $p|a$. \square

Note that p prime is essential: Fermat's little theorem fails for $p = 4$ (as $2^4 - 4$ is not divisible by 4). Question: are there non-primes for which Fermat's little theorem works? We will return to this issue...

4. Kernels and normal subgroups

Recall (from homework) the concept of conjugation: if $a \in G$ then the conjugation process takes any $g \in G$ to aga^{-1} . We now define:

DEFINITION IV.27. If H is a subgroup of G , the *conjugate* of H with respect to a is the set of conjugates of elements of H

$$aHa^{-1} := \{aga^{-1} \mid g \in H\}.$$

For example, if $G = \text{Sym}(\triangle)$ then $H = \{1, a\}$ has conjugate subgroups H , $\{1, b\}$ and $\{1, c\}$. (This follows with the help of the multiplication table).

REMARK IV.28. (1) We recall that $aga^{-1}ag'a^{-1} = agg'a^{-1}$ and so products of conjugates are conjugates. In particular, a conjugates e_G to e_G and the inverse of aga^{-1} is $ag^{-1}a^{-1}$.

(2) We further recall that conjugation by a provides an inner automorphism ψ_a of G and that implies that aHa^{-1} is a subgroup of G .

(3) If two subgroups H, H' are conjugate with $aHa^{-1} = H'$, they are isomorphic to one another via the assignment $h \mapsto aha^{-1}$, and in particular have the same size, the same number of generators, and for any fixed k the same number of elements of order k .

(4) If G is Abelian, $aha^{-1} = h$ for all $a, h \in G$. Thus, in any Abelian group G any subgroup h is fixed element by element by every inner automorphism. So, any subgroup H has only itself as conjugate.

DEFINITION IV.29. Suppose $\phi: G \rightarrow G'$ is a morphism. Set $\ker(\phi) = \{g \in G \mid \phi(g) = e_{G'}\}$ and call it the *kernel* of ϕ .

THEOREM IV.30. *For any morphism $\phi: G \rightarrow G'$, the kernel $\ker(\phi)$ is a subgroup of G . Moreover, $\ker(\phi)$ is stable under the inner automorphism ψ_a for all $a \in G$.*

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 $\phi(g_1 \cdot_G g_2) = \phi(g_1) \cdot_{G'} \phi(g_2) = e_{G'} \cdot_{G'} e_{G'} = e_{G'}$. So, $g_1 \cdot_G g_2$ is also in $\ker(\phi)$.

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 some $x \in G$. All elements of xHx^{-1} have the form xgx^{-1} with $g \in \ker(\phi)$. Then we need to show that xgx^{-1} is also in the kernel of ϕ . So we test it:

$$\begin{aligned} \phi(x \cdot_G g \cdot_G x^{-1}) &= \phi(x) \cdot_{G'} \phi(g) \cdot_{G'} \phi(x^{-1}) \\ &= \phi(x) \cdot_{G'} e_{G'} \cdot_{G'} \phi(x^{-1}) \\ &= \phi(x) \cdot_{G'} \phi(x^{-1}) \\ &= \phi(x \cdot_G x^{-1}) \\ &= \phi(e_G) = e_{G'}. \end{aligned}$$

So, indeed we have $xgx^{-1} \in H$ and H is stable under any conjugation. \square

DEFINITION IV.31. If $H \subseteq G$ is a subgroup that is stable under all inner automorphisms, we call H a *normal subgroup*.

REMARK IV.32. (1) Being normal in this technical sense 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. (The word “normal” gets used very often in mathematics to describe an unnormal situation. For example, normal vectors are perpendicular, and that is far from common).

(2) 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.

(3) The trivial subgroup $\{e_G\}$ is always normal since $aea^{-1} = e$.

(4) The other stupid subgroup of G , all of G , is also always normal. Indeed, aG is always G since for all $g \in G$ we also have $a^{-1}g \in G$ and so aG contains $aa^{-1}g = g$. For similar reasons, $Ga^{-1} = G$, and so $aGa^{-1} = G$ as well. Normality is only interesting when H is neither G nor $\{e_G\}$.

EXAMPLE IV.33. The kernel of any morphism is normal as we proved in the theorem above.

EXAMPLE IV.34. The subgroup $\{e, a\} \subseteq \text{Sym}(\triangle)$ is not normal. Indeed, $a \in H$ but $\ell a \ell^{-1} = c \ell^{-1} = cr = b$ is not in H . (We already noted above that $\{e, a\}$ has two conjugate subgroups, $\{e, b\}$ and $\{e, c\}$).

One can check that in $\text{Sym}(\triangle)$ the only ones stable under all conjugations are the trivial group $\{e\}$, the rotation subgroup $\{e, \ell, r\}$, and the whole group.

EXAMPLE IV.35. Let $G = \text{Gl}(2, \mathbb{R})$ be the 2×2 invertible matrices with real entries, with matrix multiplication as group operation. Let $\phi: G \rightarrow \mathbb{R}^\times$ 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. It follows that the *special linear group* $\mathrm{Sl}(2, \mathbb{R}) := \{g \in \mathrm{Gl}(2, \mathbb{R}) \mid \det(g) = 1\}$ is a normal subgroup.

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 (to us as aficionados of the abstract group structure) what the permuted n things are. We usually assume they are the numbers $1, \dots, n$.

Note that S_n has $n!$ elements.

EXAMPLE V.2. We have met S_3 as $\text{Sym}(\triangle)$. 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

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

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 what used to be in the bucket labeled by the top row is going to the bucket labeled by the bottom row. So, the content of bucket B (below) is replaced by the content of A (above it) and so on.

Suppose we compose the moves r and a to form ra .

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

It takes some practice to read this product correctly. The important bit is that one carries it out *right to left*, like when applying functions to an argument. So, if you want to know what this product does to the bucket of letter B you first check what the right factor $\begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix}$ does to bucket B . And you find, it sends its content

to bucket C . Next, you ask what the left factor $\begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix}$ does to stuff in the C bucket, so you look at the column labeled C on top. And it says that stuff in bucket C is being moved to bucket B , since under the C is a B . So, combining both steps, bucket B content first moves to bucket C and then back to bucket B . So in the product one should have B above B , which is exactly right.

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

DEFINITION V.3 (Cycle notation). There is another way to write permutations called *cycle notation*. You start by opening a parenthesis and choose any letter, for example A : we now have “(A ”. Next, record where stuff from bucket 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 stuff from bucket C goes, and under r that is towards B . So we continue to “(A, C, B ”. But stuff from bucket B now is moved to bucket A and that “closes the cycle”, so we write “(A, B, C)”.

If a cycle closes before you have written 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 and write the result as another cycle? 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 bucket 1. Under $(3, 5)$ it goes to position 1 (because 1 does not show up in the cycle $(3, 5)$), then bucket 1 content 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 at all, so position 2 is the final destination of 1. So we start writing the product as “(1, 2”.

Next we redo this all with bucket 2. Under $(3, 5)$, stuff in bucket 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, stuff in bucket 2 moves to bucket 3. So we are now at “(1, 2, 3”.

Restart with bucket 3. Under $(3, 5)$, 3 moves to bucket 5, and under $(2, 3, 4, 1)$ bucket 5 stays put. Then at the end $(1, 4, 5)$ move 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 answer are such that no number occurs in more than one cycle. Disjoint cycles are preferable since we “understand” better.
- For example, the order of any cycle (in the group theory sense) is its own length: if you rotate a bunch of k people on a round table by one seat, 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 the 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. So 6 iterations is the least number of interactions that brings everyone back to his own seat.

To illustrate usefulness of disjoint cycles, $(1,2,3)(2,3,4)=(1,2)(3,4)$ has order 2, not 3 (the lack of disjointness messes with things!)

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 writable as $(1,2)(1,2)$.)

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

LEMMA V.6. *If you take a permutation σ and write it as product of transpositions, then (although the number of transpositions is not determined by σ) the residue class $k + 2\mathbb{Z}$ of the number k of transpositions is determined by σ . In other words, if the number of transpositions is odd (resp. even) for σ for some way of writing σ in terms of transpositions, then it is odd (resp. even) for all such ways.*

Before we embark on a proof, one more concept:

DEFINITION V.7. Let σ be a permutation of $1, \dots, n$. 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 words, if you write $\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ \sigma_1 & \sigma_2 & \dots & \sigma_n \end{pmatrix}$ 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 Lemma V.6: If σ can be written in two ways as product of transpositions,

$$t_1 t_2 \cdots t_l = \sigma = t'_1 \cdot t'_2 \cdots t'_{k'}$$

then multiplying the left hand side by the inverse of the right hand side we arrive at a way to write $\sigma \cdot \sigma^{-1} = e$ as product of transpositions. What we want to show is that $k - k'$ is even, which is the same as proving that $k + k'$ is even. So we shall show that e cannot be written as a product of an odd number of transpositions. In order to do so we will show that if some permutation σ is composed with a transposition then its parity changes.

Imagine, for $1 \leq i < j \leq n$, composing the transposition (i, j) with the permutation $\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ \sigma_1 & \sigma_2 & \dots & \sigma_n \end{pmatrix}$ and we count the change in the disorder.

Let's say the output of σ is the sequence

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

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

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

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 with 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 does not change either.

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

- (1) If $s_k < s_i$ and $s_k < s_j$ then $[i, k]$ is a switch in σ but $[k, j]$ is not. On the other hand, $[k, j]$ is a switch in $(i, j)\sigma$ while $[i, k]$ is not.
- (2) If $s_k < s_i$ and $s_k > s_j$ then $[i, k]$ and $[k, j]$ are both a switch in σ , but neither of $[i, k]$ or $[k, j]$ is a switch in $(i, j)\sigma$.
- (3) If $s_k > s_i$ and $s_k < s_j$ then neither of $[i, k], [k, j]$ is a switch in σ and both $[i, k], [k, j]$ 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 but $[i, k]$ is not, while $[i, k]$ is a switch in $(i, j)\sigma$ but $[k, j]$ is not.

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

Finally, consider the pair i, j . If it is not a switch 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 taken modulo 2 must agree with the parity of σ . So, the number of transpositions used is even if and only if the parity of σ is even. And since parity is defined without any recourse to transpositions, it does not depend on how you write σ as such product! \square

DEFINITION V.8. Let A_n be the *alternating group*, defined as the subgroup of S_n that contains exactly the 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 $e \leftrightarrow \text{even}$, $f \leftrightarrow \text{odd}$. Note that D_2 can be viewed as $(\mathbb{Z}/2\mathbb{Z}, +)$. That means there is a morphism

$$\text{sign}: S_n \rightarrow (\mathbb{Z}/2\mathbb{Z}, +)$$

that sends even permutations to $0 \bmod 2\mathbb{Z}$, 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 among all groups.

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

PROOF. Take the base set for the permutations to be all the elements of G . So, imagine the elements are all lined up and numbered. Then take $g \in G$ and

note that as a set gG is the same as G by the Cancellation Property in groups. Then left-multiplication by g simply changes the order in which the elements of G are lined up. We read this as a permutation σ^g , and so each $g \in G$ gives rise to a permutation of the elements of G . Multiplication in by g followed by multiplication by g' corresponds to the composition σ^g followed by $\sigma^{g'}$, so we have the morphism property. Finally, knowing σ^g allows to recover g because you just have to see into what bucket the identity $e \in G$ went: the label of that bucket is g since $ge = g$. \square

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” $e, \leftrightarrow, \updownarrow, \curvearrowright$. Now ask what the effect of multiplying by the group elements is on the sequence $\{e, \leftrightarrow, \updownarrow, \curvearrowright\}$. We find

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

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

$$\begin{aligned} e &\text{ becomes } () \\ \updownarrow &\text{ becomes } (e, \updownarrow)(\leftrightarrow, \curvearrowright) \\ \leftrightarrow &\text{ becomes } (e, \leftrightarrow)(\updownarrow, \curvearrowright) \\ \curvearrowright &\text{ becomes } (e, \curvearrowright)(\updownarrow, \leftrightarrow). \end{aligned}$$

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

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

as subgroup of S_4 .

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. 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 H is a normal subgroup if it has no conjugate subgroups aside from 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 $aha^{-1} = h$ (although this certainly *could* happen).

Let us also recall that H can be used to make H -shaped 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 $(aH) * (bH) = abH$. Let's 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 mentioned 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 $(gh)(g'h') \in gg'H$. (Multiplication should not depend on the specific representative we picked; if it does, multiplication would destroy cosets). But

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

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 do this for transpositions. And since H stays H when you arbitrarily permute 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 element 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 either is made only of even or only odd permutations.

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 it 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 $\text{Sym}(\triangle) = S_3$.

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 $() \in E$, $(1, 2) \in \gamma$, $(1, 3) \in \beta$, $(2, 3) \in \alpha$,

$(1, 2, 3) \in \lambda$ 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 because all 6 elements keep the number 4 fixed and only permute 1,2,3. 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 for G/H are then

$$\begin{aligned}\bar{0} &= \{0 + 24\mathbb{Z}, 6 + 24\mathbb{Z}, 12 + 24\mathbb{Z}, 18 + 24\mathbb{Z}\}, \\ \bar{1} &= \{1 + 24\mathbb{Z}, 7 + 24\mathbb{Z}, 13 + 24\mathbb{Z}, 19 + 24\mathbb{Z}\}, \\ \bar{2} &= \{2 + 24\mathbb{Z}, 8 + 24\mathbb{Z}, 14 + 24\mathbb{Z}, 20 + 24\mathbb{Z}\}, \\ \bar{3} &= \{3 + 24\mathbb{Z}, 9 + 24\mathbb{Z}, 15 + 24\mathbb{Z}, 21 + 24\mathbb{Z}\}, \\ \bar{4} &= \{4 + 24\mathbb{Z}, 10 + 24\mathbb{Z}, 16 + 24\mathbb{Z}, 22 + 24\mathbb{Z}\}, \\ \bar{5} &= \{5 + 24\mathbb{Z}, 11 + 24\mathbb{Z}, 17 + 24\mathbb{Z}, 23 + 24\mathbb{Z}\}.\end{aligned}$$

So, for example, the last of these cosets contains all numbers that leave rest 5 when divided by 6. If we recall that we are supposed to use addition as operation in G/H , 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 also 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. \diamond

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

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

If we hope that we can make $\{E, R, L\}$ a quotient 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 $\mathbb{Z}/m\mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z}$ 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 $\psi: G \rightarrow G'$ 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 out what this quotient group has to do with ϕ in concrete terms.

In order to prepare for things, let us introduce the following concept.

DEFINITION VI.6. If $m, p, q \in \mathbb{Z}$ and $m = pq$ then consider the subgroup of $\mathbb{Z}/m\mathbb{Z}$ consisting of the cosets of the multiples of q . We call this group the *inflation of $\mathbb{Z}/p\mathbb{Z}$ by the factor q* and denote it $q \star (\mathbb{Z}/p\mathbb{Z})$.

Let's try to digest this. In $\mathbb{Z}/m\mathbb{Z}$ there are $m/q = p$ different cosets that belong to integers that are multiples of q , namely $0+m\mathbb{Z}, q+m\mathbb{Z}, 2q+m\mathbb{Z}, \dots, (p-1)q+m\mathbb{Z}$. (We note that $pq+m\mathbb{Z} = 0+m\mathbb{Z}$). So, this subgroup we are talking about has p elements. Moreover, we can make an isomorphism between $\mathbb{Z}/p\mathbb{Z}$ and $q \star (\mathbb{Z}/p\mathbb{Z})$ by sending $a+p\mathbb{Z}$ in $\mathbb{Z}/p\mathbb{Z}$ to $qa+m\mathbb{Z}$ in $\mathbb{Z}/m\mathbb{Z}$. In other words, $\mathbb{Z}/p\mathbb{Z} \simeq q \star (\mathbb{Z}/p\mathbb{Z})$ as groups; the only difference is that instead of counting $1, 2, 3, \dots, p-1, p=0$ we count $1q, 2q, \dots, (p-1)q, pq=0$. Everything has been inflated by a factor of q .

EXAMPLE VI.7. Let $G = \mathbb{Z}/28\mathbb{Z}$, $G' = \mathbb{Z}/42\mathbb{Z}$ and $\psi: G \rightarrow G'$ 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 criterion (IV.2.1), “multiplication by 9” does indeed give a function from $\mathbb{Z}/28\mathbb{Z}$ to $\mathbb{Z}/42\mathbb{Z}$ 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 $\begin{pmatrix} \bar{0} & \bar{14} \\ \bar{14} & \bar{0} \end{pmatrix}$. 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 $0, 9, 18, 27, 36, 45, \dots$. 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 $3 + 42\mathbb{Z}, 6 + 42\mathbb{Z}, \dots, 0 + 42\mathbb{Z}$ is a group with addition and sits inside $\mathbb{Z}/42\mathbb{Z}$. It is called the *image* of ψ , written $\text{im}(\psi)$, 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 $\text{im}(\psi)$ 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 $\mathbb{R}^n, \mathbb{R}^m$ 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 $\text{rk}(A)$. And to top it all off, you learned that rank plus nullity gives n . In new and fancy terms this can be phrased as “the kernel of A is a vector space of dimension $n - \text{rk}(A)$, and $\mathbb{R}^n/\ker(A)$ is a vector space of dimension $n - (n - \text{rk}(A)) = \text{rk}(A)$. This quotient is precisely the column space of A , a vector space of dimension $\text{rk}(A)$ just like $\mathbb{R}^n/\ker(A)$.

THEOREM VI.8 (The isomorphism theorem). *If*

$$\psi: G \rightarrow G'$$

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

$$\bar{\psi}: G/\ker(\psi) \simeq \text{im}(\psi)$$

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

Here, the group operation in G/H is $(aH)(bH) = abH$ and the operation in $\text{im}(\psi)$ 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, while all of $H = \ker(\psi)$ is crunched down to $e_{G'}$, basically by definition. Therefore if you want to relate stuff in 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 back 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 $G/\ker(\psi)$ has to do with $\text{im}(\psi)$. And the function $\bar{\psi}$ that 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,

$$G/\ker(\psi) \simeq G'.$$

EXAMPLE VI.9. 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 $\ker(\psi) = \{a + 15\mathbb{Z} \mid 5a + 25\mathbb{Z} = 0 + 25\mathbb{Z}\}$. This requires that $25 \mid 5a$ 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 $\text{im}(\psi)$ of ψ are the cosets $\{5a + 25\mathbb{Z}\}$. That is a group of 5 elements. We know (as 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 then, $\text{im}(\psi)$ looks like $\mathbb{Z}/5\mathbb{Z}$ inflated by a factor of 5.

So, the isomorphism theorem says that $(\mathbb{Z}/15\mathbb{Z})/(\ker(\psi)) \simeq 5 \star (\mathbb{Z}/3\mathbb{Z}) \simeq \mathbb{Z}/5\mathbb{Z}$.

- (2) More generally, let $n|km$ and consider the morphism $k: \mathbb{Z}/m\mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z}$ 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 minimal 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 $\text{lcm}(n, k)$, and of course that smallest ak that is a multiple of $\text{lcm}(n, k)$ is just $\text{lcm}(n, k)$. It follows that the corresponding a is $\text{lcm}(n, k)/k$ and so equals $n/\gcd(n, k)$ since in general $xy = \text{lcm}(x, y) \cdot \gcd(x, y)$.)

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

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 $\nu := n/\gcd(k, n) = \text{lcm}(k, n)/k$.

Altogether, the isomorphism theorem says:

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

Note that $\kappa \cdot \nu = (mk/\text{lcm}(n, k)) \cdot (\text{lcm}(k, n)/k) = m$ as it should (it should, since on both sides there should be equally many elements).

- (3) In the previous items, normality came for free since the groups were Abelian. Let now G be the unit quaternions, with multiplication table

$$\begin{pmatrix} 1 & i & j & k & -1 & -i & -j & -k \\ i & -1 & k & -j & -i & 1 & -k & j \\ j & -k & -1 & i & -j & k & 1 & i \\ k & j & -i & -1 & -k & -j & i & 1 \\ -1 & -i & -j & -k & 1 & i & j & k \\ -i & 1 & -k & j & i & -1 & k & -j \\ -j & k & 1 & -i & j & -k & -1 & -i \\ -k & -j & i & 1 & k & j & -i & -1 \end{pmatrix}$$

This group arises as the unit vectors in a 4-dimensional real vector space with bases $1, i, j, k$ on which a product has been defined that allows to multiply vectors with vectors and gives you vectors.¹

Note that in the multiplication table you can find various sub-tables that we know something about. Restriction to columns and rows 1 and 5 is the multiplication table of $\{\pm 1\}$. The restriction to rows and columns 1, 2, 5, 6 gives the multiplication table for the complex numbers $\{\pm 1, \pm i\}$ and if you look closely you recognize this as equivalent to the table for $(\mathbb{Z}/4\mathbb{Z}, +)$ since i generates the group $\{\pm 1, \pm i\}$.

In any event, the center Z of the unit quaternion group (the elements that commute with everyone else) consists of the elements $\{\pm 1\}$ as is easy to see. By definition, the center of any group is normal in that group.

The quotient of the quaternions 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 of G relative to Z , 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.

The isomorphism theorem says in this case: $G/Z \simeq KV_4$.

¹You have seen something like this twice before in life: the usual product of real numbers on \mathbb{R}^1 , and the product on \mathbb{R}^2 that you get when you read \mathbb{R}^2 as the complex numbers \mathbb{C} and use complex multiplication. The remarkable thing is that the three product structures mentioned (real, complex, quaternion) allow to divide (except by the zero vector of course). In this sense it is very unlike the cross product on \mathbb{R}^3 which also turns pairs of vectors into vectors but which has zero divisors and so does not allow division. The only other case where such product with division can be found is on \mathbb{R}^8 with the “octonion” product. It might be noted that these products get progressively worse: \mathbb{C} cannot be ordered in a way compatible with the product; quaternion multiplication is not commutative; octonion multiplication is not associative.

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 $T: V \rightarrow W$ from one vector space to another such that $T(\vec{v} + \vec{v}') = T(\vec{v}) + T(\vec{v}')$ 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 $B_V = \{b_1^V, \dots, b_n^V\}$ and $B_W = \{b_1^W, \dots, b_m^W\}$ 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 $B_V \cdot c^{B_V}(\vec{v}) = \sum c(\vec{v})_i^{B_V} b_i^V$. Here, the upper index “ B_V ” emphasizes that this is the coefficient vector relative to the basis B_V .

Recall that if $T: V \rightarrow W$ 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 $c^{B'_V}(\vec{v}) = Q_V \cdot c^{B_V}(\vec{v})$.

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 for a transformation (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 of 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 echelon 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 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 echelon form, which is practically the transpose of the row reduced echelon form of the transpose of A . Now imagine what the row reduced echelon 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 subgroup 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. This matrix is called the *Smith normal form* of the input matrix. Its special feature is: the only nonzero entries are on the diagonal, and from upper left to lower right the diagonal entries divide each other.

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 multiple 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 $G/H = (\text{trivial group}) \times (\mathbb{Z}/3\mathbb{Z}) \times ((\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z}))$.

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). \diamond

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, \dots, 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*

$$Q = \mathbb{Z}/n\mathbb{Z} \quad \text{with } n = \text{ord}(g)$$

as a quotient of \mathbb{Z} by a suitable subgroup. What this means is that there is an assignment

$$\mathbb{Z} \rightarrow \langle g \rangle$$

that sends $k \in \mathbb{Z}$ to g^k . This is sort of like an exponential map and clearly has the morphism property (according to exponential laws). Since $n = \text{ord}(g)$, the multiples mn of n are all sent to $g^{mn} = (g^n)^m = e^m = e$. And if k is sent to e then k is a multiple of n . It follows that there is actually an isomorphism

$$\mathbb{Z}/n\mathbb{Z} \rightarrow \langle g \rangle$$

sending $k + n\mathbb{Z}$ to g^k .

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 $\mathbb{Z}^k \rightarrow Q$ 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 factors of the form $(\mathbb{Z}/1\mathbb{Z})$ in item (6) of the theorem does not change anything.

All parts except the last one follow from what we have done and said in examples. At the end of the section I explain why the last part is true (namely, 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 $G = KV_4 = \{e, \updownarrow, \leftrightarrow, \curvearrowright\}$. 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 $\pi: \mathbb{Z}^3 \rightarrow KV_4$ that sends $(1, 0, 0)$ to \updownarrow , $(0, 1, 0)$ to \leftrightarrow and $(0, 0, 1)$ to \curvearrowright .

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 $\updownarrow, \leftrightarrow, \curvearrowright$ 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

$\pi((1, 0, 0) + (1, 0, 0)) = \pi((1, 0, 0)) \cdot \pi((1, 0, 0)) = \uparrow \cdot \uparrow = e$, 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 $\uparrow \cdot \leftrightarrow = \curvearrowright$, 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 far 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

$$\begin{array}{c|c|c|c} e & \uparrow & \leftrightarrow & \curvearrowright \\ \hline (0 + 2\mathbb{Z}, 0 + 2\mathbb{Z}) & (1 + 2\mathbb{Z}, 0 + 2\mathbb{Z}) & (0 + 2\mathbb{Z}, 1 + 2\mathbb{Z}) & (1 + 2\mathbb{Z}, 1 + 2\mathbb{Z}) \end{array}$$

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:

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

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 non-trivially, that is not so thrilling and 3, 7 only appear in the last elementary divisor. \diamond

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, \dots, g_m . Then G has a presentation $\pi: \mathbb{Z}^m \twoheadrightarrow G$ 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 .*

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. The structure of G can be read off from Part 2 of FTFGAG.

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 multiplication: $\sigma^g(g') = gg'$. 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

$$\begin{aligned}\lambda: G \times X &\rightarrow X, \\ (g, x) &\mapsto \lambda(g, x).\end{aligned}$$

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 the action on X to be compatible with the group multiplication:

$$\lambda(g, \lambda(g', x)) = \lambda(gg', x).$$

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 $\rho: X \times G \rightarrow X$ 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

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

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.

In 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 $\text{orb}_G(x)$,

$$\text{orb}_G(x) := \{y \in X \mid \exists g \in G \text{ with } \underbrace{gx}_{=\lambda(g,x)} = y\}.$$

So in simplified notation, $\text{orb}_G(x) = Gx$.

- 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 ,

$$\text{Stab}_G(x) = \{g \in G \text{ with } gx = x\},$$

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 $\text{Stab}_G(x) = G$, 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 $\text{Fix}_X(g)$ for the points $x \in X$ for which $gx = x$.

REMARK VIII.3. (1) To be in the same orbit is an equivalence relation. Indeed, if $x' \in \text{orb}_G(x)$ then there is $g \in G$ with $gx = x'$. But then $(g^{-1})x' = (g^{-1})(gx) = (g^{-1}g)x = ex = x$ according to the group action rules. So, being in each other's orbit is symmetric. Of course it is reflexive (since $ex = x$, x is always in its own orbit). Finally, if $gx = x'$ and $g'x' = x''$ then $g'gx = x''$ and so being in each other's orbit is transitive.

(2) Under the relation “being in each other's orbit”, the equivalence classes are precisely the orbits.

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

We consider some examples, concrete and abstract.

EXAMPLE VIII.4. Let $G = \text{Sym}(\triangle)$ 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 $\text{Stab}_G(x)$ consists of the identity e and the A -flip a , since the other 4 elements b, c, ℓ, r of G all move the A -vertex 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 all 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 symmetries of a cube, and let G be the rigid symmetry group of a cube. (This is the subgroup of all symmetries of the cube consisting of just the rigid motions that are cube symmetries). It turns out that $|\overline{G}| = 48$, 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). The rotations are: the identity, the 4×2 rotation by 180 about the big diagonals, the 3×3 rotations by 90, 180 and 270 about the lines that link centers

of opposite faces, and the 6×1 rotations by 180 about the lines that link centers of opposite edges.

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 arise as the reflections on the 3 planes that contain x , the center of the cube, and one vertex adjacent to x . In particular, $|\text{Stab}_G(x)| = 3$ and $|\text{Stab}_{\overline{G}}(x)| = 6$.

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 (since it must fix the center of the cube).

The 2×4 non-trivial rotations that fix a big diagonal have 2 fixed points each. The identity of G has 8 fixed points. The 2×4 non-rigid 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 $agH = gH$. For example, if $g = e$, the condition $geH = eH$ becomes $g \in H$, so the stabilizer of the “point” eH in X is exactly H . In general, 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 in 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 $g'gH = gH$. So, g' is the “ a ” from above and 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 multiplication with g' precisely when gHg^{-1} contains 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 G as then all subgroups H are normal.

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

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

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

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

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

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

Fix $x, y \in X$ in the same G -orbit of X . Then $\text{Stab}_G(x)$ and $\text{Stab}_G(y)$ are conjugate subgroups as you show in homework. So in particular, they have the same size.

Next, define a clustering of the elements of G by requiring that g, g' are in the same cluster if and only if $gx = g'x$. Note that the elements of $g \cdot \text{Stab}_G(x)$ all end up in the same cluster since they all send x to gx ; these sets are just the left cosets relative to $H := \text{Stab}_G(x)$. One then checks (easy but detailed) that g, g' belonging to different H -cosets rules out the possibility of $gx = g'x$. So, the clusters all are of the form $g \cdot \text{Stab}_G(x)$, and in particular all of the same size as $\text{Stab}_G(x)$. So, G is partitioned into clusters of size $|\text{Stab}_G(x)|$, 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, $|G| = |\text{orb}_G(x)| \cdot |\text{Stab}_G(x)|$.

(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 Lagrange's: " $|G| = |G/H| \cdot |H|$ ". In reverse, this theorem and its proof is simply Lagrange applied to G and its subgroup $H := \text{Stab}_G(x)$).

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

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 $\text{Fix}_X(g)$ is the points of X that are unmoved by g , so $gx = x$. Let us also write X/G for the orbit space of X under G . This is just the set of all orbits, the notation suggesting 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 number of G -orbits on X (i.e., the size of the orbit space X/G) is*

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

Again, I won't give a very formal proof but only the main ideas. Let us count $\sum_{g \in G} |\text{Fix}_X(g)|$ 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 of the form $\text{Fix}_X(g)$ and so the number of all such pairs is precisely $\sum_{g \in G} |\text{Fix}_X(g)|$. But if we cluster by x , then each cluster has the form $\{g \in G | gx = x\}$ and that is exactly $\text{Stab}_G(x)$. So, if we now sum this over all x we get $\sum_{x \in X} |\text{Stab}_G(x)|$. Of course, these two counts must agree:

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

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 $|G| = |\text{Stab}_G(x)| \cdot |\text{orb}_G(x)|$. So, restricting the sum to the orbit of x , we get $\sum_{x \in \text{orb}_G(x)} |\text{Stab}_G(x)| = \sum_{x \in \text{orb}_G(x)} |G|/|\text{orb}_G(x)| = |G| \sum_{x \in \text{orb}_G(x)} 1/|\text{orb}_G(x)| = |G|$.

So, orbit by orbit, the expression $\sum_{x \in X} |\text{Stab}_G(x)|$ 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 $|G| \cdot |X/G| = \sum_{x \in X} |\text{Stab}_G(x)|$. Combined with the last display above, this proves 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 the six numbers on each of 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. (This is a physical cube, so only the rigid motions count).

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 faces 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 of them). So there are no g at all with a 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 $|\text{Fix}_X(e)| = |X| = 720$. All other summands belong to a g without fixed points and contribute 0. So the formula says $|X/G| = \frac{1}{24}(720 + 0 + \dots + 0) = 30$.

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
 - examples: symmetry groups, KV_4 , $GL(n, \mathbb{R})$, vector spaces, $(\mathbb{Z}/n\mathbb{Z}, +)$, $U(n)$, free groups, \mathbb{Z}^n
 - Abelian groups, cyclic groups
 - order of a group, and of elements in a group
 - subgroup
 - product group $G \times H$
 - $\text{Aut}(G)$, the relabelings of G that preserve the Cayley table, is a group with composition of symmetries as multiplication.
- Week 3
 - the Euler ϕ -function
 - the number of elements in the cyclic group C_n of n elements that have order d (distinction for $d|n$ and $d \nmid n$)
 - the number of subgroups of a given size in C_n
 - the number of different 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 $a = a_1 \cdots a_k$ then $C_{a_1} \times C_{a_2} \times \cdots C_{a_k} = C_a$ provided that the a_i are pairwise coprime
 - solving simultaneous equations $x \bmod n = a \bmod n, a \bmod m = b \bmod m$ when m, n coprime
 - $U(mn) = U(m) \times U(n)$ if coprime
 - $U(p^k)$ is cyclic if $p > 2$ is prime
 - $U(2)$ and $U(4)$ are cyclic, but $U(2^k)$ is not cyclic when $k > 2$
 - $|U(p^k)| = p^{k-1}(p-1)$, and why
 - $\phi: (\mathbb{Z}/mn\mathbb{Z}) \rightarrow (\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$ via $1 + mn\mathbb{Z} \mapsto (1 + m\mathbb{Z}, 1 + n\mathbb{Z})$ is isomorphism provided m, n coprime
 - know how to manufacture the inverse map explicitly (by solving simultaneous equations)
- Week 4
 - left and right cosets of G relative to the subgroup H ; coset space G/H

- morphisms $\psi: G \rightarrow G'$ and a list of examples
- know how to test whether multiplication by $k \in \mathbb{N}$ gives a morphism $\mathbb{Z}/m\mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z}$
- conjugation by a , $g \mapsto aga^{-1}$
- inner automorphisms $\phi_a: G \rightarrow G$, sending $g \mapsto aga^{-1}$
- properties of cosets: either equal or disjoint; union is G ; all same size
- Lagrange: $\text{ord}(g) \mid \text{ord}(G)$; $|G| = |H| \cdot |G/H|$
- 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
 - transposition = 2-cycle
 - disorder as number of switches
 - odd/even: parity
 - sign of a permutation
 - the alternating group as the kernel of $\sigma \mapsto \text{sign}(\sigma)$, a group morphism from S_n to $(\pm 1, \cdot)$.
 - every group is a subgroup of a permutation group
- Week 6
 - suppose here H is normal. Then G/H can be made a group, $(g_1H) \cdot (g_2H) = (g_1g_2H)$. That this works is precisely because H is normal.
 - the kernel of a morphism is a normal subgroup
 - if $\phi: G \rightarrow G'$ is a morphism, denote H the kernel. Then G/H is isomorphic to the image of ϕ (and 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, \dots, d_m divide one another, $d_i \mid d_{i+1}$
 - 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 $d_1 \mid \dots \mid d_m$ 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
- Week 8
 - rules for a group action
 - know what $\text{Stab}_G(x)$, $\text{orb}_G(x)$, $\text{Fix}_X(g)$ are, and how to find them
 - orbit space X/G
 - orbit-stabilizer theorem
 - Burnside theorem

Some practice problems:

- Show by induction that $1 + 2 + \dots + n = n(n+1)/2$.
- Compute the gcd of 342 and 543 as a linear combination of the two inputs.
- Find the number of physical symmetries of a regular octahedron. Do they form an Abelian group? Find 6 subgroups with 4 elements each.
- Write up the Cayley table for the symmetries of a square.
- What do you know about the Cayley table of an Abelian group?
- How many subgroups does a cyclic group of order 10 have?
- Find the orders of all elements in the symmetry group of the regular pentagon.
- Discuss the order of the elements in the group $(\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/6\mathbb{Z})$.
- Identify the automorphisms of $(\mathbb{Z}/2\mathbb{Z}) \times \mathbb{Z}/2\mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$ with the 3×3 invertible matrices with entries in $\mathbb{Z}/2\mathbb{Z}$. (That is, explain how this identification works). Challenge: find the number of elements in this group.
- Compute $\phi(111111)$. (Note that $3 \cdot 37 = 111$).
- How many elements of $\mathbb{Z}/1000\mathbb{Z}$ have order 20?
- How many subgroups of $\mathbb{Z}/1000\mathbb{Z}$ have order 20?
- Is $(\mathbb{Z}/10\mathbb{Z}) \times (\mathbb{Z}/25\mathbb{Z})$ cyclic?
- Find $n \in \mathbb{N}$ as small as possible such that $n \bmod 24 = 3 \bmod 24$ and $n \bmod 35 = 17 \bmod 35$.
- Explain why $(\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/7\mathbb{Z})$ is $\mathbb{Z}/21\mathbb{Z}$. Find an explicit morphism ψ from the former to the latter, and determine the preimage $\psi^{-1}(4 + 21\mathbb{Z})$ under this morphism.
- Let G be the physical symmetries of a regular tetrahedron. Let H be the subgroup that consists of those symmetries that do not move a certain vertex. Is H normal in G ?
- If ψ is a morphism from $\mathbb{Z}/20\mathbb{Z}$ to $\mathbb{Z}/75\mathbb{Z}$, list all elements that could be $\psi(1 + 20\mathbb{Z})$. If you know that $\psi(3 + 20\mathbb{Z})$ is $45 + 50\mathbb{Z}$, what can you say about $\psi(1 + 20\mathbb{Z})$?
- If G is the physical symmetries of the regular dodecahedron, and H is the subgroup that fixes one particular vertex, compute the size of the coset space G/H .
- Let $\sigma = (123)(456)$ in cycle notation. Let $a = (135)(246)$ in cycle notation. Find the conjugate of σ by a .
- Let G be the invertible upper triangular 2×2 matrices with entries in $\mathbb{Z}/3\mathbb{Z}$. Find $|G|$. Determine whether the diagonal invertible matrices form a normal subgroup.
- Find the disorder and sign of the permutation $(1, 2, 3, 4, 5, 6, 7, 8)$, in cycle notation.
- If G is the group of all symmetries of the square, and if H is the rotations of the square, determine the structure of G/H . (You can take for granted that H is normal in G).
- Let G be the 2×2 upper triangular matrices with entries in $\mathbb{Z}/5\mathbb{Z}$. Find $|G|$ and also find the size of the subgroup H of such matrices with determinant $1 + 5\mathbb{Z}$. Explain why H is normal in G . Challenge: can you find the structure of the quotient group G/H ?

- Find the Smith normal form of $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$.
- Which of the following could be the elementary divisors of a matrix? $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 2, 5\}$, $\{1, 2, 6\}$.
- Write $\mathbb{Z}/5040\mathbb{Z}$ as product of cyclic groups that cannot be factored further.
- Find the number of different Abelian groups with 5040 elements.
- Let G be the group of invertible 2×2 matrices with rational entries. These matrices act naturally on the \mathbb{Q} -vectors of length 2 by multiplication from the left. Describe the stabilizer of the vector $(1, 0)$. Is this a normal subgroup?
- Let G be the 2×2 invertible matrices with entries in $\mathbb{Z}/5\mathbb{Z}$. We let G act on the vectors of length 2 with entries in $\mathbb{Z}/5\mathbb{Z}$ by left multiplication. Show that this action is transitive.

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 $+: R \times R \rightarrow R$ called *addition* and a second binary operation $\cdot: R \times R \rightarrow R$ 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 in force: $r(s+t) = r \cdot s + r \cdot t$ 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.

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

by subtracting $a \cdot 0$ on both sides

$$0 = a \cdot 0 = a \cdot (1 + (-1)) = a \cdot 1 + a \cdot (-1),$$

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 do 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 inverses *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 $a' =$

$a' \cdot e = a' \cdot (a \cdot a'') = (a' \cdot a) \cdot a'' = e \cdot a'' = a''$. Recall that ring elements with inverses in the ring are called *units*, see Definition I.23.

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 modeled 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 contrast 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. On the other hand, the identity matrix is a neutral element for multiplication, so this ring has a 1.
- Another collection of rings are the *polynomial rings* $\mathbb{K}[x_1, \dots, x_n]$ over a chosen coefficients field \mathbb{K} . These are commutative rings, and their elements are the polynomials in the variables x_1, \dots, 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 for which both the real and imaginary parts are integers. 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: $(a + b\sqrt{-1}) + (c + d\sqrt{-1}) = (a + c) + (b + d)\sqrt{-1}$. Multiplication 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 ring is called 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 $R[x_1, \dots, x_n]$ in the case that R has zero-divisors. You might want to check these three claims explicitly by finding one example of zero-division 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 $1_R, 1_R+1_R, 1_R+1_R+1_R, \dots$. 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 $\underbrace{1 + 1 + \dots}_{c \text{ copies}} = 0$ and c is the characteristic (the smallest positive such c). Suppose $c = mn$ factors. Then let $e_m = \underbrace{1 + 1 + \dots}_{m \text{ copies}}$ and $e_n = \underbrace{1 + 1 + \dots}_{n \text{ copies}}$. 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 factorization of c , then $1 < m, n < c$ and this makes it impossible that $e_m = 0$ or $e_n = 0$. We conclude c does not factor and is therefore a prime number. \square

DEFINITION IX.8. A commutative ring that has multiplicative inverses for each nonzero element is called a *field*. (If in a non-commutative ring all nonzero elements are units, it is called a *skew-field*. We will not talk about these).

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, \dots, r_t be the complete list of nonzero elements of R . Then ar_1, \dots, ar_t is another list of elements of R . 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. It follows that the second list is of the same length as the original one, and in particular must be a permutation of the first list. (This is where finiteness is used: a subset of a finite set that has as many elements as the whole set must agree with the whole set. In contrast, if R were infinite we could not argue like this. For example, the integer multiples of 2 are not a permutation of the integers, although to each integer corresponds exactly one even integer). 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 $a \cdot r_i = 1_R$. \square

REMARK IX.10. A postscript of this proof goes like this: let R have p elements. By the theorem, we are looking at a field and so the nonzero elements are a group with multiplication as group operation. This multiplicative group has $p-1$ elements. So Lagrange says that if a is a nonzero element of R then its multiplicative order divides $p-1$. In particular, there is a power c with $c|(p-1)$ 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 . We will show later that the units of a field actually form a cyclic group.

EXAMPLE IX.11. In the same way as the Gaussian integers are an extension of the integers, one can make 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 $\bar{2} + \bar{1}\sqrt{2}$ is $\frac{\bar{1}}{\bar{2} + \bar{1}\sqrt{2}} = \frac{\bar{2} - \bar{1}\sqrt{2}}{(\bar{2} + \bar{1}\sqrt{2})(\bar{2} - \bar{1}\sqrt{2})} = \frac{\bar{2} - \bar{1}\sqrt{2}}{\bar{4} - \bar{2}} = \bar{1} - \bar{2}\sqrt{2} = \bar{1} + \bar{1}\sqrt{2}$. And indeed, $(\bar{2} + \bar{1}\sqrt{2})(\bar{1} + \bar{1}\sqrt{2}) = \bar{1} + \bar{0}\sqrt{2}$.

CHAPTER X

Week 9/10: Ideals and morphisms

Recall that we insist that our rings are commutative, and all have a multiplicative identity 1. (All rings have a neutral element for $+$, which we always write as 0, since R with $+$ is an Abelian group).

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

$$f(r_1 \cdot_R r_2) = f(r_1) \cdot_{R'} f(r_2).$$

Examples of such things abound.

- the inclusions $\mathbb{Z} \hookrightarrow \mathbb{Q} \hookrightarrow \mathbb{R} \hookrightarrow \mathbb{C}$ and the inclusions $\mathbb{Z} \hookrightarrow \mathbb{Z}[\sqrt{-1}] \hookrightarrow \mathbb{C}$;
- the surjection $\mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z}$ sending k to $k + n\mathbb{Z}$ for any n ;
- if $m|n$, the surjection $\mathbb{Z}/n\mathbb{Z} \rightarrow \mathbb{Z}/m\mathbb{Z}$ sending $k + n\mathbb{Z}$ to $k + m\mathbb{Z}$;
- complex conjugation;
- the “conjugation” $\mathbb{Z}[\sqrt{2}] \rightarrow \mathbb{Z}[\sqrt{2}]$ sending $a + b\sqrt{2}$ to $a - \sqrt{2}$ and any similar constructs;
- the polynomial map $\mathbb{C}[x, y] \rightarrow \mathbb{C}[t]$ 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* $\epsilon_a: \mathcal{O} \rightarrow \mathbb{R}$ 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, such as $\mathbb{Z}/3\mathbb{Z}$. If $\text{char}(R) = p > 0$ is prime, there is the *Frobenius* morphism $\text{Frob}: R \rightarrow R$ that sends $r \in R$ to $\text{Frob}(r) = r^p$. 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 $f: R \rightarrow R'$ 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, \dots, 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, \dots, f_k . In other words, I contains precisely all the possible linear combinations $r_1 f_1 + \dots + r_k f_k$ where r_1, \dots, r_k run through all 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.

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

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

PROOF. Note that the kernel of a ring morphism $f: R \rightarrow R'$ is the set of elements that map to zero, and thus is a subgroup of R since f is a group morphism. Now take $x \in I$ and $r \in R$. Then $f(x) = 0$ and so $0 = f(x)f(r) = f(rx)$ by the morphism property. It follows that $rx \in \ker(f) = I$. \square

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

DEFINITION X.7. Let $I \subseteq R$ be an ideal. The *quotient ring* (or also *factor ring*) R/I is the group R/I together with multiplication $(x + I)(y + I) = xy + I$. There is an induced morphism $\pi: R \rightarrow R/I$ 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 $f: R \rightarrow R'$ 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).
- In contrast, the *preimage* $f^{-1}(J')$ is always an ideal of R . This is seen as follows. Since f is a group morphism, the preimage of J' 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 $\text{char}(R) = n$ then there is a natural morphism $\mathbb{Z}/n\mathbb{Z} \hookrightarrow R$ induced by sending $1 \in \mathbb{Z}$ to $1 \in R$ and using the morphism rule.

The main structure theorem for ideals says:

THEOREM X.8 (Structure of Quotient Rings). *If I is an ideal of R then under the natural surjection $\pi: R \rightarrow R/I$ 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 ideal $\{0 + 6\mathbb{Z}\}$ 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 ideal in 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 with the ring. For example, the multiples of 3 form a prime ideal in \mathbb{Z} , but in $\mathbb{Z}[\sqrt{-2}]$ you can factor $3 = (1 + \sqrt{-2})(1 - \sqrt{-2})$ and neither of the two factors is a multiple of 3. Similarly, $3\mathbb{Z}[\sqrt{-2}]$ can also not be maximal.

On the other hand, maximal ideals can also get “too big”: in \mathbb{Q} the multiples of 3 are all of \mathbb{Q} , which does not qualify as maximal ideal.

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 . If a ring R' has only two ideals, one has to be the ideal $\langle 0_{R'} \rangle$, and the other the ideal $R' = \langle 1_{R'} \rangle$. For $R' = R/I$, $\langle 0_{R'} \rangle = I/I$. 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 them interesting ideals.

Recall from Definition IX.8 that rings with all nonzero elements invertible are called fields.

LEMMA X.12. *If R is a commutative ring with identity,*

$$[\text{every } r \neq 0 \text{ is a unit}] \Leftrightarrow [R \text{ has exactly 2 ideals}].$$

In particular, fields are domains.

PROOF. Suppose the ring has exactly two ideals. 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 $x \neq 0$, I is not $\langle 0 \rangle$. By the 2-ideal-hypothesis, $I = R$. This means in particular that $1_R \in I$ and so 1_R is a multiple of x . But then x must be invertible.

If every nonzero element in R is invertible then as soon as an ideal contains a nonzero element, it also contains 1_R and thus equals R . So such R have only two ideals.

If a field R is not a domain, then it must have at least one pair of zero-divisors $0 = ab$ for some $a \neq 0 \neq b$ in R (see Definition IX.4). But as a field, R contains an inverse for a and then $a^{-1}ab = a^{-1}0$ leads to $b = 0$, a contradiction to zero-divisors. \square

Examples of fields include $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ but also things like $\mathbb{Z}/p\mathbb{Z}$ with p prime:

LEMMA X.13. *If $p \in \mathbb{Z}$ is prime then $\mathbb{Z}/p\mathbb{Z}$ is a field and conversely.*

PROOF. We look at the morphism $\mathbb{Z} \rightarrow \mathbb{Z}/p\mathbb{Z}$ 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 any 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 algorithm 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 $1 + p\mathbb{Z} = (a + p\mathbb{Z})(x + p\mathbb{Z})$. This implies $I/p\mathbb{Z}$ 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 not divide m (since n is not unit). So $\mathbb{Z}/p\mathbb{Z}$ can't be a field. \square

More generally,

PROPOSITION X.14. *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. \square

THEOREM X.15. *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 $g = \gcd(a, b)$ 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, \dots, 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 $a_1, a_2, \dots, a_n, \dots$. We know that

$$\gcd(a_1) \geq \gcd(a_1, a_2) \geq \gcd(a_1, a_2, a_3) \dots \geq 0.$$

Since no infinite strictly descending sequences can exist in \mathbb{N} , it follows that this sequence of \geq symbols reaches a point (say, when the index is k) from where onward each \geq is actually a $=$.

What this means is that $\gcd(a_1, \dots, a_k)$ divides a_{k+1}, a_{k+2}, \dots . But then a_{k+1}, a_{k+2}, \dots are already in the ideal generated by a_1, \dots, a_k and we can say that

$$\langle a_1, a_2, \dots, a_n, \dots \rangle = \langle a_1, \dots, a_k \rangle = \langle \gcd(a_1, \dots, a_k) \rangle$$

is generated by one element.

The second part is a homework problem. \square

DEFINITION X.16. 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*, short PID.

REMARK X.17. You will prove in homework that ideals in $\mathbb{Z}/n\mathbb{Z}$ are also all principal.

CHAPTER XI

Week 10, Euclidean rings

In this chapter we look exclusively at domains. As before, rings are commutative (unless expressly indicated not to be) and have a 1. We start with discussing polynomial rings.

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, \dots 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 a 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,

$$\sum_{i=0}^{\infty} r_i x^i = \sum_{i=0}^{\infty} r'_i x^i$$

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 $r_0 + r_1 x + \dots + r_d x^d$ instead of $\sum_{i=0}^{\infty} r_i x^i$.

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 $0x^0 + 0x^1 + 0x^2 + \dots$ and 1-element $1x^0 + 0x^1 + 0x^2 + \dots$.

REMARK XI.2. The degree function satisfies:

- $\deg(fg) = \deg(f) + \deg(g)$;
- $\deg(f + g) \leq \max(\deg(f), \deg(g))$;
- $\text{lc}(fg) = \text{lc}(f) \cdot \text{lc}(g)$.

The last item implies that if R is a domain then also $R[x]$ is a domain since $fg = 0$ implies $\text{lc}(f)\text{lc}(g) = 0$.

1. Euclidean rings

DEFINITION XI.3. A domain has a *Euclidean measure* if there is a function $\delta: R \setminus \{0\} \rightarrow \mathbb{N}$ 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 and both nonzero, 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 $|b|$.
- 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 under long 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 (which means it has a Euclidean algorithm that algorithmically allows to find $\gcd(a, b)$ as linear combination $ax + by$ of a and b with coefficients a, b in R).

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 $\gcd(a, 0) = 1 \cdot a$. 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 $a_i = b_i q_i + r_i$ and $\delta(b_i) > \delta(r_i)$, 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 gcd 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. \square

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 .

EXAMPLE XI.6. Let $a = 3x^3 + 4x + 3, b = 4x^2 + 2x + 4$ in $\mathbb{Z}/5\mathbb{Z}[x]$. (Strictly speaking, I should write bars over every number, but maybe we can live without that for a moment).

We have (keep in mind that $2 \cdot 3 = 4 \cdot 4 = 1$)

$$\begin{aligned}
 a &= \frac{3}{4}xb + (3x^3 + 4x + 3 - (3x^3 + 4x^2 + 3x)) \\
 &= \frac{3}{4}xb + x^2 + x + 3 \\
 &= \frac{3}{4}b + \frac{1}{4}b + (x^2 + x + 3 - (x^2 + 3x + 1)) \\
 &= (\frac{3}{4}x + \frac{1}{4})b + (3x + 2).
 \end{aligned}$$

It follows that $\gcd(a, b) = \gcd(b, 3x + 2)$. We next divide b by $c := 3x + 2$.

$$\begin{aligned}
 b &= 3xc + (4x^2 + 2x + 4 - (3x(3x + 2))) \\
 &= 3xc + (4x^2 + 2x + 4 - (4x^2 + x)) \\
 &= 3xc + x + 4 = 3xc + 2c.
 \end{aligned}$$

It follows now that $\gcd(a, b) = \gcd(b, c) = c$.

EXAMPLE XI.7. Let's compute \gcd of $x^{10} - 1$ and of $x^6 - 1$ in $\mathbb{Q}[x]$.

$$\begin{aligned}
 x^{10} - 1 &= x^4(x^6 - 1) + (x^4 - 1). \\
 x^6 - 1 &= x^2(x^4 - 1) + (x^2 - 1). \\
 x^4 - 1 &= x^2(x^2 - 1) + (x^2 - 1),
 \end{aligned}$$

and so

$$x^4 - 1 = x^2(x^2 - 1) + 1(x^2 - 1) = (x^2 + 1)(x^2 - 1) + 0.$$

So, $\gcd(x^{10} - 1, x^6 - 1) = \gcd(x^6 - 1, x^4 - 1) = \gcd(x^4 - 1, x^2 - 1) = \gcd(x^2 - 1, 0) = x^2 - 1$.

Moreover, $x^2 - 1 = (x^6 - 1) - x^2(x^4 - 1) = (x^6 - 1) - x^2((x^{10} - 1) - x^4(x^6 - 1)) = (1 - x^4)(x^6 - 1) - x^2(x^{10} - 1)$ gives the \gcd as linear combination, by working our computations backwards.

In Euclidean rings, a lot is like in \mathbb{Z} .

THEOREM XI.8. *In Euclidean rings, all ideals are principal.*

PROOF. This goes parallel to the proof of Theorem X.15, 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, \dots of generators of an ideal we have $\langle a_1, a_2, \dots \rangle = \langle \gcd(a_1, a_2), a_3, a_4, \dots \rangle = \langle \gcd(a_1, a_2, a_3), a_4, a_5, \dots \rangle$ and $\delta(a_1) \geq \delta(\gcd(a_1, a_2)) \geq \delta(\gcd(a_1, a_2, a_3)) \geq \dots$ since these \gcd 's divide one the next.

This decreasing sequence has to level off, since it can't go down indefinitely. It follows that from some index i onward, $\delta(\gcd(a_1, \dots, a_i)) = \delta(\gcd(a_1, \dots, a_j))$ for all $j > i$. This in turn implies that a_{i+1}, \dots are all divisible by $g = \gcd(a_1, \dots, a_i)$. But then the ideal generated by all a_k is the same as the ideal of just a_1, \dots, a_i , and this is just the set of all multiples of g . \square

EXAMPLE XI.9. Let I be the ideal generated by $f = x^3 + x^2 + x - 3$ and $g = x^2 + x - 2$ in $\mathbb{Q}[x]$. With the Euclidean algorithm one computes that $\gcd(f, g) = x - 1$. It follows that I consists exactly of the multiples of its generator $x - 1$.

THEOREM XI.10. *In a Euclidean ring, "prime" and "irreducible" are the same concepts.*

PROOF. Recall that prime things are always irreducible (but not always the other way round). So we need to show that here 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 $ax + py = g$. 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. \square

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 Euclideaness, 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 generated by 2 and x , $I = \{2a + xb \mid a, b \in R\}$. Let us 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 ,

$$2 = fg, \quad x = fh,$$

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 this 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, $x = fh = f k x$ 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. \square

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

$$[R \text{ is a ED}] \Rightarrow [R \text{ is a PID}] \Rightarrow [R \text{ is a UFD}] \Rightarrow [R \text{ is a domain}].$$

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.

DEFINITION XI.13. A *norm* on a ring R is a function $N: R \rightarrow \mathbb{N}$ for which $N(rs) = N(r) \cdot N(s)$, and $[N(r) = 0] \Leftrightarrow [r = 0]$, and $[N(r) = 1] \Leftrightarrow [r \text{ is a unit}]$.

(We met this concept earlier in HW).

EXAMPLE XI.14. (1) The ring $R = \mathbb{Z}[\sqrt{-5}]$ 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 factorizations 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 in R as $2 = rs$ 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 $N(r) = 2$. 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*.

EXERCISE XI.15. In the ring $\mathbb{Z}/7\mathbb{Z}[x]$, compute the gcd between $\bar{3}x^3 + \bar{1}x^2 - \bar{4}x + \bar{1}$ and $\bar{1}x^3 - \bar{4}x^2 + \bar{1}x - \bar{4}$.

DEFINITION XI.16. 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 multiplication 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 $z \mapsto z/1$.

LEMMA XI.17. If $f: R \rightarrow S$ is a ring morphism, and if \mathfrak{p} is a prime ideal of S , then the preimage $f^{-1}(\mathfrak{p})$, the set of elements of R that land in \mathfrak{p} under f , is a prime ideal of R .

PROOF. Let $\mathfrak{P} = f^{-1}(\mathfrak{p})$ and suppose it is not a prime ideal. That suggests the existence of $a, b \in R$, not in \mathfrak{P} , but with $ab \in \mathfrak{P}$. If ab is in \mathfrak{P} , then (since all elements of \mathfrak{P} are sent to \mathfrak{p} under f), $f(ab) \in \mathfrak{p}$. But $f(ab) = f(a) \cdot f(b)$ shows that $f(a) \cdot f(b)$ is in the prime ideal \mathfrak{p} . It follows that one of $f(a), f(b)$ must be in \mathfrak{p} . Say, $f(a) \in \mathfrak{p}$. Then a is sent to $f(a) \in \mathfrak{p}$ under f , and so $a \in \mathfrak{P}$. That is exactly what we needed to show that \mathfrak{P} is prime. \square

REMARK XI.18. Note that the preimage of a maximal ideal need not be maximal. For example, in the inclusion $\mathbb{Z} \hookrightarrow \mathbb{Q}$ the zero ideal of \mathbb{Q} is maximal (as \mathbb{Q} is a field), but its preimage, the zero ideal in \mathbb{Z} is not maximal.

The notion of a ring of fractions comes up in the proof of the next result.

THEOREM XI.19 (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 $x^2 - 1$ 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.20. *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 . And with it one can show that polynomials in $R[x]$ that factor in $\mathbb{K}[x]$ will also factor in $R[x]$.

With the lemma in hand one can prove:

LEMMA XI.21. *If R is a UFD then a factorization of $f \in R[x]$ in $\mathbb{K}[x]$ always can be reworked into factorization in $R[x]$.*

Uniqueness of the factorization in $R[x]$ is then rather easy. You might look at the proof of the Gauß Lemma in any textbook, if you are curious. \square

COROLLARY XI.22 (Rational Root Test). *Let $f(x) = c_n x^n + \dots + c_1 x + c_0$ be a polynomial with coefficients in \mathbb{Z} . If $f(x)$ has any root in \mathbb{Q} , then the root must have the form $\frac{m}{(c_n)^k}$ for some k -th power of c_n .*

In particular, if the leading coefficient of f is ± 1 then any rational root of $f(x)$ must be integer, and any such root must divide the constant term a_0 .

PROOF. Suppose f has a rational root $r = a/b$. If b is not ± 1 , then some prime power p^t divides b and then p^{tn} divides the denominator of $c_n r^n$. But at best, $p^{t(n-1)}$ can divide the denominator of any other term of $f(r)$ since the coefficients c_i are integers. In particular, the sum of all trailing terms in $f(r)$ will be a fraction whose denominator is at best divided by $p^{t(n-1)}$. It follows that the coefficient c_n must erase at least t powers of p . In other words, $p^t | c_n$. The first part then follows.

If the lead coefficient is ± 1 then it cannot erase any fraction, and so r^n must not be a proper fraction, hence must be an integer and so also r is an integer.

Let r be an integer root of f , so $f(r) = 0$. Then $c_n r^n + \dots + c_1 r = -c_0$ is a multiple of r . \square

This corollary simplifies searches for rational roots. For example, the only *possible* rational roots of $x^8 - 34x^6 + 25x^5 - 2x^4 + 127x^2 - 89x - 1$ are $1, -1$. And plugging them in one finds they are not, so f has no rational root.

If one searches for common solutions of several polynomials, one can play them off one against the other.

EXAMPLE XI.23. Let us look for the common rational roots of $f(x) = x^x + x^2 - 3x - 6$ and of $g(x) = x^5 - x^4 - 11x^3 - 13x^2 + 9x + 15$. The lead term in both cases is 1, so rational roots can only be divisors of the constant terms. The constant term of f says that a rational root can only be one of $\pm 1, \pm 2, \pm 3, \pm 6$. That of g says that rational roots must be one of $\pm 1, \pm 3, \pm 5, \pm 15$. In both sets are only $\pm 1, \pm 3$. It is easy to see that ± 1 are not roots of $f(x)$ as 6 is bigger than the sums of the other coefficients. One checks explicitly that ± 3 are not roots either, so no common rational root exists.

With the help of the Euclidean algorithm one computes $\gcd(f, g) = x^2 + x + 3$, which does not factor over \mathbb{Q} .

CHAPTER XII

Week 11/12: 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 every $r \in R$ there is an *evaluation morphism*

$$\varepsilon_r: R[x] \rightarrow R$$

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 $f(x) = a_0 + a_1x + \dots + a_dx^d$, d the degree of f , and suppose $\varepsilon_r(f) = 0$. Since $\varepsilon_r(x - r) = 0$ 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 an f_k that is constant, you can't keep the iteration going.

We have $\varepsilon_r(f) = \varepsilon_r(f_1) = \varepsilon_r(f_2) = \dots = \varepsilon_r(f_k)$ and f_k is a constant. But as a constant, plugging in has no effect. So $\varepsilon_r(f)$ is the number f_k . This says that the remainder f_k that you get when you divide $f(x)$ by $x - r$ is precisely the value of $f(x)$ at input $x = r$. (By *division of f by g* for any two polynomials f, g we mean the finding of an equation $f(x) = g(x) \cdot h(x) + r(r)$ with r of minimal possible degree.)

We have now shown:

LEMMA XII.1 (Remainder Lemma). *Let $f(x) \in R[x]$ and choose $r \in R$. The value $f(r)$ is the remainder of division of $f(x)$ by $x - r$.* \square

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 multiplicity.*

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 $f(x) = (x - r_1)(x - r_2) \cdots (x - r_k)f_k(x)$, 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 that is not on our list. The evaluation map ε_r sends $f(x) = (x - r_1)(x - r_2) \cdots (x - r_k)f_k(x)$ to $0 = f(r) = (r - r_1) \cdots (r - r_k)f_k(x)$. Since R is a domain, this product equaling zero can only happen if one of the factors is zero. So, r must be one of the r_i (since we assumed that f_k has no roots in R), and so the number of roots (counted with multiplicity) cannot exceed the degree of f . \square

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 irreducible. 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 finitely 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 $f(x) \bmod p$ is irreducible and of the same degree as f , then f is irreducible in $\mathbb{Z}[x]$ and even in $\mathbb{Q}[x]$.*

PROOF. In proving the Gauß Lemma we found 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 equal to $\deg(f)$. That now means that one of g, h has degree zero before we went modulo p , and so is an integer.

We have shown 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 must be a unit and we are done. \square

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 homework.

Moreover, it is pretty complicated to predict whether the reduction modulo some prime p of a polynomial $f(x)$ will give you something irreducible. For example, $x^2 - 2$ is irreducible over \mathbb{Z} (since $\sqrt{-2}$ is not an integer) and factors modulo the prime p if and only if $p = 2$ or p is of the form $8k \pm 1$ for some $k \in \mathbb{N}$. And this is just a quadratic polynomial, asking for a root of 2...

(For the curious: it is not so simple to find out why 2 is a square modulo primes of the form $8k \pm 1$. The idea is: suppose $p = 8k + 1$. Then $U(p)$ has $8k$ people in it, and is a cyclic group. So there is some element that generates the group, and its k -th power a is nonzero in $U(p)$ and satisfies has order 8. Take $b = a + a^{-1}$. This is not zero because the equation $a + a^{-1} = 0$ is tantamount to $a^2 + 1 = 0$, so that $a^2 = -1$ which would indicate that a has order 4, not 8. Consider $b^2 = (a + a^{-1})^2 = a^2 + 2 + a^{-2} = a^{-2}(a^4 + 2a^2 + 1)$. As the order of a is 8, $a^4 = -1$. Then it follows that $b^2 = 2a^2/a^2 = 2$. If $p = 8k - 1$, the issue is more tricky, as there will be no a with $a^8 = 1$ while $a^4 \neq 1$. (The order of $U(p)$ will be $8k - 2$, which is not even divisible by 4). So, make a field extension that catches a primitive 8th root a of 1. The point is that $b = a + a^{-1}$ has on one side $b^2 = 2$ as before, and on the other side satisfies $b^p = b$ as one can easily check using that $a^8 = 1$. But the latter property implies that b , which was only known to be in the extension, actually lives in $\mathbb{Z}/p\mathbb{Z}$, since $GF(p, 1)$ are the people with $b^p = b$ in any extension of $\mathbb{Z}/p\mathbb{Z}$.

REMARK XII.7. If you want to solve a quadratic equation $x^2 + ax + b = 0$ then the standard trick is to complete the square: $(x + a/2)^2 + q = a^2/4 = 0$, and then take roots. It is useful to note that this can be done not just with rational numbers, but also over any field where 2 is not zero. For example, in $\mathbb{Z}/11\mathbb{Z}[x]$, $x^2 + x + 9 = 0$ can be solved via the quadratic formula: $x_{1,2} = -(1/2) \pm \sqrt{(1/4) - 9}$. The thing is of course, that one must make sense of what is written. Indeed, since $2 \cdot 6 = 11 + 1$, 2 and 6 are inverses modulo 11. Therefore, $-1/2 = -6 = 5$ modulo 11. So we now can say $x_{1,2} = 5 \pm \sqrt{25 - 9} = 5 \pm 4$. Thus, the roots are in this case 1 and 9 modulo 11.

It can of course happen that the root cannot be solved. For example, $x^2 + x + 8 = 0$ does not factor in $\mathbb{Z}/11\mathbb{Z}$, since the root formula says that $x_{1,2} = 5 \pm \sqrt{25 - 8}$ and 17 is not a square modulo 11. (This means, that no number of the sort $17 + 11k$, $k \in \mathbb{N}$, is the square of an integer. Checking such statements can be tedious for large modulo numbers, and is subject of Gauss' Quadratic Reciprocity Law, which you might want to look up.)

It would be good to have a test for irreducibility based on reduction modulo a prime where one can see right away that it will work.

THEOREM XII.8 (Eisenstein Criterion). *Let $f(x) = a_0 + \dots + a_d x^d \in \mathbb{Z}[x]$ be of degree d such that the gcd of all the coefficients is 1. If there is a prime $p \in \mathbb{Z}$ with*

- (1) $p \nmid a_d$,
- (2) $p \mid a_i$ for $i = 0, \dots, d - 1$,
- (3) $p^2 \nmid a_0$,

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 $p \nmid a_d$, the degree of $f \bmod p$ is also d . In fact, since p divides all coefficients except the top one, $f(x) \bmod p = x^d \bmod p$. So, in $\mathbb{Z}/p\mathbb{Z}[x]$, which

is a Euclidean domain, and thus a UFD, the only factorizations of $f \bmod p = (g \bmod p)(h \bmod p)$ are of the type $g = x^{d-k} \bmod p, h = x^k \bmod p$. 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 $\mathbb{Z}[x]$. Irreducibility over $\mathbb{Q}[x]$ follows now from the Gauss Lemma. \square

EXAMPLE XII.9. Let $f(x) = x^n - p$. It fits the Eisenstein conditions, so is irreducible over $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.

DEFINITION XII.10. For $n \in \mathbb{N}$ let $\Phi_n(x) \in \mathbb{Z}[x]$ be the n -th cyclotomic polynomial, defined as the factor of $x^n - 1$ that do not divide $x^m - 1$ for any earlier $m < n$.

- EXAMPLE XII.11. (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 $x^p - 1 = (x - 1)(x^{p-1} + x^{p-2} + \dots + x + 1)$ and it turns out that the second factor is irreducible, and thus

$$\Phi_p(x) = x^{p-1} + x^{p-2} + \dots + x + 1 \quad \text{if } p \text{ is prime.}$$

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 $((y + 1)^p - 1)/y = y^{p-1} + \binom{p}{1}y^{p-2} + \dots + \binom{p}{k}y^{p-k-1} + \dots + \binom{p}{p-1}y^0$. Clearly, $\binom{p}{p-1} = p$ is a multiple of p , and 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 and preserves the property of having interesting factors).

LEMMA XII.12. For $k \in \mathbb{N}$ with $0 < k < p$, the number $c_{p,k} := \binom{p}{k}$ is a multiple of p .

PROOF. You learned in discrete math that the number $c_{p,k}$ is the number of ways to pick k things from p distinct things, 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 p , 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. \square

REMARK XII.13. The roots of $x^n - 1$ are the n -th roots of unity. They form a multiplicative group of size n , and are spaced out regularly along the unit circle at every $2\pi/n$. The first one, ω_n at $2\pi/n$, is clearly a generator, multiplication with ω amounts to rotation by $2\pi/n$. It follows that this group, which we denote μ_n , is cyclic of order n , and as we know such groups to be isomorphic to \mathbb{Z}/n . The generators of μ_n are the powers $(\omega_n)^k$ that satisfy $\gcd(k, n) = 1$. The corresponding roots are known as n -th primitive roots of unity. The n -th cyclotomic polynomial

has as roots exactly the n -th primitive roots of unity. You might want to check that explicitly for $n = 1, \dots, 6$.

2. Making new fields from old

Recall that a field is a ring in which every nonzero element has an inverse. 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.14. If \mathbb{F} is a field and $f(x) \in \mathbb{F}[x]$ is an irreducible polynomial, then the quotient ring $\mathbb{F}[x]/\langle f \rangle$ is the *Kronecker extension* of \mathbb{F} by f . We denote it $\text{Kron}(f, \mathbb{F})$.

LEMMA XII.15. If $f \in \mathbb{F}[x]$ is irreducible, $\text{Kron}(f, \mathbb{F})$ is actually a field.

PROOF. We need to show that every nonzero element of $\text{Kron}(f, \mathbb{F})$ has an inverse. Take a coset $\bar{g} \in \mathbb{F}[x]/\langle f \rangle$ nonzero. That means in particular, that $g \in \mathbb{F}[x]$ is not a multiple of f . Since f is irreducible, then $\gcd(f, g) = 1$. As $\mathbb{F}[x]$ is Euclidean, we can write $af + bg = 1$ for some $a, b \in \mathbb{F}[x]$. Reading this equation in $\text{Kron}(f, \mathbb{F})$ gives $\bar{g} \cdot \bar{b} = \bar{1}$ and shows that \bar{b} is the inverse of \bar{g} in $\text{Kron}(f, \mathbb{F})$. \square

EXAMPLE XII.16. If $\mathbb{F} = \mathbb{R}$ and $f(x) = x^2 + 1$ then $\mathbb{F}[x]/\langle f \rangle$ is a 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 $\overline{a + bx}$ 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, and $\text{Kron}(f, \mathbb{F}) = \mathbb{C}$ follows.

DEFINITION XII.17. If \mathbb{F} is any field, a field \mathbb{E} that contains \mathbb{F} is a *field extension* of \mathbb{F} .

Field extensions can be “large” or “small”. The notion of measure comes from linear algebra: an extension \mathbb{E} of \mathbb{F} will always be a vector space over \mathbb{F} .

DEFINITION XII.18. The dimension of an extension \mathbb{E} as vector space over \mathbb{F} is the *degree* of the extension. It can be in \mathbb{N} , or infinite.

EXAMPLE XII.19. (1) $\mathbb{R} \subseteq \mathbb{C}$ is an extension of degree 2, with basis $\{1, \sqrt{-1}\}$.

(2) $\mathbb{Q} \subseteq \mathbb{C}$ is an infinite extension (since \mathbb{Q} is countable and \mathbb{C} is not). The same is true for $\mathbb{Q} \subseteq \mathbb{R}$.

Another way to make new fields is as follows.

DEFINITION XII.20. 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 the intersection of all fields that contain \mathbb{F} and β . Clearly, $\mathbb{F}(\beta)$ is inside \mathbb{E} .

THEOREM XII.21 (Kronecker Extension Theorem). Let $f \in \mathbb{F}[x]$ be an irreducible polynomial.

The Kronecker extension $\text{Kron}(\mathbb{F}, f) = \mathbb{F}[x]/\langle f \rangle$ is a field. It can be viewed as a vector space over \mathbb{F} , and as vector space 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 $\text{Kron}(\mathbb{F}, f)$.

PROOF. By Lemma XII.15, Kronecker extensions are always fields. From the definition, it is clear that $\text{Kron}(\mathbb{F}, f)$ has an \mathbb{F} -vector space basis given by $1, x, \dots, x^{\deg f - 1}$.

Now we prove the last claim. Let us make a ring morphism $\pi: \mathbb{F}[x] \rightarrow \mathbb{F}(\beta)$ by sending x to β and any element of \mathbb{F} to itself. The kernel is the polynomials $g(x)$ for which $g(\beta) = 0$, and clearly $f(x)$ is in this kernel. Let $g(x) \in \ker(\pi)$. If $g(\beta) = f(\beta) = 0$, and if $h(x) = \gcd(f(x), g(x))$, then also $h(\beta) = 0$. (This is because $\mathbb{F}[x]$ is Euclidean, and so h is a linear combination of f and g). Since f is irreducible, this gcd can only be 1 or f itself. If it were f then f should divide g which would mean that $\bar{g} = \bar{0}$ in $\text{Kron}(f, \mathbb{F})$. In any other case, this gcd must be 1, and so 1 is a linear combination of f and g . But that cannot be since supposedly f, g both have root β , but the polynomial 1 does not. Thus, all kernel elements are multiples of f .

It follows that there is a morphism of fields, $\text{Kron}(f, \mathbb{F}) \rightarrow \mathbb{E}$, sending x to β and elements of \mathbb{F} to themselves (viewed as elements of \mathbb{E}). This morphism is injective since its kernel are the polynomials that have root β , and we already killed in $\text{Kron}(f, \mathbb{F})$ all of them since we already killed all multiples of f . So we can view $\text{Kron}(f, \mathbb{F})$ as sitting inside \mathbb{E} .

Since $\mathbb{F}(\beta)$ is supposed to be the smallest field inside \mathbb{E} containing both β and \mathbb{F} , we conclude that $\mathbb{F}(\beta)$ sits inside $\text{Kron}(f, \mathbb{F})$. On the other hand, under this morphism, the coset of x is sent to β , and all other elements of $\text{Kron}(f, \mathbb{F})$ are simply polynomials in x with coefficients in \mathbb{F} . By the field axioms (closed under \pm and \cdot), all such expression must be in $\mathbb{F}(\beta)$. In other words, $\mathbb{F}(\beta)$ can't really be smaller than $\text{Kron}(f, \mathbb{F})$. That finishes the proof of the last claim. \square

EXAMPLE XII.22. Let $\mathbb{F} = \mathbb{Z}/3\mathbb{Z}$ and choose $f(x) = x^2 + x - 1$. We study the Kronecker extension $R = \mathbb{F}[x]/\langle f \rangle$. Note that plugging in shows that f has no roots in $\mathbb{Z}/3\mathbb{Z}$ and hence must be irreducible since any factorization of a cubic requires at least one linear factor, which would point at a root.

Denote α the coset of x in R . Then $\alpha^2 + \alpha - \bar{1} = 0$ in R , which can be read as $\alpha^2 = \bar{1} - \alpha$. So, powers of α higher than the first power can be replaced by lower powers. So, R is a $\mathbb{Z}/3\mathbb{Z}$ -vector space spanned by $\bar{1}$ and α , and the elements of R are the $9 = 3^2$ expressions

$$0, \bar{1}, \bar{2}, \alpha, \bar{1} + \alpha, \bar{2} + \alpha, 2\alpha, \bar{1} + 2\alpha, \bar{2} + 2\alpha.$$

Moreover, $\alpha^2 + \alpha - \bar{1} = 0$ in R means that the polynomial $y^2 + y - 1$ has a root in R , namely α . (This kind of event 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: $y^2 + y - \bar{1} = (y - \alpha)(y + \alpha + \bar{1})$. (You might want to check this: $(y - \alpha)(y + \alpha + \bar{1}) = y^2 + y(-\alpha + \alpha + \bar{1}) + (-\alpha)(\alpha + \bar{1})$. The linear term is fine, and for the constant term observe that it equals $-\alpha^2 - \alpha$. But as $\alpha^2 + \alpha - \bar{1} = 0$, this constant term is $-\bar{1}$.)

The moral of the example is :if you make a Kronecker extension, a previously irreducible polynomial will acquire at least one new root in the Kronecker type extension field. This suggests that you can, iteratively, 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.23. 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 $\bar{1} \cdot \alpha^3 + \bar{0} \cdot \alpha^2 + \bar{0} \cdot \alpha^1 - \bar{2} \cdot \alpha^0 = \bar{0}$, 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 $\bar{1}, \alpha, \alpha^2$.

We know, that $f(y)$ has a root in R called α . So, $(y - \alpha)$ divides $f(y)$, and one computes by long division that $y^3 - 2 = (y - \alpha)(y^2 + \alpha y + \alpha^2)$.

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 which we view as elements of the intermediate extension R . Now take this expression and plug it in 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)

$$\alpha^2(b^2 + 2ac + b + 1) + \alpha(2a^2 + 2bc + c) + \bar{1}(c^2 + 4ab + 2a).$$

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 we know there are no rational numbers a, c such that $(2a/c)^3 = 2$. It follows that such a, b, c as stipulated cannot exist, and so the quadric does not have any further roots in R . If one wants to split $x^3 - 2$ completely, one should now repeat this process with R and $y^2 + \alpha y + \alpha^2$ as inputs. This leads to

$$\frac{\mathbb{Q}[x, y]}{\langle x^3 - 2, y^2 + xy + x^2 \rangle}$$

which as a \mathbb{Q} -vector space is spanned by $\bar{1}, \bar{x}, \bar{x}^2, \bar{y}, \bar{xy}, \bar{x^2y}$. The polynomial $t^3 - 2$ can be factored as $(t - \bar{x})(t - \bar{y})(t - \bar{x} - \bar{y})$.

CHAPTER XIII

Week 12/13: Splitting fields and extension towers

1. Splitting fields

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 with coefficients from \mathbb{E} .

We let $\text{Split}(f)$ stand for any splitting field of f that is minimal with this respect.

Let us prove that such things exist, along the lines of Example XII.23:

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 as much as you can. Then you are left with proving that you can split any irreducible polynomial. So assume from the start that f is irreducible.

Let \mathbb{E} be the Kronecker extension $\text{Kron}(\mathbb{F}, f)$. 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. \square

Note that this also proves that $\text{Split}(f)$ 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 $\text{Split}(f)$, each version is isomorphic to any other, and that any such isomorphism can be arranged to identify the copy of \mathbb{F} that each splitting field contains. One says that the various versions of $\text{Split}(f)$ are *isomorphic over \mathbb{F}* .

Note also that the worst case scenario is that we need to make Kronecker extensions to polynomials of degrees $\deg(f), \deg(f) - 1, \dots, 3, 2$.

EXAMPLE XIII.3. If $f(x) = x^2 - 2$ with $\mathbb{F} = \mathbb{Q}$, then $\text{Split}(f) = \text{Kron}(\mathbb{Q}, f)$. Indeed, $x^2 - 2$ does not split over \mathbb{Q} , so $\text{Split}(f)$ is not \mathbb{Q} . On the other hand, $\text{Kron}(\mathbb{Q}, f)$ contains one root β of f , and dividing $x^2 - 2$ by $x - \beta$ leaves a linear polynomial, namely $x + \beta$. So, f splits over $\text{Kron}(\mathbb{Q}, f)$. It follows that we can find a copy of $\text{Split}(f)$ inside $\text{Kron}(\mathbb{Q}, f)$.

On the other hand, as a vector space over \mathbb{Q} , $\text{Kron}(\mathbb{Q}, f)$ is 2-dimensional, with basis $\{1, \beta\}$. So $\text{Split}(f)$, 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 $\text{Kron}(\mathbb{Q}, f)$. Since we know that $\text{Split}(f)$ can't be \mathbb{Q} , it must be $\text{Kron}(\mathbb{Q}, f)$.

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 $\text{Kron}(\mathbb{F}, f)[x]$.

EXAMPLE XIII.4. As we have seen in Example XII.23, the splitting field for $x^3 - 2$ is a 6-dimensional vector space over \mathbb{Q} , since the quadratic factor $x^2 + \beta x + x^2$ obtained by dividing $x^3 - 2$ by $x - \beta$ remains irreducible over $\text{Kron}(\mathbb{Q}, x^3 - 2)$. 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 $\text{Kron}(\mathbb{F}, f)$ is the degree of the polynomial f , since $\text{Kron}(\mathbb{F}, f)$ has the basis $1, x, x^2, \dots, x^{\deg f - 1}$.

EXAMPLE XIII.6. $\text{Kron}(\mathbb{Q}, x^3 - 2)$ is degree 3 over \mathbb{Q} ; $\text{Split}(\mathbb{Q}, x^3 - 2)$ is degree 6 over \mathbb{Q} ; $\text{Split}(\mathbb{F}, f)$ 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 at all, and is thus irreducible.

Let β be the Kronecker root for f in $\mathbb{K} := \text{Kron}(\mathbb{F}, f)$. 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 of them you find as follows.

LEMMA XIII.8. If β is a root of $f(x)$ and the field has characteristic p , then β^p is also a root.

PROOF. In characteristic p , freshman's dream for p -th powers holds: $f(x)^p = f(x^p)$, since the Binomial Theorem assures that all the "missing" terms are multiples of p and this count as zero. Then plug in $x = \beta$: you get $0 = f(\beta)^p = f(\beta^p)$. \square

On the last root in the example: we proved that β being root implies $\gamma := \beta^p$ being root. But by the same argument, then γ^p is a root as well. Let's check what that means here. Since $\gamma = \beta^2$ here, $\gamma^p = \beta^4 = \beta(\beta^3) = \beta(-\beta - \bar{1})$ since $\beta^3 + \beta + \bar{1} = 0$. Since $\beta(-\beta - \bar{1}) = -\beta^2 - \beta = \beta^2 + \beta$ as we are in characteristic 2, this last expression is a root of $x^3 + x + 1$.

Alternatively, you can argue as follows: 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.

2. Roots with multiplicity

Some polynomials, like $x^2 + 2x - 3 = (x+3)(x-1)$, have all their roots distinct. For polynomials $x^2 + bx + c$ of degree 2, we know that this happens exactly when the *discriminant* $b^2 - 4c$ 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 $b = 6, c = 9$ we find that $x^2 + bx + c = x^2 + 6x + 9 = (x+3)^2$ 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 $x^3 + bx^2 + cx + d$, the discriminant is $18bcd - 4b^3d + b^2c^2 - 4c^3 - 27d^2$.

DEFINITION XIII.9. Let $f(x) \in \mathbb{F}[x]$ be any polynomial with coefficient in the field \mathbb{F} . If β is a root of $f(x)$ we say that it has *multiplicity* m provided that $(x-\beta)^m$ divides $f(x)$, but $(x-\beta)^{m+1}$ does not divide $f(x)$.

As one sees by examples, it is often interesting to take a polynomial over one ring R and ask for its roots in a bigger ring. For example, we know that we need to look inside \mathbb{C} for roots of $x^2 + 1$ since $x^2 + 1$ does not factor over \mathbb{Z}, \mathbb{Q} or \mathbb{R} .

On the other hand, some strange things can happen in finite characteristic.

EXAMPLE XIII.10. 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 . In particular, $x^2 + 1$ factors over $\mathbb{Z}/2\mathbb{Z}$ and has double root $1 + 2\mathbb{Z}$.

Stranger yet, there are polynomials that are irreducible and yet have multiple roots in a suitable larger ring.

EXAMPLE XIII.11. 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 $(x - \sqrt[3]{t})^3 = x^3 - 3x^2\sqrt[3]{t} + 3x\sqrt[3]{t}^2 - t = x^3 - t$ 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. 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 "dash" in the theorem below denotes taking the derivative according to the rules of calculus: product rule and power rule. (You might ask "What other rules might I possibly want to use for a derivative, 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.12. *Let \mathbb{F} be a field and $f(x) \in \mathbb{F}[x]$. Then $f(x)$ has a double or higher root in some (perhaps unknown) 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 $f'(x) = 2(x - r)g(x) + (x - r)^2 g'(x) = (x - r) \cdot [2g(x) + (x - r) \cdot g'(x)]$ 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; in other words, assume that some irreducible polynomial $g(x)$ of positive degree divides both f and f' . Then let \mathbb{E} be the Kronecker extension $\text{Kron}(f, \mathbb{F})$. 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 $f(x) = h(x)(x - \beta)$ in the polynomial ring $\mathbb{E}[x]$. 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} . \square

REMARK XIII.13. 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 $\gcd(f, f') = \gcd(f, 0) = f$.

DEFINITION XIII.14. 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.

REMARK XIII.15. If $p(x), q(x)$ are two polynomials, then Sylvester invented a trick to predict whether p and q have a common root in any field at all. It goes like this: write $p(x) = a_d x^d + \dots + a_1 x + a_0$ and $q(x) = b_e x^e + \dots + e_1 x + e_0$. Now form a matrix

$$\begin{pmatrix} a_d & a_{d-1} & \cdots & a_0 & & & \\ & a_d & a_{d-1} & \cdots & a_0 & & \\ & & \ddots & & & \ddots & \\ & & & a_d & a_{d-1} & \cdots & a_0 \\ b_e & b_{e-1} & \cdots & b_0 & & & \\ & b_e & b_{e-1} & \cdots & b_0 & & \\ & & \ddots & & & \ddots & \\ & & & b_e & b_{e-1} & \cdots & b_0 \end{pmatrix}$$

Here we have e rows with the coefficients of p and d rows with the coefficients of b , each new copy sliding over by one. This is the *Sylvester matrix* of p, q .

The determinant of the Sylvester matrix is zero if and only if p, q have a common root (out there, somewhere), says a theorem of Sylvester. It is called the *resultant* of p and q . The resultant of p and its derivative p' is the *discriminant* of p . The discriminant of $ax^2 + bx + c$ is, for example, $a(4ac - b^2)$. Numerically, it is equal to $a_d^e \cdot b_e^d$ times the product of all differences $\alpha_i - \beta_j$ where α_i is a root of p and β_j is a root of q (taken in some extension field).

Week 13/14: Minimal polynomials and finite fields

It is time to introduce some notation regarding field and ring extensions.

DEFINITION XIV.1. 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 .

Note that $\mathbb{F}(\alpha)$ may be considerably larger than $\mathbb{F}[\alpha]$ since it also must contain all inverses to the elements of $\mathbb{F}[\alpha]$. For example, $\mathbb{Q}(\pi)$ is not just polynomials in π but also fractions of such polynomials.

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.2. If $\mathbb{F} \subseteq \mathbb{E}$ is a field extension it is called *finite* if the bigger field has a finite vector space basis over the smaller one. Whether the extension is finite or not, we write $[\mathbb{E} : \mathbb{F}]$ for the vector space dimension of \mathbb{E} over \mathbb{F} and call it the *degree of the extension*.

For example, \mathbb{C} is finite over \mathbb{R} with basis $1, \sqrt{-1}$. As another example, if f is irreducible of degree d over \mathbb{F} then $\text{Kron}(f, \mathbb{F})$ is finite, with basis $1, \bar{x}, \dots, x^{d-1}$ over \mathbb{F} (since f allows to reduce d -th and higher powers to lower ones).

DEFINITION XIV.3. Suppose $\mathbb{F} \subseteq \mathbb{E}$ is any extension, and pick $\alpha \in \mathbb{E}$. We say that α is *finite over* \mathbb{F} , or also α is *algebraic over* \mathbb{F} , if the infinitely many powers α, α^2, \dots are linearly dependent over \mathbb{F} . Otherwise, we call the α *transcendental* over \mathbb{F} .

Surely, if an extension $\mathbb{F} \subseteq \mathbb{E}$ is finite, then any $\alpha \in \mathbb{E}$ is finite over \mathbb{F} , since $\mathbb{F}(\alpha)$ is no bigger than \mathbb{E} .

For example, $\sqrt{2}$ is algebraic, and π is transcendental, over \mathbb{Q} .

If α is finite over \mathbb{F} then there is at least one nontrivial linear combination $\sum_{i=0}^d c_i \alpha^i = 0$, where all coefficients c_i come from \mathbb{F} and at least one of them is nonzero. Of all the possible such expressions, there will be exactly one that has the smallest possible degree, and then we can modify it so it has leading coefficient 1. (A polynomial with leading coefficient 1 is called *monic*).

DEFINITION XIV.4. In the context of the preceding paragraph, let $\text{minpol}_{\mathbb{F}}(\alpha)$ be the monic polynomial $\sum_{i=0}^d c_i x^i$ of minimal degree for which substituting α for x gives an expression that evaluates to zero. We call it the *minimal polynomial of* α *over* \mathbb{F} and the corresponding d the *degree of* α *over* \mathbb{F} .

For example, the minimal polynomial of $\sqrt{-1}$ over \mathbb{R} is $x^2 + 1$. Indeed, the given polynomial is monic and plugging in $\sqrt{-1}$ for x yields zero. On the other hand, a lower degree polynomial with real coefficients cannot have root $\sqrt{-1}$ since such a thing would imply that $\sqrt{-1}$ should be a real number. One can find minimal polynomials by doing explicit computations.

Recall that we defined the vector space dimension of \mathbb{E} over \mathbb{F} as the degree of the extension, and wrote $[\mathbb{E} : \mathbb{F}]$.

LEMMA XIV.5. *For an iterated extension, $\mathbb{F} \subseteq \mathbb{F}' \subseteq \mathbb{F}''$, we have a formula*

$$[\mathbb{F}'' : \mathbb{F}] = [\mathbb{F}'' : \mathbb{F}'] \cdot [\mathbb{F}' : \mathbb{F}].$$

This is actually 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 $\mathbb{Q}[\sqrt[2]{2}, \sqrt[3]{2}, \dots, \sqrt[k]{2}]$ 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 to have *some* minimal polynomial. (My guess is that it has degree equal to a number of about 40 digits).

There are many field extensions that are not algebraic. For example, $\mathbb{Q} \subseteq \mathbb{R}$ is not finite, because the powers $1, \pi, \pi^2, \dots$ 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 $\mathbb{Q}[x]$. The algebraic extensions are the ones that we will focus on.

It is immediate that finite extensions are algebraic, but it does not quite work the other way. For example, let $\overline{\mathbb{Q}}$ be the field generated by all roots to all polynomials with rational coefficients. Then this is (very large and not finite, but) algebraic over \mathbb{Q} . Let's try to digest that. The main part of the magic lies in the following

LEMMA XIV.6. *If $\mathbb{F} \subseteq \mathbb{E}$ is an extension such that there is a list of elements $\{e_i\}_{i \in I}$ of \mathbb{E} , indexed by some set I , such that*

- $\mathbb{E} = \mathbb{F}(\{e_i\}_{i \in I})$, and
- each e_i is algebraic over \mathbb{F} ,

then every element of \mathbb{E} is algebraic over \mathbb{F} .

PROOF. Take any element e of \mathbb{E} . Then since \mathbb{E} is the field extension generated by the set $\{e_i\}_{i \in I}$, e can be written as a fraction of two polynomials in the e_i . As such it only involves finitely many of the e_i . Since each e_i is algebraic (and hence $\mathbb{F}(e_i)$ finite) over \mathbb{F} so is $\mathbb{F}(e_1, \dots, e_k)$ for any k . It follows that the sub-extension $\mathbb{F}(e)$ is also finite. It follows that the infinitely many powers of e must satisfy a linear relation in the finite-dimensional \mathbb{F} -vector space $\mathbb{F}(e)$. \square

For an algebraic but non-finite example, look at the field that you get when you start with \mathbb{Q} and then throw in all n -th roots of 2 ($n = 2, 3, \dots$). It is algebraic but not finite. That 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 unpleasanties 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. We already discussed that this will happen in the lemma above, and here we offer a different view.

Another example of a finite extension is the following. \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 $\text{Kron}(\alpha, \mathbb{F}) = \mathbb{F}[x]/\langle f \rangle$ which is a vector space of dimension $\deg(f)$ over \mathbb{F} spanned by $1, x, \dots, x^{\deg(f)-1}$.

It follows from all these thoughts that:

LEMMA XIV.7. *For $f \in \mathbb{F}[x]$, since $\deg(f)$ is finite, one can iterate Kronecker extensions to find a splitting field that sits inside an extension of \mathbb{F} of degree equal to the factorial of $\deg(f)$.* \square

EXAMPLE XIV.8. If $\mathbb{F} = \mathbb{Q}$ and $f = x^3 - 2$, then we can build splitting fields one step at a time. $\mathbb{F}' := \text{Kron}(\mathbb{F}, f) = \mathbb{Q}[x]/\langle f \rangle$ 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 $\mathbb{F}'' := \text{Kron}(\mathbb{F}', x^2 + x\beta + \beta^2)$ 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 $\mathbb{F}'' = \text{Split}(\mathbb{F}, f)$.

EXAMPLE XIV.9. It is becoming clear that it is useful to know irreducible polynomials over our favorite fields. Let's look at the irreducible ones of degrees 1, 2, 3, 4 over $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$.

Every linear polynomial (with nonzero coefficient for x) is irreducible; these are here $x + \bar{0}$ and $x + \bar{1}$.

A quadric polynomial factors if and only if it has a root. The only quadric (with a nonzero coefficient for x^2) with no root is $x^2 + x + \bar{1}$. Indeed, we need the constant coefficient to be nonzero (or else x would divide the polynomial) and $x^2 + \bar{1} = (x + \bar{1})^2$ since we are in characteristic 2. Note as an aside that $f(x) = f(x + 1)$. This is forced on us, since f is the only irreducible quadric over $\mathbb{Z}/2\mathbb{Z}$, and $f(x)$ is reducible if and only $f(x + c)$ is for any constant.

A cubic will factor if and only if it has a root, and the cubics without root are $x^3 + x + \bar{1}$ and $x^3 + x^2 + \bar{1}$. Indeed, again we need a nonzero constant term, and one checks easily that the two given ones have no root. Having no root, that cannot factor since a degree three polynomial must be irreducible or have at least one linear factor. In reverse, the only missing cubic is $x^3 + x^2 + x + \bar{1}$ and that factors as $(x + \bar{1})^3$ as one checks.

For quartics, the irreducible ones must have nonzero leading and constant term. They also cannot have exactly 4 nonzero terms, since such a polynomial gives $\bar{4} = \bar{0}$ when you plug in $\bar{1}$. The candidates are $x^4 + x^3 + \bar{1}$, $x^4 + x^2 + \bar{1}$, $x^4 + x + \bar{1}$ and $x^4 + x^3 + x^2 + x + \bar{1}$. We already know they have no root, and hence no linear factor. But they might be the product of two irreducible quadrics. However, there is only one irreducible quadric, namely $q = x^2 + x + \bar{1}$ and $q^2 = x^4 + x^2 + \bar{1}$. It follows that there are three irreducible quartics, $x^4 + x^3 + \bar{1}$, $x^4 + x + \bar{1}$ and $x^4 + x^3 + x^2 + x + \bar{1}$. The former two trade places under $x \mapsto x + 1$ whereas the last one stays unchanged.

It should be stressed that over $\mathbb{Z}/3\mathbb{Z}$ this all looks quite different.

2. Finite Fields

EXAMPLE XIV.10. Let $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$ and take $f(x) = x^3 + x + 1$, $g(x) = x^3 + x^2 + 1$; both are irreducible.

$\text{Kron}(\mathbb{F}, f) = \mathbb{F}[x]/\langle f \rangle$ has $8 = 2^3$ elements $\{\bar{0}, \bar{1}, \bar{x}, \bar{x} + \bar{1}, \bar{x}^2, \bar{x}^2 + \bar{1}, \bar{x}^2 + \bar{x}, \bar{x}^2 + \bar{x} + \bar{1}\}$. The same is true for $\text{Kron}(\mathbb{F}, g) = \mathbb{F}[y]/\langle g \rangle$, 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 the Kronecker root for f , so $\alpha = x \bmod \langle f \rangle$. Write β for the Kronecker root of g , so $\beta = y \bmod \langle g \rangle$.

From the Kronecker construction we know that α is a root of f ; who else might be a root? Division gives $f(x) : (x - \alpha) = x^2 + \alpha x + (\alpha^2 + 1) =: f_2(x)$. 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 $f(\alpha) = 0$.)

Another way to look at this is that since we are in characteristic 2, whenever α is a root, then so is α^2 and then also $(\alpha^2)^2$ which we can reduce to $\alpha^2 + \alpha$ using that $f(\alpha) = 0$.

Either way, it follows that $\text{Kron}(\mathbb{F}, f)$ is actually the splitting field of f over \mathbb{F} .

¹

Now plug $y - 1$ into $f(x)$. You get $(y - 1)^3 + (y - 1) + 1 = y^3 + y^2 + 1 = g(y)$. So, g has roots equal to those of f shifted up by 1. They are therefore $\alpha + 1, \alpha^2 + 1, \alpha^2 + \alpha + 1$.

In particular, $\text{Kron}(\mathbb{F}, f)$ is also the splitting field for $g(x)$. So there is no real difference between $\alpha + 1$ in $\text{Kron}(\mathbb{F}, f)$ and $\beta \in \text{Kron}(\mathbb{F}, g)$. There is only one degree 3 extension of \mathbb{F} .

REMARK XIV.11. Here is an amusing computation that explains the previous example. Any degree 3 extension \mathbb{E} 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 in \mathbb{E} 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 $x^8 - x$. Of course, x and $x - 1$ are linear and $\mathbb{Z}/2\mathbb{Z}$ already contains the roots to $x - 1$ and $x - 0$.

So, any field of 8 elements contains all roots to f , and all roots to g . In conclusion, there is only one field with 8 elements. Moreover, since \mathbb{E} is the Kronecker extension to \mathbb{F} along *any* irreducible cubic, all of these irreducible cubics must divide $x^8 - x$. And indeed, $x^8 - x = x(x + 1)(x^3 + x + 1)(x^3 + x^2 + 1)$.

We now generalize this example.

¹You might want to consider the following example to realize how unusual this is. Over the reals, $x^3 + x + 1$ has no critical point since $3x^2 + 1$ is always positive. So, it has at most one real root. Since we are looking at a cubic, it also has at least one real root. We conclude that $x^3 + x + 1$ has exactly one real root r , and two properly complex ones c_1, c_2 . It follows that the Kronecker extension $\mathbb{Q}(r)$ is not the splitting field of $x^3 + x + 1$ since it cannot contain c_1, c_2 .

THEOREM XIV.12. Let $\mathbb{F} = \mathbb{Z}/p\mathbb{Z}$ be the field with p elements, p a prime number.

(1) Choose $e \in \mathbb{N}_{\geq 1}$. Then there exists an irreducible polynomial of degree e , and hence a field $\text{GF}(p, e)$ with p^e elements given by the Kronecker extension of \mathbb{F} by any such polynomial.

(2) All elements $a \in \text{GF}(p, e)$ are roots of the polynomial $x^{p^e} - x$, and hence $x^{p^e} - x$ completely splits over $\text{GF}(p, e)$. In other words, $\text{GF}(p, e)$ is the splitting field of the polynomial $x^{p^e} - x$.

(3) Every element of $\text{GF}(p, e)$ is equal to its p^e -th power, and so every element has a p^e -th root in $\text{GF}(p, e)$.

(4) One has $\text{GF}(p, 1) = \mathbb{Z}/p\mathbb{Z}$.

(5) The degree of the field extension $[\text{GF}(p, e) : \text{GF}(p, 1)]$ is e .

(6) In consequence, $\text{GF}(p, e)$ is the Kronecker extension $\text{Kron}(\mathbb{F}, g(x))$ for every irreducible polynomial of degree e over $\mathbb{Z}/p\mathbb{Z}$.

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). Inside this field, denote the set of roots for $x^{p^e} - x$ by $\text{GF}(p, e)$. We plan to show that this subset of the splitting field is a field all by itself.

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 $\text{GF}(p, e)$ is closed under $+$ and $-$, and under multiplication and division. So, this set of roots is a field.²

Splitting: By construction, f has all its roots in $\text{GF}(p, e)$, so $\text{GF}(p, e)$ is the smallest field over which f splits. It follows that $\text{GF}(p, e)$ is the splitting field.

Powers: Since $f(a) = 0$ for all $a \in \text{GF}(p, e)$, each element agrees with its own p^e -th power.

(1) 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 already in $\mathbb{Z}/p\mathbb{Z}$ and we need no extension.

Size and Degree: 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 $\text{GF}(p, e)$. So, $\text{GF}(p, e)$ is full of single roots of $f(x)$ and so must have p^e elements (it is a splitting field!). It follows that $[\text{GF}(p, e) : (\mathbb{Z}/p\mathbb{Z})] = e$.

Uniqueness: 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 $\text{Kron}(\mathbb{F}, g) = \text{GF}(p, e)$. □

COROLLARY XIV.13. The nonzero elements $U(p, e)$ of $\text{GF}(p, e)$ 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 finite Abelian group, $U(p, e)$ can be written as $(\mathbb{Z}/a_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/a_k\mathbb{Z})$ with elementary divisors $a_1|a_2|\cdots|a_k$, by FTFGAG. If q is an element of this product

²This 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 $x^2 - 2$...

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 all elements of $\text{GF}(p, e)$ are roots to $x^{a_k+1} - x$. But such a polynomial can have only $a_k + 1$ roots, and we know $\text{GF}(p, e)$ is the set of these roots, and there are p^e such roots. So $p^e = a_k + 1$. So $a_k = p^e - 1$. But $a_1 \cdot a_2 \cdots a_k$ should be $p^e - 1 = |U(p, e)|$, and that means that $k = 1$ and so $U(p, e)$ is cyclic, isomorphic to $(\mathbb{Z}/(p^e - 1)\mathbb{Z}, +)$. \square

It is natural to ask when finite fields sit inside one another.

EXAMPLE XIV.14. Let $p = 2$ and take $f(x) = x^4 + x + 1$. Since $f(\bar{0}) = f(\bar{1}) = \bar{1} \in \mathbb{Z}/2\mathbb{Z}$, 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 $x^2 + x + 1$ does not divide $f(x)$. So $f(x)$ is irreducible and $\text{Kron}(\mathbb{Z}/2\mathbb{Z}, f)$ has $2^4 = 16$ elements.

By the theorem, $\text{Kron}(\mathbb{Z}/2\mathbb{Z}, f) = \text{GF}(2, 4)$.

Let α be the Kronecker root and compute explicitly: $(\alpha^2 + \alpha + 1)^2 + (\alpha^2 + \alpha + 1) + 1 = 0$ in $\text{Kron}(\mathbb{Z}/2\mathbb{Z}, f)$. This says that $\alpha^2 + \alpha + 1$ is a root in $\text{GF}(2, 4)$ of the irreducible quadric $x^2 + x + 1$. In particular, $\text{GF}(2, 4)$ contains $\text{Kron}(\mathbb{Z}/2\mathbb{Z}, x^2 + x + 1) = \text{GF}(2, 2)$.

EXAMPLE XIV.15. Let $p = 2$ and take $f(x) = x^3 + x + 1$, an irreducible cubic in $\mathbb{Z}/2\mathbb{Z}[x]$. Take $\text{GF}(p, 3)$ to be its splitting field $\text{Kron}(\mathbb{Z}/2\mathbb{Z}, f)$. The 8 elements of $\text{GF}(2, 3)$ 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 $\text{GF}(2, 2)$ in here. Since elements of $\text{GF}(2, 2)$ 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 $\text{GF}(2, 3)$. 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 except for $\mathbb{Z}/2\mathbb{Z} = \text{GF}(2, 1)$ can be in $\text{GF}(2, 3)$ and $\text{GF}(2, 2)$ simultaneously.

We now investigate the general case.

THEOREM XIV.16. $\text{GF}(p, e)$ sits inside $\text{GF}(p', e')$ if and only if $p = p'$ and $e|e'$.

PROOF. Suppose $\text{GF}(p, e) \subseteq \text{GF}(p', e')$. In $\text{GF}(p, e)$ we have that $1 + \dots + 1$ (p copies) gives 0. In $\text{GF}(p', e')$, this is so with p' copies. If $p \neq p'$ then $\gcd(p, p') = 1$ copies of 1 also amount to zero by the familiar argument involving linear combinations and the Euclidean algorithm. We conclude $p = p'$ is necessary.

Suppose now $\text{GF}(p, e)$ sits inside $\text{GF}(p, e')$. Then $\text{GF}(p, e')$ is a vector space over $\text{GF}(p, e)$. Since one field has p^e elements, and the other has $p^{e'}$ elements, $p^{e'}$ should be a power of p^e , and that means that e' should be a multiple of e .

Conversely, suppose $e|e'$ and look for $\text{GF}(p, e)$ inside $\text{GF}(p, e')$. Write $e' = de$ and calculate

$$x^{p^{e'}} - x = (x^{p^e} - x)(x^{p^{(d-1)e}} + x^{p^{(d-2)e}} + \dots + x^{p^{1e}} + x^{p^0}).$$

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. \square

It is a natural question to ask "if $\text{GF}(p, e)$ is a subfield of $\text{GF}(p, e')$, how do we best identify the smaller field? (So far we only know that it must be in there somewhere).

COROLLARY XIV.17. *If $e|e'$, the subfield $\text{GF}(p, e)$ inside $\text{GF}(p, e')$ consists of exactly those elements of $\text{GF}(p, e')$ that satisfy $x^{p^e} = x$.*

One can obtain elements in $\text{GF}(p, e)$ by raising any element of $\text{GF}(p, e')$ to the power $(p^{e'} - 1)/(p^e - 1)$.

PROOF. That $\text{GF}(p, e)$ is inside $\text{GF}(p, e')$ comes from the preceding theorem. Since any element in any version of $\text{GF}(p, e)$ is a root of $x^{p^e} - x$, selecting the ones that do this is the right strategy.

By Corollary XIV.13, $U(p, e')$ is cyclic of order $p^{e'} - 1$. Since $U(p, e)$ of size $p^e - 1$ sits inside $U(p, e')$ it must be so that $U(p, e)$ is also cyclic and comprised of the $(p^{e'} - 1)/(p^e - 1)$ -powers of the elements of $U(p, e')$, compare Theorem III.8. \square

EXAMPLE XIV.18. Consider the containment $\text{GF}(2, 2) \subseteq \text{GF}(2, 4) = \text{Kron}(\mathbb{Z}/2\mathbb{Z}, x^4 + x + 1)$. Take the 16 elements of $\text{GF}(2, 4)$ and raise them to the power $(2^4 - 1)/(2^2 - 1) = 5$. Since $a^5 = a(a^4) = a(-a - 1) = a^2 + a$, things of this form must be elements of $\text{GF}(2, 2)$, aside from the two elements of $\mathbb{Z}/2\mathbb{Z} = \text{GF}(2, 1)$ that sit in every $\text{GF}(2, e)$. For example, if $\alpha = \bar{x}$ is the Kronecker element, then $\alpha^2 + \alpha$ is in $\text{GF}(2, 2)$, and so is $(\alpha^2 + \alpha) + (1)$ since both summands are. These must be the $2^2 = 4$ elements of $\text{GF}(2, 2)$ as it sits inside $\text{GF}(2, 4)$.

EXAMPLE XIV.19. 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

$$\text{GF}(2, 1) \subseteq \text{GF}(2, 2) \subseteq \text{GF}(2, 4) \subseteq \text{GF}(2, 8) \subseteq \text{GF}(2, 24),$$

$$\text{GF}(2, 1) \subseteq \text{GF}(2, 3) \subseteq \text{GF}(2, 6) \subseteq \text{GF}(2, 12) \subseteq \text{GF}(2, 24),$$

and additional containments $\text{GF}(2, 2) \subseteq \text{GF}(2, 6)$ and $\text{GF}(2, 4) \subseteq \text{GF}(2, 12)$.

EXAMPLE XIV.20. Let $\mathbb{F} = \text{GF}(2, 3) = \text{Kron}(\mathbb{Z}/2\mathbb{Z}, x^3 + x + 1)$. We have $(x^8 - x = (x - 0)(x - 1)(x^3 + x + 1)(x^3 + x^2 + 1))$. Let α be the Kronecker root. The minimal polynomials of the 8 elements of \mathbb{F} over $\mathbb{Z}/2\mathbb{Z}$ are:

- x for $\bar{0}$, $x + \bar{1}$ for $\bar{1}$;
- $x^3 + x + \bar{1}$ for α , α^2 and $\alpha + \alpha$;
- $x^3 + x^2 + \bar{1}$ for $\alpha + \bar{1}$, $\alpha^2 + \bar{1}$ and $\alpha^2 + \alpha + \bar{1}$.

One checks explicitly that the Frobenius (here, the squaring map) turns $\alpha \rightarrow \alpha^2 \rightarrow \alpha^2 + \alpha$, $\alpha + \bar{1} \rightarrow \alpha^2 + \bar{1} \rightarrow \alpha^2 + \alpha + \bar{1}$, and fixes both $\bar{0}$ and $\bar{1}$. This illustrates that the Frobenius moves about the roots of the same irreducible polynomials.

CHAPTER XV

Week 14: Galois

1. The Frobenius

In a ring of characteristic $p > 0$ (such as in $\text{GF}(p, e)$ or indeed any ring containing $\mathbb{Z}/p\mathbb{Z}$), we have

$$(a + b)^p = a^p + b^p$$

since the intermediate terms arising from the binomial theorem all are multiples of p , hence zero. It follows that

$$\begin{aligned} \text{Frob}: \text{GF}(p, e) &\rightarrow \text{GF}(p, e), \\ \gamma &\mapsto \gamma^p \end{aligned}$$

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 morphism.

THEOREM XV.1. (1) *The p -Frobenius (p -th power map) is a field automorphism*

$$\text{Frob}: \text{GF}(p, e) \rightarrow \text{GF}(p, e)$$

for any e .

(2) *If $e'|e$ (and $\text{GF}(p, e')$ therefore sits inside $\text{GF}(p, e)$) then the Frobenius sends elements of this subfield $\text{GF}(p, e')$ into the subfield.*

(3) *The e -fold iteration of the Frobenius on $\text{GF}(p, e)$ is the identity map. One can interpret this statement as the existence of a group morphism*

$$\begin{aligned} F: \mathbb{Z}/e\mathbb{Z} &\rightarrow \text{Aut}(\text{GF}(p, e)), \\ t \bmod e\mathbb{Z} &\mapsto \text{Frob}^t(-) \end{aligned}$$

to the field automorphism group of $\text{GF}(p, e)$.

(4) *The elements of $\text{GF}(p, e)$ that are fixed by Frob are exactly the elements of $\text{GF}(p, 1) = \mathbb{Z}/p\mathbb{Z}$. More generally, the elements of $\text{GF}(p, e)$ that are fixed under the k -fold iteration of the p -th power map are precisely the elements of $\text{GF}(p, \gcd(e, k))$. Specifically, if $e'|e$ then the fixed points in $\text{GF}(p, e)$ under $\text{Frob}^{e'}$ are exactly the elements on $\text{GF}(p, e')$.*

(5) *Suppose $\alpha \in \text{GF}(p, e)$ is the root of an irreducible 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 $\text{GF}(p, e)$ of $f(x)$. In other words,*

The orbit of $\alpha \in \text{GF}(p, e)$ under the Frobenius is the set of all roots of the minimal polynomial of α over $\mathbb{Z}/p\mathbb{Z}$.

(6) *If $k \in \mathbb{N}$, the orbit of α under $\text{Frob}^k(-)$ on $\text{GF}(p, e)$ are the roots of the minimal polynomial of α over $\text{GF}(p, \gcd(e, k))$.*

PROOF. We know that Frob is additive and multiplicative, $(\alpha \pm \alpha')^p = (\alpha^p \pm \alpha'^p)$, $(\alpha\alpha')^p = \alpha^p\alpha'^p$. If $\alpha^p = \alpha'^p$ then $(\alpha - \alpha')^p = 0$ and since a field has no zerodivisors,

$\alpha = \alpha'$. So Frob is injective. But then the image of Frob has p^e elements, and hence fills out the target field. So, Frob is bijective and hence an isomorphism, permuting the elements of $\text{GF}(p, e)$.

All elements in $\text{GF}(p, e)$ satisfy $\alpha^{p^e} = \alpha$. If $e'|e$, the elements of $\text{GF}(p, e')$ inside $\text{GF}(p, e)$ 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, $\text{Frob}(\alpha)$ belongs to $\text{GF}(p, e')$ again. So the isomorphisms that the Frobenius induces on the various fields $\text{GF}(p, -)$ are compatible with inclusions.

The e -fold iteration of Frob sends $\alpha \in \text{GF}(p, e)$ to $\alpha^{p^e} = \alpha$, so it is the identity on $\text{GF}(p, e)$. It follows that we can read Frob as a group action of $\mathbb{Z}/e\mathbb{Z}$ on the elements of $\text{GF}(p, e)$ via $\lambda(t \bmod e\mathbb{Z}, \alpha) \mapsto \text{Frob}^t(\alpha) = \alpha^{p^t}$.

Let $g = \gcd(k, e)$, and view it as linear combination $g = ak + be$ with $a, b \in \mathbb{Z}$. Then if $\text{Frob}^k(\alpha) = \alpha$, and we always have $\text{Frob}^e(\alpha) = \alpha$, we conclude that $\text{Frob}^g(\alpha) = \text{Frob}^{ak+be}(\alpha) = \text{Frob}^{ak}(\text{Frob}^{be}(\alpha))$ is the result of applying b iterates of $\text{Frob}^e(-)$ to α , followed by a iterates of $\text{Frob}^k(-)$. None of them changes α , so $\text{Frob}^g(\alpha) = \alpha$. It follows that α belongs to $\text{GF}(p, \gcd(g, e))$. On the other hand, since g divides k , $\text{GF}(p, g) \subseteq \text{GF}(p, k)$. It follows the two fields are equal.¹

Since Frob^e is the identity on $\text{GF}(p, e)$, Frob^k acts the same way as $\text{Frob}^{\gcd(e, k)}$. (You should make sure you believe this before going on. It can be seen via the Euclidean algorithm: $\text{Frob}^k(\alpha) = \text{Frob}^{k-e}(\alpha)$; 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 $\mathbb{Z}/p\mathbb{Z} = \text{GF}(p, 1)$. If $\text{Frob}^k(\alpha) = \alpha$ then α is a root of $x^{p^{\gcd(e, k)}} - x$ and therefore belongs to $\text{GF}(p, \gcd(k, e))$.

Suppose $e'|e$, so $\text{GF}(p, e')$ sits inside $\text{GF}(p, e)$. If $f(\alpha) = 0$ and the coefficients of f come from a field $\text{GF}(p, e')$ then the coefficients c_i satisfy $\text{Frob}^{e'}(c_i) = c_i$. 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, $\text{Frob}^{e'}(\alpha)$ is a root to the same polynomial as α . Since the degree of α over $\text{GF}(p, 1)$ is the product of the degree of α over $\text{GF}(p, e')$ with e/e' , it follows that the degree of the minimal polynomial of α over $\text{GF}(p, e')$ is e/e' . This implies that iterating $\text{Frob}^{e'}$ on α makes it cycle through all the roots of f . (If it did not move through all roots, one could take the roots it moves through and construct a minimal polynomial of lower degree, which cannot be). \square

EXAMPLE XV.2. Let $p = 3$ and $e = 4$. There are 81 elements in $\text{GF}(3, 4)$. An example of an irreducible polynomial of degree 4 is $x^4 - x^3 - 1$, so we can view $\text{GF}(3, 4)$ as $\text{Kron}(\mathbb{Z}/3\mathbb{Z}, x^4 - x^3 - 1) = \mathbb{Z}/3\mathbb{Z}[x]/\langle x^4 - x^3 - 1 \rangle$.

A slightly horrendous calculation shows that $x^3 - x$ factors as $(x)(x-1)(x+1)$ times $(x^2 + 1) * (x^2 - x - 1) * (x^2 + x - 1)$ times

$$\begin{aligned} & (x^4 - x - 1) * (x^4 + x - 1) * (x^4 - x^2 - 1) * (x^4 + x^2 - 1) * (x^4 + x^2 - x + 1) * \\ & \quad * (x^4 + x^2 + x + 1) * (x^4 - x^3 - 1) * (x^4 - x^3 + x + 1) * (x^4 - x^3 - x^2 + x - 1) * \\ & \quad * (x^4 - x^3 + x^2 + 1) * (x^4 - x^3 + x^2 - x + 1) * (x^4 - x^3 + x^2 + x - 1) * (x^4 + x^3 - 1) * \\ & \quad * (x^4 + x^3 - x + 1) * (x^4 + x^3 - x^2 - x - 1) * (x^4 + x^3 + x^2 + 1) * \\ & \quad * (x^4 + x^3 + x^2 - x - 1) * (x^4 + x^3 + x^2 + x + 1). \end{aligned}$$

¹It is likely that one of a, b is negative, say $a < 0$. In that case, “ a applications of Frob” refers to $|a|$ applications of the inverse map of the Frobenius, which we have shown to be a bijection.

The polynomials on this list are irreducible. (Recall: over $\mathbb{Z}/2\mathbb{Z}$ there were only 3 irreducible quartics. Here we have a lot more. (In fact, there are formulae that prescribe the number of irreducible polynomials of degree d over $\mathbb{Z}/p\mathbb{Z}$.) In particular, there are 3 irreducible linear polynomials, 3 irreducible quadratics, 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 copy of $\text{GF}(3, 3)$ is inside $\text{GF}(3, 4)$.)

Note that $3 \cdot 1 + 3 \cdot 2 + 18 \cdot 4 = 3 + 6 + 72 = 81$ is the degree of $x^{3^4} - x$, as it should.

So the roots of $x^{81} - x$, which are precisely the elements of $\text{GF}(3, 4)$, come in 3 types:

- there are three elements of $\text{GF}(3, 1)$: encoded as the roots to $x = 0, x - 1 = 0, x + 1 = 0$;
- there are six elements of $\text{GF}(3, 2)$ that are not in $\text{GF}(3, 1)$: they come in pairs of the roots of $x^2 + 1 = 0, x^2 - x - 1 = 0, x^2 + x - 1 = 0$;
- all other elements in $\text{GF}(3, 4)$ are not in $\text{GF}(3, 2)$: these come in quadruplets as the roots of the 18 irreducible quartics. There are 72 total of these.

Let us take the irreducible quadric $x^2 + 1$, and let α be the Kronecker root of $\text{GF}(3, 4) = \text{Kron}(\mathbb{Z}/3\mathbb{Z}, x^4 - x^3 - 1)$ for $f(x) = x^4 - x^3 - 1$. In other words, $\alpha = \bar{x}$. Let's try to find a copy of $\text{GF}(3, 2)$ inside this field. This would require, for example, finding the roots to $x^2 + 1$ (one of the three irreducible quadratics 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 $\text{GF}(3, 4)$ the copy of $\text{GF}(3, 2)$ 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:

$$\begin{aligned} 0^3 &= 0, \\ 1^3 &= 1, \\ 2^3 &= 8 = 2, \\ (\alpha^3 + \alpha^2 + 1)^3 &= \alpha^9 + \alpha^6 + 1 = \dots = 2\alpha^3 + 2\alpha^2 + 2, \\ ((\alpha^3 + \alpha^2 + 1) + 1)^3 &= \dots = 2\alpha^3 + 2\alpha^2 + 0, \\ ((\alpha^3 + \alpha^2 + 1) + 2)^3 &= \dots = 2\alpha^3 + 2\alpha^2 + 1, \\ (2(\alpha^3 + \alpha^2 + 1))^3 &= \dots = \alpha^3 + \alpha^2 + 1, \\ (2(\alpha^3 + \alpha^2 + 1) + 1)^3 &= \dots = \alpha^3 + \alpha^2 + 2, \\ (2(\alpha^3 + \alpha^2 + 1) + 2)^3 &= \dots = \alpha^3 + \alpha^2 + 0 \end{aligned}$$

So, the third-power map flips them about in pairs. The 3 pairs correspond to the roots of the 3 irreducible quadratics above.

Now let us look what the Frobenius does to general elements of $\text{GF}(3, 4)$, those that do not live in smaller fields. As a starter, we look at what happens to α itself under iterates of Frob. By definition, $\text{Frob}(\alpha) = \alpha^3$ and we leave it like that since we can't rewrite polynomials of degree less than four.

Then $\text{Frob}(\text{Frob}(\alpha)) = \alpha^9$ 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 $(x^4 - x^3 - 1)^3 = (x^3)^4 - (x^3)^3 - 1 = 0$, 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 $\text{GF}(3, 4)$. Three elements are fixed points, there are three 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).

From another angle, the nonzero elements of $\text{GF}(3, 4)$ form a cyclic group with multiplication by Corollary XIV.13. Note that $|\text{GF}(3, 4)| = 3^4 = 80 + 1$. So, with multiplication as operation, we can view $U(\text{GF}(3, 4))$ as the hours on a clock with 80 hours total.

The Frobenius takes 3rd powers here, so that corresponds to multiplying a given hour by 3. The elements for which $\text{Frob}^4(\alpha) = \alpha$ are the ones for which $\alpha^{81} = \alpha$, and since we have 80 elements in the group, this is all of them.

The elements for which $\text{Frob}^1(\alpha) = \alpha$ are those whose square is the element $\bar{1}$, and so apart from 0, the elements of $\text{GF}(3, 4)$ that are fixed under the Frobenius are $\pm\bar{1}$.

The elements for which $\text{Frob}^2(\alpha) = \alpha$ are the ones that satisfy $\alpha^9 = \alpha$. Apart from 0, these are those among the 80 units that are 10-th powers of a generator. And in particular, there are 8 of them, apart from the element 0.

The elements that satisfy $\text{Frob}^3(\alpha) = \alpha$ must have $\alpha^{27} = \alpha$, or $\alpha^{26} = 1$. But since $26 = 2 \cdot 13$ and 13 is coprime to 3, that is exactly the same ones for which $\alpha^2 = 1$, which are exactly those that satisfy $\text{Frob}^1(\alpha) = \alpha$.

2. Review

- Chapter 9:

- commutative ring
- zerodivisor, domain
- extension ring, $\mathbb{Z}[\sqrt{-1}]$, $\mathbb{Z}[\sqrt{a}]$, $\mathbb{Q}[\sqrt{a}]$, $\mathbb{Z}/m\mathbb{Z}[\sqrt{a}]$
- characteristic
- unit, field

- Chapter 10:

- ideal, proper ideal
- ring morphism
- kernel, factor/quotient ring

- structure theorem on quotient rings
- prime ideal, maximal ideal, fields
- PID, UFD
 - Chapter 11:
 - long division
 - Euclidean measure, algorithm, ring
 - Gauss Lemma
 - Rational Root Test
 - Chapter 12:
 - evaluation morphism
 - root of polynomial, Remainder Lemma
 - in a domain, number of roots of a polynomial bounded by degree
 - irreducible polynomial
 - Eisenstein criterion
 - cyclotomic polynomial, roots of unity and generators for the group of unit roots
 - field extensions, Kronecker construction, $\mathbb{F}(\beta)$
 - Chapter 13:
 - splitting field
 - extension degree
 - theorem on extension towers
 - multiplicity of root
 - derivative test, (in)separable extension
 - Chapter 14:
 - minimal polynomial of β over \mathbb{F}
 - algebraic and transcendental extensions
 - finite and infinite extensions
 - finite fields
 - $\text{GF}(p, e) = \text{Split}(\mathbb{Z}/p\mathbb{Z}, x^{p^e} - x)$, containments, and the structure theorem
 - $U(p, e)$ is cyclic
 - Chapter 15:
 - the Frobenius automorphism and its iterates
 - the structure theorem on the action of Frobenius power Frob^k on $\text{GF}(p, e)$

Applications

In this section we will discuss how the theory of fields can help in answering several (very) old questions and problems in mathematics.

3. Geometric Constructions

3.1. The Delian Problem. Delos is an island in the Cyclades archipelago in Greece. According to Plutarch, a plague struck Delos and the Oracle of Delphi announced that in order for the plague to be overcome, Delos needed to build a new altar, double the size of the existing one, for the Oracle. The existing altar was a regular cube. The citizens asked Plato what the point of the assignment was, and he interpreted it as follows.

PROBLEM XV.3. *Suppose a cube of side length a is given, together with a ruler, a compass, and enough paper to work on. Construct a line segment b such that the cube with side length b has double the volume of the original.*

Obviously, b/a should be the cubic root of 2. The relation to field theory goes as follows.

Since we are given nothing but the cube (and in particular no measuring device), one can read this as assume that the length of the side of the cube is the unit of measurements, and construct a segment of length equal to cubic root of 2. So, from now on we assume that $a = 1$.

Note that with ruler and compass it is relatively simple to find segments of length $2, 3, \dots$ by drawing a long line and repeatedly marking intervals of length 1. Almost as easily, one can take fractions geometrically. Say, you wanted to make $3/5$. Draw two lines from a common center C , on one mark three units distance from C , on the other 5. Now draw a line L that links the two extreme marked points on the two rays. Recall that one can construct parallels to any line through any given point with ruler and compass. Draw a line parallel to L through the first marked point on the line with 5 marked points. This line meets the other ray at a distance $3/5$ from C . It follows we can construct segments of all rational lengths.

Suppose you took two line segments of length x and 1, and join them on a common line L at the point J . Let M be the midpoint of the joined segment. Then draw a circle C about M of radius $(x + 1)/2$ and draw a line L' perpendicular to L through the point J . It meets the circle C in two points, whose distance is exactly \sqrt{x} from J . (Because a theorem of Euclid says that this distance squared is the product of the lengths of the two segments on the line L). Thus, ruler and compass allow to make line segments of length \sqrt{x} for any x that has already been constructed.

So, let's say \mathbb{F} is the collection of all real numbers α so that either α or $-\alpha$ is the length of a line segment that we can construct with ruler and compass. We

know that iterated square roots, or square roots of linear expressions in numbers that we can construct, are all in there. What we want to know is whether the cubic root of 2 is in there. For that we have to understand what sorts of constructions are possible.

From the get-go, we can introduce a coordinate system, and we know that we start with all points that have rational coordinates. With ruler and compass, one can make new points by

- (1) intersecting 2 lines passing each through already constructed points,
- (2) intersecting 2 circles about already constructed centers with already constructed radii,
- (3) a line and a circle like above.

In the first and last case, we are solving a degree 2 equation or a linear equation, so the new lengths we have access to are in a degree 2 extension of the currently available lengths. In the middle case, we are looking for solutions to a system of equations like

$$\begin{aligned}(x-a)^2 + (y-b)^2 &= c^2, \\ (x-d)^2 + (y-e)^2 &= f^2.\end{aligned}$$

Subtracting the equations gives a linear equation in x, y and if you use it to replace y in the first equation then the new x appears as the solution of a degree 2 equation. Altogether:

The collection of constructible coordinates after a finite number of constructions are exactly the elements of an iterated extension of \mathbb{Q} where at each stage a degree 2 extension is performed. Moreover, any extension of this type can be constructed.

We can now address the Delian Problem. Indeed, in order to solve it, we need to construct a root to $x^3 - 2 = 0$. We know that a root of $x^3 - 2$ cannot be rational, so the cubic $x^3 - 2$ is irreducible over \mathbb{Q} . That means that $\mathbb{Q}(\sqrt[3]{2})$ is a degree three extension from \mathbb{Q} . But no such field can lie inside a constructible field since all constructible fields have extension degree a power of 2, and 3 does not divide a power of 2. So, doubling the cube cannot be done by ruler and compass.

3.2. Trisecting angles. Suppose we are given an angle α drawn in the plane, in addition to a line segment of length 1. It is easy to cut the angle in half using ruler and compass (review 8th grade geometry!). Is it possible to trisect it?

In order to find the answer, note first that having the picture of an angle is the same as having line segments with length $\sin(\alpha), \cos(\alpha)$. Indeed, draw a circle about the vertex of the angle with radius 1 and then draw a perpendicular from one intersection point to the other leg of the angle.

There is a triple angle cosine formula:

$$\cos(\alpha) = 4\cos(\alpha/3) - 3\cos(\alpha/3).$$

Thus, in order to trisect α we need to be able to find a solution to $4x^3 - 3x - \cos(\alpha) = 0$.

The general expectation is that this should require a degree 3 extension, and thus not be possible. However, this requires that the given cubic be irreducible, and it is not always that. For example, we know very well that we can construct 60

degree angles, and so surely we can trisect 180 degrees. Note that then $\cos(\alpha) = -1$ and so the equation becomes $4x^3 - 3x + 1 = 0$ and factors as $(x + 1)(2x - 1)^2$. In particular, $\cos(180/3) = 1/2$ requires no extension at all. In other words, some angles can be trisected.

One can show by example that some angles cannot be trisected with ruler and compass. For example, 60 degree angles cannot be trisected. Here is the check: $\cos(60) = 1/2$, so we are looking at the roots of $4x^3 - 3x - 1/2 = 0$, or $8x^3 - 6x + 1 = 0$. Clearly x can't be 0, so we can introduce $y = 1/x$ and then find that our equation amounts, by dividing out x^3 , to $y^3 + 6y^2 - 8 = 0$. This is a monic cubic with integer coefficients. The rational root theorem says that its rational roots must be divisors of 8. Trying them all out, this cubic is shown to be irreducible. So, $\cos(20)$ is a degree-3-extension away from \mathbb{Q} and thus cannot lie inside any constructible field.

In particular, there cannot be a general mechanism to trisect angles, since if there were one, it should work on $2\pi/6$ and the above shows that it cannot work in this case.

3.3. Regular n -gons. We learn in 8th grade how to construct regular triangles, quadrangles, and hexagons with ruler and compass. ("Regular" means: all sides equal, all angles equal). What about 5, 7, ...?

If we view the vertices of a (suitably sized) regular n -gon as the complex roots of the equation $x^n = 1$ then we see that we really need the roots to $x^{n-1} + \dots + x + 1 = 0$. Moreover, it allows the following observation. Suppose we have roots constructed to $x^n = 1$ and to $x^m = 1$. Then we also can construct those to $x^a = 1$ where $a = \text{lcm}(m, n)$. Indeed, if you assume first that m, n are coprime, then this means we have produced angles $2\pi/m$ and $2\pi/n$. There are $c, d \in \mathbb{Z}$ with $mc + nd = \text{gcd}(m, n)$, which under division by mn gives $c/n + d/m = 1/\text{lcm}(m, n)$. So, going c steps with angle $2\pi/n$ counterclockwise and then d steps counterclockwise with angle $2\pi/m$ we arrive at an angle of $2\pi/mn$. We thus reduce from general natural n to prime powers.

Now go back to $x^{n-1} + \dots + x + 1 = 0$. Finding roots of this requires a field extension of degree $n - 1$, or maybe a multiple of $n - 1$. Since we can only do degree 2 in any step, $n - 1$ needs to be a power of 2. That is a necessary requirement, but it may not suffice. The trouble is that not every extension of degree pq can be viewed as a 2-stage extension, a degree p extension followed by a degree q extension.²

Gauss proved that every regular n -gon can be constructed which a) has $n - 1$ of the form 2^{2^c} for some $c \in \mathbb{N}$, or b) is a product of such (unequal) numbers n as in part a), times an arbitrary power of 2. Wantzel proved later that a regular n -gon can be constructed only in these cases; Gauss had already stated the "only if".

Letting $c = 0, 1, 2, 3, 4$ we find primes 3, 5, 17, 257, 65537. These number are called Fermat primes (if they are prime). The 5 displayed ones are the only known ones.

For example, it should be possible to construct a regular pentagon. Here is what you do. We want to solve $x^4 + x^3 + x^2 + x + 1 = 0$. In order to do constructin, we need to break this degree 4 extension into two quadratic ones. Let ω be the first primitive 5-th root of unity, $\omega = \cos(2\pi/5) + \sqrt{-1} \sin(2\pi/5)$. Then $y := \omega + \omega^4$ is purely real (make a picture!). Note that $y^2 = \omega^2 + \omega^8 + 2\omega\omega^4 = \omega^2 + \omega^3 + 2 = (\omega^2 + \omega^3 + 1) + 1$.

²Over finite fields this does actually happen, since there is only one field of p^e elements. But over \mathbb{Q} there are many different extensions of the same degree.

Since $\omega^4 + \dots + \omega + 1 = 0$, we get $y^2 = -y + 1$. Solving this geometrically give $y = 2\cos(2\pi/5)$, and so we have one coordinate of ω . To get the other, make an extension to solve $s^2 + (y/2)^2 - 1 = 0$ for s (the sine-cosine relation). That makes a total degree 4 extension and we have what we need to draw ω .

The construction of the regular pentagon was known to the Greeks, at least. Gauss, when he was 19, showed that you can construct a 17-gon.

4. Solving equations by radicals

It is an interesting question to ponder how one can describe the roots $\lambda_1, \dots, \lambda_n$ of a polynomial

$$f(x) = x^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$$

in terms of the coefficients. Let \mathbb{F} be the field $\mathbb{Q}(a_n, \dots, a_0)$. If f is linear, then the field of solutions $\mathbb{F}(\{\lambda_i\}_1^n)$ is just \mathbb{F} . If f is a quadric, then

$$\lambda_{1,2} = \frac{-a_1}{2} \pm \sqrt{\frac{a_1^2 - 4a_0}{4}}.$$

We interpret this to mean that as “the roots can be obtained from the coefficients of f by a) field operations, b) taking square roots.

If f is a cubic $x^3 + a_2x^2 + a_1x + a_0$, the first thing to do is to shift $y = x + a_2/3$ and to rewrite f in terms of y . The result is of the shape $y^3 + ay + b$, called *compressed form*. Fundamentally, a rational cubic can have one or three real roots (and in unfortunate circumstances 2, one of which is a double root). Which of these arises is predicted by the *discriminant* $\Delta(f) := (a/3)^3 + (b/2)^2$. If Δ is negative, there will be three distinct real roots; if it is positive there will be just one real root; if it is zero there will be a double and a single real root, or one triple real root. In the case $\Delta < 0$, let z be the cubic root of $-b/2 + \sqrt{\Delta}$. If R is the real part of z and I the imaginary part of z then the roots of f are $2R, -R \pm I\sqrt{3}$. If $\Delta > 0$ then one real root is $(-b/2 + \sqrt{\Delta})^{1/3} + (-b/2 - \sqrt{\Delta})^{1/3}$. One can now divide x minus this root out of $f(x)$, and solve the remaining quadric.

We interpret this as “the roots of a cubic equation can be expressed in terms of the coefficients, involving field calculations and 2nd and 3rd roots. These ideas seem to go back to at least del Ferro (around 1500), but Tartaglia and Cardano are often mentioned in this context.

For degree 4 equations (quartics), one first shifts to form a depressed quartic. There is a somewhat monstrous formula to express the roots of such quartic in the terms of the roots of an associated cubic polynomial whose coefficients arise from those of f by taking two square roots. So again, the root can be expressed in terms of the coefficients with the help of field operations, and the taking of 2nd, 3rd, 4th roots. This is usually credited to Cardano, but was known earlier to Ferrari (around 1550).

The natural question that arises is whether it is always possible to express the solutions to a polynomial in terms of its coefficients, using only field operations and the process of taking radicals (= (higher) roots). The concept of degree of an extension will not be helpful, because we want this for all n , and taking a radical like solving $x^k = a$ is a degree k extension. Much of the answer lies in the next theorem.

DEFINITION XV.4. Let $\mathbb{F} \subseteq \mathbb{E}$ be a field extension. We write $\text{Gal}(\mathbb{E}, \mathbb{F})$ for the set of all field automorphisms $\mathbb{E} \rightarrow \mathbb{E}$ that leave \mathbb{F} fixed, element by element.

For example, $\text{Gal}(\mathbb{C}, \mathbb{R})$ has two elements, the identity of \mathbb{C} , and complex conjugation. (Any automorphism of \mathbb{C} that fixes \mathbb{R} also fixes the equation $x^2 + 1 = 0$ and so also fixes the collection of roots to this equation. (It does not need to fix the roots individually, since they are not in \mathbb{R}). Since $x^2 + 1 = 0$ has 2 roots, there can be at most $2!$ isomorphisms of \mathbb{C} that fix \mathbb{R} . Indeed, $2!$ is the number of permutations on 2 elements, and if you know where $\sqrt{-1}$ goes you know the whole morphism).

Note that $\text{Gal}(\mathbb{E}, \mathbb{F})$ is a group, the operation being composition of morphisms.

THEOREM XV.5. *Let $\mathbb{F} \subseteq \mathbb{E}$ be an extension and assume that \mathbb{E} is the splitting field of some polynomial $f \in \mathbb{F}[x]$ for which $\gcd(f, f') = 1$. (It is immaterial, which polynomial). In particular, \mathbb{E} is finite over \mathbb{F} .*

Suppose \mathbb{K} is an intermediate field $\mathbb{F} \subseteq \mathbb{K} \subseteq \mathbb{E}$. To it one may associate the group $\text{Gal}(\mathbb{E}, \mathbb{K})$ of automorphisms of \mathbb{E} that leave \mathbb{K} fixed. Obviously, $\text{Gal}(\mathbb{E}, \mathbb{K}) \subseteq \text{Gal}(\mathbb{E}, \mathbb{F})$ since membership in the former group comes with more conditions.

On the other hand, suppose G is a subgroup of $\text{Gal}(\mathbb{E}, \mathbb{F})$. Then the set of all elements of \mathbb{E} that are fixed under all elements of G turns out to be a field. Obviously, it is between \mathbb{E} and \mathbb{F} .

The Galois correspondence says that this is a 1-1 order reversing bijection:

$$\begin{aligned} G \subseteq \text{Gal}(\mathbb{E}, \mathbb{F}) &\mapsto \{\text{fixed points of } G \text{ in } \mathbb{E}\} \\ \text{Fix}_{\mathbb{E}}(G) &\leftarrow \mathbb{E} \supseteq \mathbb{K} \subseteq \mathbb{F}. \end{aligned}$$

In this correspondence, the subgroup G of $\text{Gal}(\mathbb{E}, \mathbb{F})$ is normal if and only if the corresponding intermediate field $\mathbb{K} = \text{Fix}_{\mathbb{E}}(G)$ is a splitting field of some polynomial in $\mathbb{F}[x]$. If that is indeed the case, then the three Galois groups that arise satisfy

$$\text{Gal}(\mathbb{E}, \mathbb{F}) / \text{Gal}(\mathbb{E}, \mathbb{K}) = \text{Gal}(\mathbb{K}, \mathbb{F}).$$

Now suppose that $\mathbb{E} \subseteq \mathbb{F}$ comes about as an iteration of *radical extensions*, extensions of the sort $\mathbb{K} \rightsquigarrow \text{Split}(\mathbb{K}, x^k - a)$ for some $a \in \mathbb{K}$. The Galois group to this extension for $a = 1$ is the group of automorphisms of the group of k -th roots of unity, and as such isomorphic to $\mathbb{Z}/\phi(k)\mathbb{Z}$, where $\phi(k)$ is the Euler ϕ -function that counts cosets mod k that are relatively prime to k . In particular, this Galois group is cyclic and thus Abelian.

If \mathbb{K} already contains the k -th roots of unity, then the Galois group to the extension $\text{Split}(\mathbb{K}, x^k - a)$ can be shown to be Abelian. It follows that *if an extension is a sequence of extensions each of which is solving an equation $x^k = a$ for various k and a , then the resulting Galois group contains a sequence of subgroups*

$$\{1\} = G_0 \subsetneq G_1 \subsetneq \dots \subsetneq G_t = \text{Gal}(\mathbb{E}, \mathbb{F})$$

such that each G_{i-1} is normal in G_i , and each quotient G_i/G_{i-1} is Abelian. (This consequence requires some thought on intersecting filtrations with subgroups).

Now consider the case of a quintic, $x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0$ over \mathbb{Q} and let \mathbb{E} be the splitting field. One can show that for randomly chosen rational coefficients, $\text{Gal}(\mathbb{E}, \mathbb{Q})$ is the group S_5 of permutations on the 5 roots. (It means that splitting a general quintic requires consecutive Kronecker extensions, one of degree 5, one of degree 4, one of the degree 3 and finally one of degree 2). If one could solve for the roots in terms of radicals of the coefficients, then S_5 should allow for a chain of subgroups as above. However, a purely group-theoretic result shows that the only normal subgroups of S_5 are S_5 itself, the alternating

group A_5 , and the trivial group $\{1\}$. This, the last step of such a chain must be $G_{t-1} = A_5 \subsetneq S_n = G_t$. And a similar group-theoretic result says that A_5 is simple (it has no normal subgroups except itself and the trivial subgroup). So, at the finest, our chain is

$$\{1\} \subsetneq A_5 \subsetneq S_5.$$

But A_5 is not Abelian, so cannot be achieved as a radical extension. Thus, a general quintic has no chance to have its roots expressible by radicals.

Various thoughts

4.1. 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.6. To get a feeling, let's try to divide 12 by 6 but modulo 8. The quotient, let's call it \bar{a} , should live in $\mathbb{Z}/8\mathbb{Z}$ and have the property that $\bar{a} \cdot \bar{6} = \bar{12}$. Of course, \bar{a} could be $\bar{2}$.

But if you list the multiples of $\bar{6}$ you find:

$$\begin{array}{llll} \bar{0} \cdot \bar{6} = \bar{0}, & \bar{1} \cdot \bar{6} = \bar{6}, & \bar{2} \cdot \bar{6} = \bar{4}, & \bar{3} \cdot \bar{6} = \bar{2}, \\ \bar{4} \cdot \bar{6} = \bar{0}, & \bar{5} \cdot \bar{6} = \bar{6}, & \bar{6} \cdot \bar{6} = \bar{4}, & \bar{7} \cdot \bar{6} = \bar{2}. \end{array}$$

So we see that there are actually *two* different cosets that compete for being a quotient $\bar{12}/\bar{6}$, namely $\bar{2}$ and $\bar{6}$. This comes from the fact that we can think of $\bar{12}$ also as $\bar{4}$.

Moreover, one can see that the people in $\mathbb{Z}/8\mathbb{Z}$ split into two classes, the cosets that are multiples of $\bar{6}$ and those that are not, where each coset that shows up at all as multiple of $\bar{6}$ shows exactly twice. \diamond

EXAMPLE XV.7. 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 $\bar{5}$ in $\mathbb{Z}/8\mathbb{Z}$ we find

$$\begin{array}{llll} \bar{0} \cdot \bar{5} = \bar{0}, & \bar{1} \cdot \bar{5} = \bar{5}, & \bar{2} \cdot \bar{5} = \bar{2}, & \bar{3} \cdot \bar{5} = \bar{7}, \\ \bar{4} \cdot \bar{5} = \bar{4}, & \bar{5} \cdot \bar{5} = \bar{1}, & \bar{6} \cdot \bar{5} = \bar{6}, & \bar{7} \cdot \bar{5} = \bar{3}. \end{array}$$

So, quite against expectations, $\bar{7}/\bar{5}$ can be found in $\mathbb{Z}/8\mathbb{Z}$, and there is exactly one answer: $\bar{3}$. In fact, as one can see, any coset in $\mathbb{Z}/8\mathbb{Z}$ can be divided by $\bar{5}$ in exactly one way.

In this section we will try to understand and predict this kind of behavior. \diamond

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.6 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.8. If \bar{a}, \bar{b} are in $\mathbb{Z}/n\mathbb{Z}$, with neither a nor b divisible by n , then they are called *zerodivisors* if $\bar{a}\bar{b} = \bar{0}$.

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.9. Let $n = 6$; then $\bar{2} \cdot \bar{3} = \bar{0}$.

Indeed, in order to prepare what is to come in a bit, let's list all multiples of $\bar{2}$:

$$\bar{2} \cdot \bar{0} = \bar{0}, \quad \bar{2} \cdot \bar{1} = \bar{2}, \quad \bar{2} \cdot \bar{2} = \bar{4}, \quad \bar{2} \cdot \bar{3} = \bar{0}, \quad \bar{2} \cdot \bar{4} = \bar{2}, \quad \bar{2} \cdot \bar{5} = \bar{4}.$$

◇

The reason that $\bar{2}$ was capable to yield $\bar{0}$ when multiplied with a nonzero coset was of course that 2 has an interesting common factor with 6. In the general case, suppose \bar{a} is a coset in $\mathbb{Z}/n\mathbb{Z}$ and we look for another element $\bar{b} \in \mathbb{Z}/n\mathbb{Z}$ such that $\bar{a}\bar{b} = \bar{0}$. 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 $\bar{a} \cdot \bar{e}$ is a multiple of $\bar{d} \cdot \bar{e} = \overline{de} = \bar{n} = \bar{0}$. But a multiple of $\bar{0}$ must be $\bar{0}$ itself.

On the other hand, pick now an a such that $\gcd(a, n) = 1$. This means by Proposition I.19 that there are integers α, β with $a\alpha + n\beta = 1$. Reading this “modulo n ”, we get $\bar{a} \cdot \bar{\alpha} + \overline{n\beta} = \bar{1}$. Naturally, $\overline{n\beta} = \bar{n} \cdot \bar{\beta} = \bar{0}$. So, $\bar{a} \cdot \bar{\alpha} = \bar{1}$. It follows that for any $b \in \mathbb{Z}$, $\bar{a} \cdot \overline{b\alpha} = \bar{b}$. 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 \bar{b} . In particular, there is only one coset that when multiplied gives $\bar{0}$ (and of course this one coset is $\bar{0}$ itself).

Putting it all together, we have proved most of the following theorem:

THEOREM XV.10. *If $\gcd(a, n) = 1$ then multiplication by a is a bijection on $\mathbb{Z}/n\mathbb{Z}$. In other words, for each $\bar{b} \in \mathbb{Z}/n\mathbb{Z}$ there is exactly one $\bar{x} \in \mathbb{Z}/n\mathbb{Z}$ such that $a \cdot \bar{x} = \bar{b}$. Yet in other words, \bar{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, \bar{a} is a zerodivisor in $\mathbb{Z}/n\mathbb{Z}$.

EXAMPLE XV.11. Let $n = 6$. The numbers a that have $\gcd(a, n) = 1$ are living in the cosets $\bar{1}$ and $\bar{5}$. Everyone is a multiple of $\bar{1}$ for obvious reasons, and everyone is also a multiple of $\bar{5}$ because $\bar{5} = -\bar{1}$.

The multiples of $\bar{2}$ are $\{\bar{0}, \bar{2}, \bar{4}\}$, and these are also exactly the multiples of $\bar{4}$. Note that each one of $\{\bar{0}, \bar{2}, \bar{4}\}$ is a multiple of both $\bar{2}$ and $\bar{4}$ in $2 = \gcd(6, 2) = \gcd(6, 4)$ ways. For example, $\bar{4} = \bar{4} \times \bar{1} = \bar{4} \times \bar{4}$ and also $\bar{4} = \bar{2} \times \bar{2} = \bar{2} \times \bar{5}$.

The multiples of $\bar{3}$ are $\bar{3}$ and $\bar{0}$, and each of $\{\bar{0}, \bar{3}\}$ arises $3 = \gcd(3, 6)$ times as multiple. For example, $\bar{3} = \bar{3} \times \bar{1} = \bar{3} \times \bar{3} = \bar{3} \times \bar{5}$. ◇

EXERCISE XV.12. For $n = 10$ and $a = 1, 2, \dots, 9$ determine

(1) which cosets in $\mathbb{Z}/10\mathbb{Z}$ are multiples of \bar{a} ;

(2) how many cosets in $\mathbb{Z}/10\mathbb{Z}$ are multiples of \bar{a} and express these numbers in terms of a and 10. ◇

Theorem XV.10 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 $\bar{2}$ one fails when the input is $\bar{1}$ or $\bar{3}$ while one gets too many suggestions when one divides $\bar{2}$ by $\bar{2}$ (namely, $\bar{1}$ and $\bar{3}$) or when one divides $\bar{0}$ by $\bar{2}$ (namely, $\bar{0}$ and $\bar{2}$).

In order to explain this behavior, we shall need the following observation:

EXERCISE XV.13. Prove that $\text{lcm}(a, b) \cdot \text{gcd}(a, b) = ab$. \diamond

Now suppose $\bar{a} \cdot \bar{x} = \bar{b}$ has at least one solution, so $ax - b$ is a multiple of n . If you added $c = n/\text{gcd}(a, n)$ to x then we calculate: $\bar{a}(x + c) = \bar{a}x + \bar{a}c = \bar{b} + \overline{(an/\text{gcd}(a, n))} = \bar{b} + \overline{\text{lcm}(a, n)} = \bar{b}$ since $\overline{\text{lcm}(a, n)} = \bar{0}$ (like any other multiple of n) represents the coset of zero. It follows that besides \bar{x} all expressions $x + i \cdot c$ are also solutions to $\bar{a}x = \bar{b}$.

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 $\bar{x}, \bar{x} + \bar{c}, \dots, \bar{x} + ((n/c) - 1)\bar{c}$ that all solve $\bar{a}x = \bar{b}$. (Of course, $n/c = \text{gcd}(a, n)$ by definition of c).

EXERCISE XV.14. Group the elements of $\mathbb{Z}/24\mathbb{Z}$ in such a way that two cosets \bar{a}, \bar{b} are in the same group exactly when their sets of multiples $\{\bar{1}a, \bar{2}a, \bar{3}a, \dots\}$ and $\{\bar{1}b, \bar{2}b, \bar{3}b, \dots\}$ agree as sets (perhaps after reordering). Describe in words each group. \diamond

REMARK XV.15. 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 $0 \leq \deg(r) \leq \deg(b) - 1$.

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. \diamond

EXERCISE XV.16. 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). \diamond

4.2. 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.10, we need to count the numbers on the list $1, \dots, n - 1$ that are coprime to n . For this, recall the Euler ϕ -function from Definition I.34.

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.

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 I.35 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 $\{0, 1, \dots, n - 1\}$ 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.17. 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 $\phi(12) = \phi(4) \cdot \phi(3)$. \diamond

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

$$(x \bmod a) = (i \bmod a) \text{ and } (x \bmod b) = (j \bmod b).$$

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 $\dots, x - 2ab, x - ab, x, x + ab, x + 2ab, \dots$. 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 $\text{lcm}(a, b) = ab/\gcd(a, b) = ab = n$. 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.18. 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 $x \bmod 36$ 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 $x \bmod 36$ 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 $x = 7 + 9\ell = 7 + 9 \cdot 3 = 34 \bmod 36$.

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 $5 = 4k \bmod 9$. (This doesn't seem so easy as before. We were rather lucky earlier, because $9 \bmod 4$ 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 $4 \cdot 8 = 32 = 5 \bmod 9$. 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. \diamond

In the following theorem, we state formally what exactly happened in the computation above, and how to accomplish it.

THEOREM XV.19 (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 $i \cdot b = 1 \bmod a$ and $j \cdot a = 1 \bmod b$. 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 $x = 2 \cdot 1 \cdot 9 + 7 \cdot 7 \cdot 4 = 18 + 196 = 214 = 34 \bmod 36$.

EXAMPLE XV.20. Let's solve

$$x = 13 \bmod 29 \text{ and } x = 8 \bmod 12.$$

Matching letters, we have $a = 29, b = 12, r = 13, s = 8$. We need to find i, j with $i \cdot 12 = 1 \bmod 29$ and $j \cdot 29 = 1 \bmod 12$. 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. \diamond

If one needs to solve three simultaneous equations,

$$x = r \bmod a; \quad x = s \bmod b; \quad x = t \bmod c,$$

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.21 (Chinese Remainder Theorem). *Let n_1, n_2, \dots, n_t be pairwise coprime numbers. Choose values a_1, \dots, 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 $N_i \cdot x_i = 1 \pmod{n_i}$. Then*

$$x = \sum_{i=1}^t x_i N_i a_i.$$

EXERCISE XV.22. Solve the simultaneous equations

- (1) $x = 1 \pmod{3}, x = 2 \pmod{5}, x = 3 \pmod{7}$.
- (2) $x = 2 \pmod{3}, x = 2 \pmod{5}, x = 3 \pmod{7}$.
- (3) $x = 1 \pmod{9}, x = 2 \pmod{5}, x = 3 \pmod{7}$.

◇

It now follows from the Chinese Remainder theorem that:

COROLLARY XV.23. *If you factor $n = a_1 \cdots a_k$ into pairwise prime numbers a_1, \dots, a_k then $\phi(n) = \phi(a_1) \cdots \phi(a_k)$.*

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

$$\phi(n) = n \cdot \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_k}\right) = \prod (p_i^{a_i} - p_i^{a_i-1}).$$

EXERCISE XV.24. (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 $n > 2$.

◇

EXERCISE XV.25. By looking at the smallest non-prime number, show that for composite numbers n the equation $a^p = a \pmod{n}$ may fail.

◇

EXERCISE XV.26. (Not easy). Formulate and prove a theorem like Fermat's little theorem in the case where p is not prime.

◇

EXERCISE XV.27. 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.28. Suppose n is a number with k digits a_k, \dots, a_0 : $n = a_k \cdot 10^k + \dots + a_1 \cdot 10 + a_0$. 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 .

◇

Index

- KV_4 , 30
- $r\%s$, 132
- Abelian, 34
- addition (in ring), 18
- additive opposite, 18
- automorphism, 52
- bijjective, 8
- Cantor, Georg, 9
- cardinality, 10
- Chinese Remainder Theorem, 133
- choose, 10
- Clay Millennium Problems, 26
- Cohen, Fred, 9
- continuum hypothesis, 9
- coprime, 24
- coset, 26
- cyclic group, 30
- dihedral group, 29
- division with remainder, 19
- element, 7
- empty set, 7
- Euclid, 23
- Euclidean Algorithm, 131
- Euclidean Algorithm, 20
- Euler ϕ -function, 24
- Fraenkel, Abraham, 9
- free group, 37
- function, 8
- Gödel, Kurt, 9
- generator, 38
- geometric series, 23
- Goldbach conjecture, 25
- group, 31
- harmonic series, 23
- Hilbert's List of 23 Problems, 26
- identity, 31
- identity (in ring), 18
- injective, 8
- inner automorphism, 52
- intersection, 10
- inverse, 31
- irreducible, 21
- Klein group, 30
- map, *see also* function
- modulo, 26
- morphism property, 35
- multiplication (in ring), 18
- negative, *see also* additive opposite
- neutral additive element (in ring), 18
- neutral element, 31
- neutral multiplicative element (in ring), 18
- norm function, 22
- order, 34
- power set, 10
- prime, 21
- Prime Number Theorem, 24
- relatively prime, 24
- Remainder Lemma, 103
- remainder under division, 19
- representative (of a coset), 26
- ring, 18, 87
- set, 7
 - element of, *see also* element
 - empty, 7
 - finite, 10
 - infinite, 10
- size of a set, 10
- source
 - of a function, 8
- special linear group, 54
- subgroup, 34
- subset, 7
- surjective, 8
- target
 - of a function, 8

union, 10
unique factorization, 21
unit, 130, 133
unit (in ring), 21

Zermelo, Ernst, 9
zerodivisor, 129, 130

Bibliography

- [Sil12] Joseph H. Silverman, *A friendly introduction to number theory, 4th edition*, Pearson Education, Inc., 2012.